CERN 95–06
22 November 1995
Vol. I

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CAS CERN ACCELERATOR SCHOOL

FIFTH ADVANCED ACCELERATOR PHYSICS COURSE

Hotel Paradise, Rhodes, Greece
20 September–1 October 1993

PROCEEDINGS
Editor: S. Turner
Vol. I

GENEVA
1995

© Copyright CERN, Genève, 1995

Propriété littéraire et scientifique réservée pour tous les pays du monde. Ce document ne peut être reproduit ou traduit en tout ou en partie sans l'autorisation écrite du Directeur général du CERN, titulaire du droit d'auteur. Dans les cas appropriés, et s'il s'agit d'utiliser le document à des fins non commerciales, cette autorisation sera volontiers accordée.

Le CERN ne revendique pas la propriété des inventions brevetables et dessins ou modèles susceptibles de dépôt qui pourraient être décrits dans le présent document; ceux-ci peuvent être librement utilisés par les instituts de recherche, les industriels et autres intéressés. Cependant, le CERN se réserve le droit de s'opposer à toute revendication qu'un usager pourrait faire de la propriété scientifique ou industrielle de toute invention et tout dessin ou modèle décrits dans le présent document.

Literary and scientific copyrights reserved in all countries of the world. This report, or any part of it, may not be reprinted or translated without written permission of the copyright holder, the Director-General of CERN. However, permission will be freely granted for appropriate non-commercial use.

If any patentable invention or registrable design is described in the report, CERN makes no claim to property rights in it but offers it for the free use of research institutions, manufacturers and others. CERN, however, may oppose any attempt by a user to claim any proprietary or patent rights in such inventions or designs as may be described in the present document.

ISSN 0007-8328
ISBN 92-9083-078-6

CERN 95-06
22 November 1995
Vol. I

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CERN ACCELERATOR SCHOOL

FIFTH ADVANCED ACCELERATOR PHYSICS COURSE

Hotel Paradise, Rhodes, Greece
20 September–1 October 1993

PROCEEDINGS
Editor: S. Turner
Vol. I

GENEVA
1995

CERN–Service d'information scientifique–RD/957–2500–novembre 1995

Abstract

The fifth CERN Accelerator School (CAS) advanced course on Accelerator Physics was given at the Paradise Hotel, Rhodes, Greece from 20 September to 1 October 1993. Its syllabus was based on the previous similar courses held at Oxford 1985, Berlin 1987, Uppsala 1989 and Noordwijkerhout 1991, and whose proceedings were published as CERN Reports 87-03, 89-01, 90-04 and 92-01, respectively. The present volumes are intended to replace and to bring up to date all the material in earlier publications. They contain not only all the lectures given in the Rhodes course but a number of important contributions to previous courses which are thought to be essential for a complete understanding of all aspects of the design and construction of particle accelerators at an advanced level. They include sections on Hamiltonian equations and accelerator optics, chromaticity and dynamic beam aperture, particle tracking, the kinetic theory, longitudinal beam optics, coherent instabilities, beam-beam dynamics, intra-beam scattering, beam cooling, Schottky noise, beam radiation, neutralisation, beam polarisation, radio-frequency quadrupoles, as well as chapters on space charge, superconducting magnets, crystal bending, beam-beam measurement and accelerator medical applications.

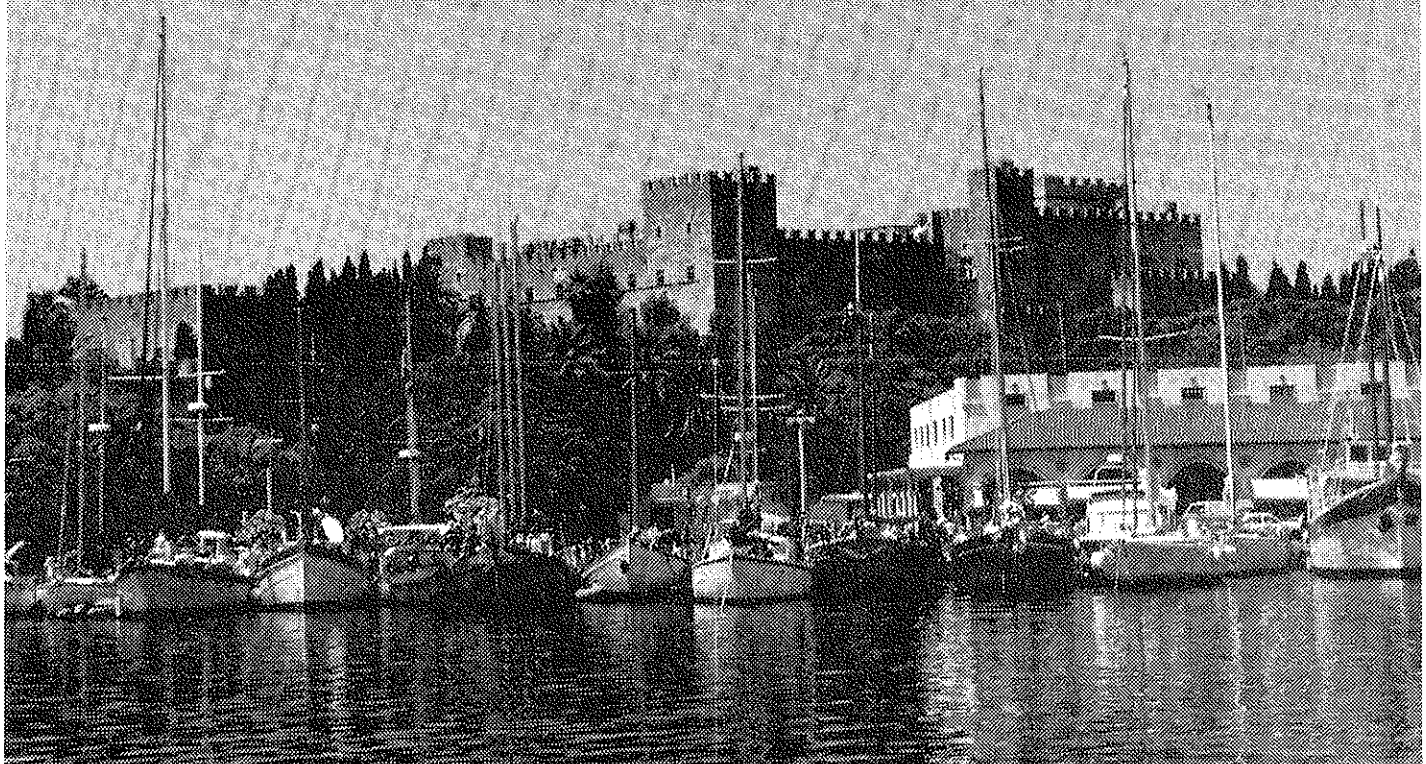
CERN ACCELERATOR SCHOOL
UNIVERSITY OF ATHENS—N.C.S.R. DEMOKRITOS

will organize a course on

ADVANCED ACCELERATOR PHYSICS

2th September—1 October, 1993
Hotel Paradise, Rhodes, Greece

This course will provide lectures for graduate and senior undergraduate students and staff members from accelerator laboratories, university departments and companies specializing in equipment for particle accelerators. It is aimed at knowledge, up to third-degree level, of physics, mathematics or engineering.



For further information and application forms, please contact:

Mrs. S. von Wartburg
CERN Accelerator School
CH-1211 Geneva 23
Switzerland

E-mail: CASRIO@CERNVM.CERN.CH
Telefax: +41 22 7824836

Head of School: E.J.N. Wilson

Programme Committee: K. Hübner, (Chairman) P.J. Bryant, C. Kounnas (Athens), Y. Goren, C. Hill, D. Möhl, S. Turner, E.J.N. Wilson

Local Organizing Committee: C. Kounnas (Athens), A. Storrs, S. Turner, S. von Wartburg

CLOSING DATE FOR APPLICATIONS: 15 JUNE 1993

Sponsored by: Greek Ministry of Culture
Greek Ministry of Industry, Energy and Technology
(General Secretariat for Research and Technology)

PROGRAMME FOR THE ADVANCED ACCELERATOR PHYSICS COURSE
Paradise Hotel, Rhodes, Greece, 20 September - 1 October 1993

Time	Monday 20 September	Tuesday 21 September	Wednesday 22 September	Thursday 23 September	Friday 24 September	Saturday 25 September	Sun. 26	Monday 27 September	Tuesday 28 September	Wednesday 29 September	Thursday 30 September
	BREAKFAST										
09.00	--	Longitudinal dynamics II	Stochastic Cooling	Vlasov equation and Landau damping	SR equilibrium beam distribution	Low emittance lattices		Wigglers	Beam-beam dynamics	Disruption and beam-strahlung	Spin rotators
09.30	-- Opening remarks										
10.00		W. Pirkl	D. Möhl	A. Hofmann	A. Ropert	A. Ropert		R. Walker	E. Keil	L. Rivkin	J.-P.Koutchouk
	COFFEE						E	COFFEE			
10.20	Basic Hamiltonian mechanics	Transverse dynamics I	Longitudinal dynamics III	Synchro betatron resonances	Betatron coupling	Particle tracking I	X	Instabilities in rings I	Instabilities in rings II	Instabilities in linacs I	Instabilities in linacs II
11.20	B. Montague	E. Wilson	W. Pirkl	G. Guignard	G. Guignard	F. Willeke	C	A. Hofmann	A. Hofmann	A. Mosnier	A. Mosnier
	MID-MORNING BREAK						U	MID-MORNING BREAK			
11.30	Longitudinal dynamics I	Hamiltonian tutorial	Transverse dynamics II	Chromaticity	Dynamic aperture	Particle tracking II	R	Wake fields	Intra-beam scattering	e^+e^- polarisation	Cures of instabilities
12.30	W. Pirkl		E. Wilson	A. Verdier	W. Scandale	F. Willeke	S	L. Palumbo	J. LeDuff	J.-P.Koutchouk	D. Boussard
	LUNCH BREAK						I	LUNCH BREAK			
16.00	Electron Cooling		Longitudinal dynamics tutorial	Transverse dynamics tutorial	Transverse dynamics tutorial		O	Poster session	Discussion	Contribution of N.Christofilos to accelerator physics	High luminosity
17.00	J. Bosser						N			A.Melissinos	J. LeDuff
	TEA	VISIT RHODES TOWN	TEA			FREE		TEA			
17.30	Canonical transformations		S.C. magnets		Crystal bending and extraction			Discussion of posters	Energy measurement by (resonant) depolarisation	NESTOR	Medical applications
18.30	E. Wilson		P. Schmüser		S. Pape-Møller				A.Melissinos	L. Resvanis	G. Kraft
	WELCOME COCKTAIL										
	EVENING MEAL									BANQUET	EVENING MEAL

Foreword

CERN's Accelerator School (CAS) was established in 1983 with the mission to preserve and disseminate the knowledge accumulated at CERN and elsewhere on particle accelerators and storage rings of all kinds. This is being carried out principally by means of a programme of basic, intermediate and advanced two-week courses on general accelerator physics aiming to bridge the gap between the level of knowledge attained with a science or engineering degree and that required for starting accelerator research work. In addition, topical courses are organized jointly with the US and Japanese Particle Accelerator Schools, while specialised courses are staged as the need or opportunity arises and in conjunction with appropriate bodies or organizations. This policy has led to the vigorous programme of courses and their proceedings shown in the table overleaf.

For the basic accelerator physics course held in 1992 the opportunity was taken to bring together written versions of the lectures presented at the school as well as revised versions of the most appropriate chapters from the proceedings of the four previous basic courses. In this way a comprehensive coverage of the topic was presented in one set of proceedings — CERN 94-01. The same procedure has been followed in preparing the present volumes. The basis is a complete set of lectures presented in the advanced accelerator physics course in Rhodes to which has been added new versions of other essential lectures from previous schools. It is hoped that this recompilation will greatly facilitate the task of those wishing to find information in the accelerator field. It is to be hoped that these present volumes remain the standard work of reference for many years to come so that only radically new material need be published separately in future proceedings. These latest volumes, together with those of the 1992 basic course span the spectrum of needs from the beginner to the expert. The recently introduced intermediate level course is based on the more advanced material from the 1992 proceedings and the less detailed material from these present volumes. Many organizations and individuals played major roles in making it possible to hold the fifth version of the CAS advanced accelerator physics course in Rhodes, Greece. In particular, we should mention the Greek Ministry of Culture, the Greek Ministry of Industry, Energy and Technology (General Secretariat for Research and Technology), the University of Athens — N.C.S.R. Demokritos and members of their staff, especially Professors P. Gemtos, A.D. Panagioutou and L.K. Resvanis, Dr. I. Siotio and particularly Dr. C. Kourkoumelis whose help and patience were invaluable in settling all the local organizational details. On this occasion, with its complete revision of the proceedings, very special thanks must go not only to the lecturers at the school who made such serious efforts to prepare, present and write-up their topics, but also to the many authors who carefully revised and up-dated topics from previous courses. The support of the CERN Management and CAS Advisory, Programme and Local Organization Committees was as good as ever and is acknowledged. Finally we thank the many institutes who supported the CAS by sending students to the course, as well as the students themselves whose interest and enthusiasm amply repaid all the effort made on their behalf.

S. Turner, Editor

List of CAS Courses and their Proceedings

Year	General Accelerator Physics Courses	Topical Courses	Specialised Courses and Workshops
1983	–	–	Antiprotons for colliding beam facilities CERN 84-15 (1984)
1984	Basic CERN 85-19 (1985)	–	Generation of high fields (ECFA and INFN Workshop) ECFA 85/91, CERN 85-07 (1985)
1985	Advanced CERN 87-03 (1987)	Nonlinear dynamics aspects of particle accelerators Lecture Notes in Physics 247 (Springer Verlag, Berlin, 1986)	–
1986	Basic CERN 87-10 (1987)	Frontiers of particle beams Lecture Notes in Physics 296 (Springer Verlag, Berlin, Heidelberg, New York, 1988)	Applied Geodesy for particle accelerators CERN 87-01 (1987) Applied Geodesy Lecture Notes in Earth Sciences 12 (Springer-Verlag, Berlin, Heidelberg, New York, 1987)
1987	Advanced CERN 89-01 (1989)	–	New developments in particle acceleration techniques (ECFA et al. Workshop) CERN 87-11, ECFA 87/110
1988	Basic CERN 89-05 (1989)	Frontiers of particle beams; Observation, diagnosis and correction Lecture Notes in Physics 343 (Springer Verlag, Berlin, Heidelberg, New York, 1989)	Superconductivity in particle accelerators CERN 89-04 (1989)
1989	Advanced CERN 90-04 (1990)	–	Synchrotron radiation and free-electron lasers CERN 90-03 (1990)
1990	Basic CERN 91-04 (1991)	Frontiers of particle beams: Intensity limitations Lecture Notes in Physics 400 (Springer-Verlag, Berlin, Heidelberg, New York, 1992)	Power converters for particle accelerators CERN 90-07 (1990)
1991	Advanced CERN 92-01 (1992)	–	RF engineering for particle accelerators CERN 92-03 (1992), 2 Vols.
1992	Basic CERN 94-01 (1994)	Frontiers of particle beams: Factories with e^+e^- rings (to be published)	Magnetic measurements and alignment CERN 92-05 (1992)
1993	Advanced Present volumes	–	RF engineering for particle accelerators Repeat of 1991 course
1994	Basic	Frontiers of accelerator technology (to be published)	Cyclotrons, linacs and their applications (to be published)
1995	Intermediate	–	Superconductivity in particle accelerators (to be published)

Contents

Vol. I

Vol. II

BASIC HAMILTONIAN MECHANICS

B.W. Montague

1. INTRODUCTION

At the time of Newton, mechanics was considered mainly in terms of forces, masses and velocities, since these quantities were directly tangible in terms of everyday experience. However, the subsequent evolution of celestial mechanics called for more compact and general methods of handling dynamical systems, and led to the increasing use of potentials as the relevant quantities in the mathematical treatment. Another important development of this period was to free classical mechanics from the constraints of specific co-ordinate systems and to facilitate the transformation from one system to another. This is not only a matter of convenience but a powerful tool for finding invariants of the motion, and a fundamental feature of the Hamiltonian formulation.

Joseph-Louis Lagrange was one of the outstanding pioneers of this development; his Méchanique Analytique laid the foundations of the analytic, in contrast with the geometrical, approach to generalised dynamics. Nevertheless he suggested, apparently as a concession to tradition, that mechanics (with the time dimension included), might be considered as the geometry of a four-dimensional space, which was in a sense a precursor of the structure of special relativity.

Generalised classical mechanics has developed considerably since the time of Lagrange and remains not only a broad and fundamental part of classical physics as a whole but has become an important part of the framework on which quantum mechanics has been formulated. The range of topics is so large that even in the restricted field of particle accelerators our coverage here of Lagrangian and Hamiltonian dynamics can only be rather limited. More detailed presentations of the subjects can be found in the Bibliography and are cited in the text wherever the need is specially manifest.

The Lagrange equations of motion can be presented in a number of different versions, depending on the nature of the problem and the form of the dynamical constraints. A detailed discussion of these topics is given by Goldstein [1] in Chapter 1; for the application to accelerator physics and the Hamiltonian formulation it is sufficient to consider a restricted set of conditions, namely the motion of charged particles in electromagnetic fields, a domain which is thoroughly covered by Jackson [2] in Chapter 12. Furthermore we shall be dealing with the motion of single particles, taking no account of the forces due to space charge.

In the Lagrangian formulation the dynamical behaviour of a system with k degrees of freedom is characterised by a set of generalised "position" co-ordinates q_k, generalised "velocity" co-ordinates $\dot{q}_k = dq_k / dt$ and the independent variable t, which is often, but not necessarily, the time variable. The "Lagrangian" or Lagrange function L of the form:

$$L(q_k, \dot{q}_k, t) \tag{1}$$

is then a function of $2k$ dynamical variables.

2. OUTLINE OF LAGRANGIAN AND HAMILTONIAN FORMALISM

In the simplest, non-relativistic case where the forces can be derived from a scalar potential V, independent of velocity, the Lagrangian takes the specific form:

$$L(q,\dot{q},t) = T(q,\dot{q},t) - V(q,t), \tag{2}$$

where T is the kinetic energy, V is the potential energy and the index k is implicit.

In the presence of electromagnetic fields, which can be time-dependent, a generalised potential, and consequently a Lagrangian, can be formulated by combining the Lorentz equation for the force with Maxwell's equations for the electromagnetic fields.

The Lorentz equation for the force $\boldsymbol{F}$ on a particle of charge e moving with velocity $\boldsymbol{v}$ in an electromagnetic field is given by

$$\boldsymbol{F} = e[\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}] \ , \tag{3}$$

where $\boldsymbol{E}$ and $\boldsymbol{B}$ are respectively the electric and magnetic fields. Now in the case of static fields the part of the force $\boldsymbol{F}$ arising from $\boldsymbol{E}$ can be derived from a scalar potential ϕ, but the presence of a magnetic field and time-variation of either, requires the introduction of a vector potential $\boldsymbol{A}$, which contributes to both electric and magnetic forces.

Since we are not taking account here of space-charge forces, only two of Maxwell's equations are required, namely:

$$\text{curl}\, \boldsymbol{E} + d\boldsymbol{B} / dt = 0 \tag{4}$$

$$\text{div}\, \boldsymbol{B} = 0 \ . \tag{5}$$

From Eq. (5) it follows that one can write:

$$\boldsymbol{B} = \text{curl}\, \boldsymbol{A} \ , \tag{6}$$

where $\boldsymbol{A}$ is the magnetic vector potential. Substituting Eq. (6) into Eq. (4) expanding and rearranging terms, it is then straightforward but somewhat lengthy to show that the non-relativistic Lagrangian for time-dependent electromagnetic fields is:

$$L(q,\dot{q},t) = T(q,\dot{q},t) - U(q,\dot{q},t) \ , \tag{7}$$

where the scalar potential V of Eq. (2) has been replaced by a generalised potential:

$$U = e\phi - \boldsymbol{A} \cdot \boldsymbol{v} \ . \tag{8}$$

The relativistic Lagrangian is not just the difference between kinetic and potential energies, and is more complicated to derive formally. A full discussion is given in Chapter 7 of Goldstein [1]; here we simply present the form appropriate to accelerator dynamics, viz:

$$L = -mc\left[c^2 - \boldsymbol{v} \cdot \boldsymbol{v}\right]^{1/2} - e\phi + e\boldsymbol{A} \cdot \boldsymbol{v} \ . \tag{9}$$

We note that in the non-relativistic limit, $v << c$, this reduces almost to the form of Eq. (7), apart from a constant $-mc^2$, which vanishes on subsequent differentiations.

In the framework of Hamiltonian theory the importance of the Lagrangian lies in the formulation of Hamilton's Principle of Stationary Action (sometimes called "least action" which is usually but not always the case). This principle states that the action integral defined by:

$$S = \int_{t_1}^{t_2} L(q,\dot{q},t)dt \ , \tag{10}$$

is an extremum for the dynamically true path of the time trajectory between t_1 and t_2, i.e.

$$\delta S = \delta \int_{t_1}^{t_2} L dt = 0 \text{ to first order} \ . \tag{11}$$

The evaluation of this by the calculus of variations, which is very clearly explained in Vol. II of the Feynmann Lectures [3], results in the Lagrangian equations of motion for a conservative system:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} = 0 \ . \tag{12}$$

The quantity $(\partial L / \partial \dot{q}_k)$ is known as the canonical momentum; in the simplest cases where there is no vector potential it is the same as the mechanical momentum, but the presence of an $\boldsymbol{A} \cdot \boldsymbol{v}$ term gives rise to a so-called magnetic momentum. One then recognises that Eq. (12) is Newton's Second Law in a more modern guise.

Summarising, for a system of k degrees of freedom the Lagrangian is a function of $2k$ dynamical variables (generalised co-ordinates and velocities) and the "time" t, k of these variables $(\dot{q}_k)$ being the "time" derivatives of the other k variables (q_k). The corresponding Lagrange equations consist of a set of k second-order differential equations describing the motion of the system.

The Hamiltonian formulation of mechanics describes a system in terms of generalised co-ordinates (q_k) and generalised momenta (p_k), the same as the canonical momenta we identified in Eq. (12) above. The change of basis from the set $(q_k, \dot{q}_k, t)$ to the set (q_k, p_k, t) is obtained through a Legendre transformation, defined by the function:

$$H(q,p,t) = \sum_k p_k \ \dot{q}_k - L(q,\dot{q},t) \ . \tag{13}$$

We can take the differential of H on the left-hand side as

$$dH = \frac{\partial H}{\partial t} dt + \sum_k \frac{\partial H}{\partial q_k} dq_k + \sum_k \frac{\partial H}{\partial p_k} dp_k \tag{14}$$

or alternatively, from the right-hand side of (13) as

$$dH = \sum_k p_k \ d\dot{q}_k + \sum_k \dot{q}_k \ dp_k - \sum_k \frac{\partial L}{\partial q_k} dq_k - \sum_k \frac{\partial L}{\partial \dot{q}_k} \ d\dot{q}_k - \frac{\partial L}{\partial t} dt \ . \tag{15}$$

Designating the canonical momenta of Eq. (12) as

$$p_k = \frac{\partial L}{\partial \dot{q}_k} \ , \tag{16}$$

the first and fourth summations cancel and the remaining terms can be identified with the corresponding terms in (14)

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \tag{17}$$

$$\dot{q}_k = \frac{\partial H}{\partial p_k} \tag{18}$$

$$\dot{p}_k = -\frac{\partial H}{\partial q_k} \ . \tag{19}$$

The function $H(q,p,t)$ is the Hamiltonian and Eqs. (18) and (19) are the Hamilton equations of motion. They are first order, 2k in number for k degrees of freedom, and show a remarkable symmetry of form between the generalised position co-ordinates q_k and their conjugate momenta p_k. This symmetry leads to very flexible transformation properties between sets of dynamical variables.

For non-relativistic motion the Hamiltonian is often, though not necessarily, the sum of potential and kinetic energies

$$H(q,p,t) = T + U \ . \tag{20}$$

A relativistic Hamiltonian for a single particle in an electromagnetic field can be derived from the Lagrangian of Eq. (9). In Cartesian co-ordinates, $k = x, y, z$, the canonical momenta given by (16) are

$$p_k = m_0 \gamma \ v_k + eA_k \tag{21}$$

and differ from the component $m_0 \gamma v_k$ of the mechanical momentum by the contribution eA_k, the electromagnetic momentum. The resulting Hamiltonian is easily shown to be

$$H(q,p,t) = e\phi + c\left[(\boldsymbol{p} - e\boldsymbol{A})^2 + m_0^2 c^2\right]^{1/2} \ , \tag{22}$$

where the p_k have been expressed in vector form.

3. SOME PROPERTIES OF THE HAMILTONIAN

From the Hamiltonian $H(q_k, p_k, t)$ the Hamilton equations of motion are obtained by

$$\dot{q}_k = \frac{dq_k}{dt} = \frac{\partial H}{\partial p_k} \tag{18}$$

$$\dot{p}_k = \frac{dp_k}{dt} = -\frac{\partial H}{\partial q_k} \ . \tag{19}$$

The Poisson bracket of any two dynamical variables $f(q_k,p_k,t)$ and $g(q_k,p_k,t)$ is defined by

$$\{f,g\} \equiv \sum_k \left\{ \frac{\partial f}{\partial q_k}\frac{\partial g}{\partial p_k} - \frac{\partial f}{\partial p_k}\frac{\partial g}{\partial q_k} \right\} . \tag{23}$$

One sees that

$$\{f,f\} \equiv 0 ,$$
$$\{g,f\} = -\{f,g\} .$$

The time derivative of f

$$\begin{aligned} \frac{df}{dt} &= \frac{\partial f}{\partial t} + \sum_k \left\{ \frac{\partial f}{\partial q_k}\dot{q}_k + \frac{\partial f}{\partial p_k}\dot{p}_k \right\} \\ &= \frac{\partial f}{\partial t} + \{f,H\} \quad \text{(from (18), (19) and (23)} . \end{aligned} \tag{24}$$

In particular, if $f = H$ we have

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} . \tag{25}$$

If H does not depend *explicitly* on time it is a constant of the motion. Any invariant of the motion not containing t explicitly, has a vanishing Poisson bracket with H.

4. CANONICAL TRANSFORMATIONS

Canonical transformations are of considerable utility in simplifying problems by an advantageous choice of co-ordinate system. In particular they can be used to reduce the number of degrees of freedom of a system by exposing invariant quantities, or quantities that are "almost invariant" apart from a small parameter, permitting perturbation theory to be applied. In non-linear systems they often enable the simple linear part of the motion to be transformed away so that the non-linear part can more easily be handled, without prejudice to the linear motion. Examples of this are treated by Schoch [4] and by Montague [5], the latter case dealing with space-charge forces.

The form of the equations is preserved in transforming between co-ordinate systems (q_k,p_k) and (Q_k,P_k). The necessary and sufficient condition for a transformation to be canonical is

$$\left[\sum_k P_k \; dQ_k - H_1(Q_k,P_k,t)dt \right] - \left[\sum_k p_k \; dq_k - H(q_k,p_k,t)dt \right] = dG \tag{26}$$

where dG is a total differential. This follows from Hamilton's variational principle

$$\delta \int_{t_1}^{t_2} L \; dt = \delta \int_{t_1}^{t_2} \left[\sum_k p_k \, \dot{q}_k - H(q_k, p_k, t) \right] dt = 0 \tag{27}$$

and Eq. (13). The canonical transformation

$$Q_k = Q_k(q_1, q_2 \ldots p_1, p_2 \ldots t) \tag{28}$$

$$P_k = P_k(q_1, q_2 \ldots p_1, p_2 \ldots t) \tag{29}$$

is derived from the generating function G which can have one of four forms:

$$G_1(q,Q,t): \quad p_k = \frac{\partial G}{\partial q_k} \; ; \quad P_k = -\frac{\partial G}{\partial Q_k} \tag{30a}$$

$$G_2(q,P,t): \quad p_k = \frac{\partial G}{\partial q_k} \; ; \quad Q_k = \frac{\partial G}{\partial P_k} \tag{30b}$$

$$G_3(Q,p,t): \quad P_k = -\frac{\partial G}{\partial Q_k} \; ; \quad q_k = -\frac{\partial G}{\partial p_k} \tag{30c}$$

$$G_4(p,P,t): \quad q_k = -\frac{\partial G}{\partial p_k} \; ; \quad Q_k = \frac{\partial G}{\partial P_k} \tag{30d}$$

For all forms the new Hamiltonian is

$$H_1(Q_k, P_k, t) = H(q_k, p_k, t) + \frac{\partial G}{\partial t} \tag{31}$$

5. INTEGRAL INVARIANTS, LIOUVILLE'S THEOREM

The action integral

$$S = \int_{t_1}^{t_2} L \; dt \tag{32}$$

is invariant under a canonical transformation. So also are the phase space integral invariants (Poincaré invariants):

$$J_1 = \iint \sum_k dp_k dq_k \tag{33a}$$

$$J_2 = \iint \iint \sum_{k,i} dp_i dq_k dq_i dq_k \tag{33b}$$

$$J_n = \int \ldots \int dp_1 \ldots dp_n dq_1 \ldots dq_n \tag{34}$$

where the integrals are taken over any arbitrary phase-space submanifold of appropriate dimension (2, 4 ... 2n respectively). For J_1, corresponding to one degree of freedom, i.e. a two-dimensional phase space, the invariance is that of area conservation in the (q,p) plane of a closed trajectory. For sub-manifolds of dimensions between 4 and (2n-2) the interpretation of the Poincaré invariants is much more complicated and we shall not discuss it further here. For a system with n degrees of freedom, the invariance of J_n is Liouville's theorem.

6. LINEAR OSCILLATOR

The energies are:

kinetic $T = \frac{m}{2}\dot{q}^2$ ($m =$ mass)

potential $U = \frac{k}{2}q^2$ ($k =$ spring constant, not to be confused with index) .

The Lagrangian is:

$$L(q,\dot{q}) = T - U = \frac{m}{2}\dot{q}^2 - \frac{k}{2}q^2$$

and the canonical momentum:

$$p = \frac{\partial L}{\partial \dot{q}} = m\ \dot{q}\ ,$$

whence $\dot{q} = \frac{p}{m}$. Using this to replace $\dot{q}$ in T, the Hamiltonian becomes:

$$H = T + U = \frac{p^2}{2m} + \frac{k\ q^2}{2}$$

and the Hamilton equations of motion are:

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m} \quad \text{and} \quad \dot{p} = -\frac{\partial H}{\partial q} = -k\ q\ .$$

These can be expressed as one second-order equation

$$\ddot{q} = \frac{\dot{p}}{m} = -\frac{k}{m}q \quad \text{or} \quad \ddot{q} + \omega^2 q = 0 \quad \text{where} \quad \omega = \sqrt{\frac{k}{m}}\ .$$

We can transform to a new co-ordinate system of action-angle variables using a canonical transformation with the generating function

$$G(q,Q) = \frac{\sqrt{k\ m}}{2}q^2 \cot Q\ . \tag{35}$$

Then using (30a) one has

$$p = \frac{\partial G}{\partial q} = \sqrt{k\ m} \cdot q \cot Q$$

and

$$P = -\frac{\partial G}{\partial q} = \frac{\sqrt{k\ m}}{2} q^2 \csc^2 Q \ ,$$

whence

$$q^2 = \frac{2P \sin^2 Q}{\sqrt{k\ m}} \quad \text{and} \quad p^2 = 2P\sqrt{k\ m} \cos^2 Q \ .$$

Also,

$$H_1 = H + \frac{\partial G}{\partial t} = H \quad (\text{since } \frac{\partial G}{\partial t} = 0) \ ,$$

so

$$H_1(Q,P) = P\sqrt{\frac{k}{m}} \cos^2 Q + P\sqrt{\frac{k}{m}} \sin^2 Q = P\sqrt{\frac{k}{m}} = \omega \ P \ .$$

The transformed Hamilton equations of motion are then

$$\dot{p} = -\frac{\partial H_1}{\partial Q} = 0 \ ; \quad (P = \text{constant and } H_1 \text{ is cyclic in } Q)$$

and

$$\dot{Q} = \frac{\partial H_1}{\partial p} = \omega \ ; \quad \left(Q = \omega t + \phi \quad \text{and} \quad q = \sqrt{\frac{2P}{m\omega}} \sin(\omega t + \phi) \right) \ .$$

This type of canonical transformation is useful in some non-linear problems to transform away the linear part of the motion and treat the non-linear part by perturbation theory.

7. SIMPLE PENDULUM

$$U = m\ g\ \ell(1 - \cos\theta)$$

$$T = \frac{m\ v^2}{2} = \frac{m\ \ell^2\ \dot{\theta}^2}{2}$$

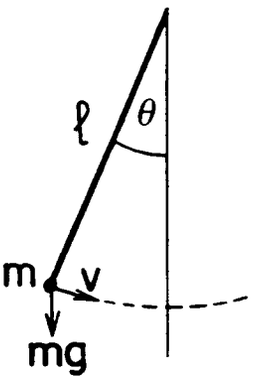
$$L(\theta, \dot{\theta}) = T - U = \frac{m\ \ell^2\ \dot{\theta}^2}{2} - m\ g\ \ell(1 - \cos\theta)$$

$$p = \frac{\partial L}{\partial \dot{\theta}} = m\ \ell^2\ \dot{\theta} \ , \quad \text{so} \quad \dot{\theta} = \frac{p}{m\ \ell^2}$$

$$H(\theta, p) = T + U = \frac{p^2}{2m\ \ell^2} + m\ g\ \ell(1 - \cos\theta)$$

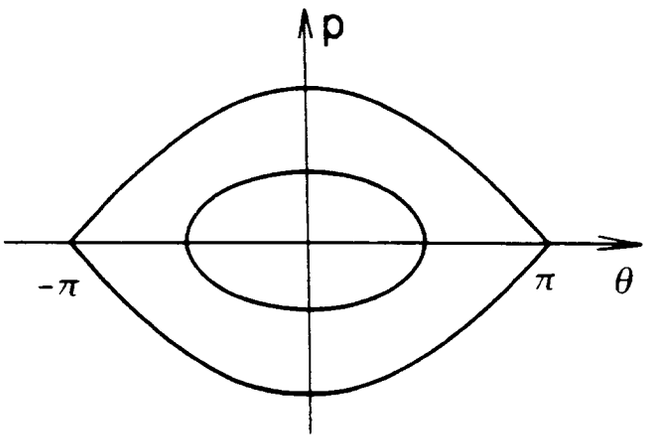
$$\dot{p} = -\frac{\partial H}{\partial \theta} = -m\ g\ \ell \sin\ \theta$$

$$\ddot{\theta} = \frac{\dot{p}}{m\ \ell^2} = -\frac{g}{\ell} \sin\ \theta \qquad \ddot{\theta} + \frac{g}{\ell} \sin\ \theta = 0$$



The Hamiltonian can be written

$$H(\theta,p)=\frac{p^2}{2m\ \ell^2}+2m\ g\ \ell\sin^2\frac{\theta}{2}=\text{ constant}\quad\left(\text{since }\frac{dH}{dt}=\frac{\partial H}{\partial t}=0\right).$$



Trajectories of constant H in (θ,p) phase plane.

This example is closely analogous to the problem of phase oscillations in a synchrotron or storage ring.

8. RELATIVISTIC HAMILTONIAN WITH ELECTROMAGNETIC FIELDS

The relativistic Lagrangian for a single particle in an electromagnetic field is, using vector notation

$$L=-m_0c^2\sqrt{1-\beta^2}+e\boldsymbol{A}\cdot\boldsymbol{v}-e\phi \tag{36}$$

where $\boldsymbol{A}$ is the vector potential and ϕ the scalar potential.

Note that $L\neq T-U$ relativistically. However, the canonical momenta p_k are still obtained from

$$p_k=\frac{\partial L}{\partial\dot{q}_k}=m_0\gamma\ v_k+eA_k \tag{37}$$

but are no longer the same as the mechanical momenta $m_0\gamma v_k$ in the presence of a velocity-dependent potential $\boldsymbol{A}$.

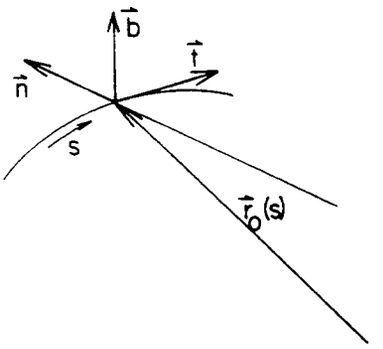
The relativistic single-particle Hamiltonian in the electromagnetic field becomes

$$H(\boldsymbol{q},\boldsymbol{p},t)=e\phi+c\left[(\boldsymbol{p}-e\boldsymbol{A})^2+m_0^2c^2\right]^{1/2}. \tag{38}$$

Since $(\boldsymbol{p}-e\boldsymbol{A})^2=(m_0c\ \beta\ \gamma)^2$, $H=e\phi+m_0c^2\gamma$ has the value of the total energy, including the rest energy.

9. APPLICATION TO AN ACCELERATOR ORBIT

We take as reference curve $\boldsymbol{r}_0(s)$ and express any neighbouring point $\boldsymbol{r}(x,y,s)$ in terms of the Frenet unit vectors



$\boldsymbol{t}(s)=$ tangent

$\boldsymbol{n}(s)=$ normal

$\boldsymbol{b}(s)=$ binormal

$$\boldsymbol{r}(\mathrm{x},\mathrm{y},\mathrm{s})=\boldsymbol{r}_0(s)+x\ \ \boldsymbol{n}(s)+y\ \ \boldsymbol{b}(s)$$

The following properties hold

$$\begin{aligned}&\frac{d\boldsymbol{r}_0}{ds}=\boldsymbol{t}\,,\qquad \frac{d\boldsymbol{t}}{ds}=\kappa\ \ \boldsymbol{n}\,,\\ &\frac{d\boldsymbol{n}}{ds}=\tau\ \ \boldsymbol{b}-\kappa\ \ \boldsymbol{t}\,,\qquad \frac{d\boldsymbol{b}}{ds}=-\tau\ \ \boldsymbol{n}\,,\end{aligned}\tag{39}$$

where $\kappa=\dfrac{1}{\rho}$ is the curvature and τ the torsion of the reference curve.

The canonically-conjugate momenta $\boldsymbol{p}$ in the co-ordinate system $\boldsymbol{r}$ are obtained from the generating function

$$\begin{aligned}&G_2(\boldsymbol{r},\boldsymbol{p})=\boldsymbol{p}\cdot\boldsymbol{r}\\ &=\boldsymbol{p}\cdot\left[\boldsymbol{r}_0(s)+x\boldsymbol{n}(s)+y\boldsymbol{b}(s)\right]\end{aligned}\tag{40}$$

which generates the identity transformation

$$p_k=\frac{\partial G}{\partial r_k}=P_k \tag{41a}$$

$$Q_k=\frac{\partial G}{\partial P_k}=r_k \tag{41b}$$

and since $\dfrac{\partial G}{\partial t}=0$, the Hamiltonian is unchanged. From Eqs. (40) and (41) we have, using (32)

$$\begin{aligned}&p_s=\frac{\partial G}{\partial s}\\ &=\boldsymbol{p}\cdot\left[\frac{d\boldsymbol{r}_0}{ds}+x\frac{d\boldsymbol{n}}{ds}+y\frac{d\boldsymbol{b}}{ds}\right]\\ &=\boldsymbol{p}\cdot\left[\boldsymbol{t}+x(\tau\ \ \boldsymbol{b}-\kappa\ \ \boldsymbol{t})-y\ \ \tau\ \ \boldsymbol{n}\right]\\ &=\boldsymbol{p}\cdot\left[(1-\kappa\mathrm{x})\boldsymbol{t}+\tau(x\ \ \boldsymbol{b}-y\ \ \boldsymbol{n})\right]\end{aligned}\tag{42}$$

If we now define the reference curve to be plane, the torsion τ vanishes. Taking also the other canonical momenta from (41a) we have

$$\begin{aligned} p_x &= \boldsymbol{p}\cdot\boldsymbol{n} \\ p_y &= \boldsymbol{p}\cdot\boldsymbol{b} \\ p_s &= \boldsymbol{p}\cdot\boldsymbol{t}(1-\kappa\ x) \end{aligned} \tag{43}$$

The vector potential $\boldsymbol{A}$ transforms similarly

$$\begin{aligned} A_x &= \boldsymbol{A}\cdot\boldsymbol{n} \\ A_y &= \boldsymbol{A}\cdot\boldsymbol{b} \\ A_s &= \boldsymbol{A}\cdot\boldsymbol{t}(1-\kappa\ x) \end{aligned} \tag{44}$$

One notes that p_S, A_S are generally *not* the components in the tangential direction of the reference curve.

The Hamiltonian (38) can now be written explicitly in terms of the canonical momenta (43) and the vector potential components (44). We choose the Coulomb gauge for Maxwell's equations to make $\phi = 0$ and obtain

$$H(q_k,p_k,t) = c\left[\left(p_x - eA_x\right)^2 + \left(p_y - eA_y\right)^2 + \left\{\frac{p_s - eA_s}{1-\kappa\ x}\right\}^2 + m_0^2c^2\right]^{1/2} \tag{45}$$

where the A_k are functions of position and time in general.

For a cyclic accelerator it is convenient to use s as the independent variable instead of t, since the fields arising from $\boldsymbol{A}$ can readily be expanded in Fourier components periodic around the circumference. Hamilton's principle in Eq. (27) can be re-expressed in terms of an integral along a space trajectory s rather than a time trajectory t, since we are considering the motion of a single particle with a one-one correspondence between time and position. The change of independent variable from t to s is then trivial and we can write Hamilton's principle in the form:

$$\delta\int_{s_1}^{s_2}\left[p_x\frac{dx}{ds} + p_y\frac{dy}{ds} - H\frac{dt}{ds} + p_s\right]ds = 0 \tag{46}$$

where we now consider the new Hamiltonian to be

$$-p_s\left(x,y,t,p_x,p_y,-H,s\right) = F \text{ say }, \tag{47}$$

and where $(t, -H)$ is a new pair of canonically-conjugate variables. The satisfying of Hamilton's variational principle (11) guarantees the validity of the transformation, and the previous results may be taken over with the appropriate changes in notation.

Rewriting (45) in the new form

$$F\left(x,y,t,p_x,p_y,-H,s\right)=-p_s$$

$$=-eA_s-(1-\kappa\ x)\left[\frac{1}{c^2}\left(H^2-m_0^2c^4\right)-\left(p_x-eA_x\right)^2-\left(p_y-eA_y\right)^2\right]^{1/2} \tag{48}$$

where numerically

$$\frac{1}{c^2}\left(H^2-m_0^2c^4\right)=m_0^2c^2\left(\gamma^2-1\right)=m_0^2c^2\beta^2\gamma^2=\boldsymbol{p}\cdot\boldsymbol{p} \tag{49}$$

is the square of the total momentum.

The Hamilton equations of motion are, following Eqs. (18) and (19)

$$x'=\frac{\partial F}{\partial p_x}\ ;\quad y'=\frac{\partial F}{\partial p_y}\ ;\quad t'=-\frac{\partial F}{\partial H} \tag{50a}$$

$$p_x'=-\frac{\partial F}{\partial x}\ ;\quad p_y'=-\frac{\partial F}{\partial y}\ ;\quad H'=-\frac{\partial F}{\partial t} \tag{50b}$$

where the primes denote $\frac{d}{ds}$.

In general the components of the vector potential A are functions of all the dynamical variables and of the independent variable s (the distance along the reference trajectory). The curvature κ, which introduces a kinematic non-linear term (usually small), is generally a function of s. With a given accelerator or storage-ring lattice these terms can be calculated for all the elements (dipoles, quadrupoles, etc.) and expanded as Fourier series in s around the circumference of the machine. The Hamiltonian F in Eq. (48) can then be expressed as a power series expansion in the dynamical variables up to any order desired. The equations of motion may be obtained from Eqs. (50); however, the Hamiltonian can conveniently be used directly for the study of resonance behaviour by selection of near-resonant terms in the machine periodicities and by judicious approximations. Although these procedures are well defined and straightforward in principle, the calculation of the power-series coefficients is in practice rather laborious for all but the simplest cases.

Certain simplifications can often be made in Eq. (48). If A is not an explicit function of t then $H' = 0$ and H is an invariant of the motion and is called a cyclic variable. This corresponds in (49) to a constant value of $m_0\beta\gamma$, t is an ignorable co-ordinate and the degrees of freedom are reduced from three to two. Such a situation arises for a coasting beam in a storage ring, and as an approximation to the motion in an accelerator on a sufficiently short time scale. Also it is sometimes permissible to approximate bending magnets, quadrupoles, etc. by piecewise constant elements with $A_X = A_Y = 0$, except at the ends where they can be expressed as δ-functions. The only component of the vector potential is

$$A_s=\frac{p_0K}{2e}\left(x^2-y^2\right) \tag{51}$$

where $p_0 = m_0c\beta\gamma$ and K is the gradient parameter. The reference orbit is the axis of the quadrupole and the curvature $\kappa = 0$. Then Eq. (48) simplifies to

$$F(x,y,p_x,p_y) = -\frac{p_0 K}{2}(x^2 - y^2) - \left[p_0^2 - p_x^2 - p_y^2\right]^{1/2}$$

The equations of motion (50) then become, assuming $p_0^2 >> p_x^2 + p_y^2$

$$\dot{p_x} = -\frac{\partial F}{\partial x} = p_0 \;\; K \;\; x \tag{52a}$$

$$\dot{p_y} = -\frac{\partial F}{\partial y} = p_0 \;\; K \;\; y \tag{52b}$$

$$x' = \frac{\partial F}{\partial p_x} = \frac{p_x}{\left[p_0^2 - p_x^2 - p_y^2\right]^{1/2}} \approx \frac{p_x}{p_0} \tag{52c}$$

$$y' = \frac{F}{p_y} = \frac{p_y}{\left[p_0^2 - p_x^2 - p_y^2\right]^{1/2}} \approx \frac{p_y}{p} \tag{52d}$$

leading to the familiar form of the equations of motion

$$x'' - K \;\; x = 0$$
$$y'' + k \;\; y = 0$$

10. ADIABATIC INVARIANTS

In Section 5 we briefly discussed integral invariants, the invariance property being with respect to canonical transformations of phase-space integrals. There is another type of invariance related to the behaviour of a system under slow changes of a parameter normally considered as constant. The original classical model of this Adiabatic Principle, formulated by Einstein around 1911, supposed a weight suspended by a string passing through a hole in the ceiling, forming a pendulum whose frequency could be slowly changed by raising or lowering the string. Einstein showed that, provided the fractional change in frequency during one oscillation period is sufficiently small, the energy divided by the frequency, which is nothing else but the action variable S, is a constant. This principle was of immediate consequence for the understanding of quantum theory and the stability of atoms under changes of environmental fields; subsequently it became of great practical significance in many other branches of physics. In accelerator theory adiabatic invariance is of fundamental importance in ensuring the stability of beams under conditions of changing parameters, as happens during the acceleration cycle in a synchrotron. In particular, an analysis of the phase-space stacking process in proton storage rings requires a full understanding of the adiabaticity conditions.

The basic principle of adiabatic invariance, as formulated in the pendulum with slowly varying frequency, is intuitively rather simple to grasp. For a constant frequency, and in the absence of damping, it is obvious that the sum of potential and kinetic energies is constant. If now the pendulum is gradually shortened, gravitational energy is being supplied to the system by an external force at the same time as the frequency is slowly changing. It seems intuitively plausible that for sufficiently slow changes something must be conserved and the main questions that arise are (i) what is conserved? and (ii) what constitutes "sufficiently slow" in this context?

Surprisingly, these questions are not always very simple to answer, and in many cases need to be examined in the framework of canonical perturbation theory. The subject of adiabatic invariance is covered at some length in Section II-7 of Goldstein [1]; here we limit the discussion to an outline of the method and its application to a simple example.

Since we are interested in systems which involve some kind of periodic or quasi-periodic behaviour, the first step is to apply canonical transformations to the Hamiltonian in order to bring it into the form of action-angle variables, as we did in Section 6 for the linear harmonic oscillator. When we derived the new Hamiltonian H_1 using the generating function $G(q,Q)$ of Eq. (35), it had the same value as the old H, since $\partial G / \partial t = 0$, G having no explicit time-dependence. If we now consider ω to be a slowly-varying function of time, such that in any particular period of oscillation $(1/\omega)(d\omega/dt) = \varepsilon << 1$, an extra term is introduced into H_1; however the transformation remains canonical and the contribution $\partial G / \partial t$ amounts to a perturbation of the original Hamiltonian, which can be handled by perturbation theory.

Supposing that in any particular period τ of oscillation $(1/\omega)(d\omega/dt) = \varepsilon$ is constant, the variation in ω over this period is $\omega = \omega_0 \exp(\varepsilon_\tau) \cong (1 + \varepsilon\tau)$ for $\varepsilon_\tau << 1$. Then it can be shown by a straightforward but rather lengthy procedure that the fractional secular change in the action variable S over the period is $\Delta S / S = 1/2(\varepsilon\tau)^2$. Thus if $(1/\omega)(d\omega/dt) = \varepsilon$ is already small in first order, the long-term changes in S are very small.

The importance of adiabatic invariance of the action variables is of great significance in many fields, some of which are discussed in detail in Section II-7 of Goldstein [1]. It is in fact the mechanism which enables a ballerina to execute an accelerating pirouette by drawing in the outstretched arms, adiabatically of course!

REFERENCES

[1] H. Goldstein, Classical Mechanics, Second Edition, Addison-Wesley (1980).

[2] J.D. Jackson, Classical Electrodynamics, Second Edition, John Wiley & Sons (1975).

[3] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman Lectures on Physics, Vol. II, ch. 19, Addison-Wesley (1965).

[4] A. Schoch, Report CERN 57-21 (1958). [An early systematic treatment of linear and non-linear resonances in the CERN PS.]

[5] B.W. Montague, Non-Linear Coupling from The Beam-Beam Force in Electron-Positron Storage Rings. Nuclear Instruments and Methods 187 (1981) 335-340, North-Holland.

NONLINEAR RESONANCES

E.J.N. Wilson
CERN, Geneva, Switzerland

Abstract
Building on classical Hamiltonian dynamics, this paper shows how a number of transformations can isolate the perturbation due to nonlinear terms in the guide field of a synchrotron. The concept of a resonance in transverse phase space is extended to include islands in phase space and the catalytic influence of synchrotron motion on beam diffusion.

1. INTRODUCTION

Hamiltonian mechanics is not everyone's cup of tea. Even experienced synchrotron designers often avoid the use of canonical transformation exercises to solve the simpler problems like the influence of sextupole magnets on betatron motion. It is true that for these problems a simple physical model can be used in which a sextupole is imagined as an element whose focusing strengths increase linearly with radius [1]. Such a tangible model related to the familiar concepts of optics is a valuable key to the door of nonlinear theory and may be used with great effect to predict resonant behaviour near a third integer resonance. One may extend this model to higher-order multipoles and into two degrees of freedom but at some point the hand waving becomes too vigorous to be credible.

Unfortunately, the topical problems of nonlinear theory applied to today's accelerator and storage rings include high-order resonances in three dimensions which can cause an inexorable beam loss over periods of hours. Typical of these problems is the influence of the beam-beam potential seen by a particle traversing an oncoming bunch in a collider. One can tune the Q values to avoid all low-order resonant conditions so that it is only the higher terms in the polynomial expansion of this potential which are important. The intuitive model is difficult to apply to polynomial terms which are typically of tenth order and it is worth taking the trouble to become adept at the Hamiltonian formalism to solve such a problem.

In the contribution which precedes this one [2], B. Montague has laid a firm foundation for the understanding of this formalism. In this contribution I shall first take his expression for the general Hamiltonian of a particle in an electromagnetic field and judiciously simplify it by approximations which are valid for large synchrotrons. This will reveal one of the advantages of the formalism, that each term in the multipole expansion of the field has a one-to-one relationship with a term in Hamiltonian.

The next procedure is to apply a number of canonical transformations which remove the lower-order linear terms in the Hamiltonian. To the uninitiated these transformations are unnerving in that they bring one further from the everyday world in which our familiar optics can be applied. However, their power lies in the fact that the nonlinearity is isolated and we are able to plot its effect as a trajectory in a phase space in which ordinary dynamics is reduced to circular trajectory. The perturbation of the nonlinearity becomes immediately visible.

To help the reader chart our course we list the canonical transformations to be applied in the form of a table. These are to be found in Ref. [2].

Table 1

Canonical Transformations

	Transformation	Purpose	Final Coordinates
1	Change of independent variable	Express as function of s not t	x, x', s
2	To action angle variables (J, ψ)	Remove variation of focusing with azimuth	J, ψ, y
3	To the coordinates of a harmonic oscillator		J_1, ψ_1, s
4	Infinitesimal point transformation	Nonlinear terms become first order perturbation	J_2, ψ_2, s
5	Change independent variable to θ	Opens the way to finding a periodic solution	J_3, ψ_3, s
6	Transformation to a rotating coordinate system	Freezes the trajectory in a stroboscopic picture of closed curves, islands and separatrices	J_4, ψ_4, s

When this series of transformations is complete we shall have travelled a long road. Many of the steps on the journey will have to be taken on trust by those not yet skilled in imagining the new canonical transformations necessary to recast variables in a simpler form. It may help to think of each transformation as an imaginative leap like solving a hard integral, a difficult step for the uninitiated but easily checked once it is identified. In the end we shall find a simple geometrical trajectory which exhibits the physical features of resonant behaviour including unstable and stable fixed points, islands of stable motion, separatrices and stochastic layers which can lead particles to ever growing amplitude and cause the diffusion which it is our aim to explain. The position and separation of these features in phase space are precisely related to the coefficients in the field expansion of the real world.

In the remainder of this report we shall explain each of these transformations in some detail and then discuss the effect of nonlinearity in creating unstable regions of phase space, on the width of resonances and on the growth of amplitude when crossing resonances. All this will be treated for motion in one degree of freedom. We shall then see how by adding a further degree of freedom, synchrotron motion, the topology of phase space is altered and sketch out how one may lead to the Chirikov limit beyond which slow but continuous beam growth occurs. A criterion for purity of magnetic field then emerges.

We leave the details of discussion of the consequences of adding a second transverse degree of freedom and a third dimension to the field shape to other authors [3] but indicate the qualitative consequences of these extensions to the one dimensional model.

In all sections of this report, we shall refer to B. Montague's contribution to the 1993 CERN Advanced Accelerator School [2] which the reader is well advised to have at hand. Another useful reference is E.D. Courant, R.D. Ruth and W.T. Weng [4].

The next section describes the link between the rigorous and general treatment of Ref. [2] and the simplified Hamiltonian which is the starting point of our discussion. The reader may choose to omit it on first reading and pass to Eq. (8).

2. THE GENERAL FORM OF THE HAMILTONIAN

It is a good idea to start from one of the fundamental expressions to be found in text books on dynamics in an electromagnetic field. We shall rather soon introduce approximations which are justified in the context of a modern synchrotron. Watch these carefully. If your particular synchrotron is a small one you may wish to review their validity.

The general Hamiltonian for a charged particle of mass, m, and charge, e, in a magnetic vector potential, A, and electric potential, ϕ, is given by Eq. (22) of Reference [2]:

$$H(q,p,t) = e\phi + c\left[(\boldsymbol{p} - e\boldsymbol{A})^2 + m_0^2c^2\right]^{1/2} \tag{1}$$

where

c is the velocity of light,

$\boldsymbol{A}$ and ϕ are functions of space and time,

$\boldsymbol{p}$ is the momentum with components, p_k, conjugate to the space coordinates, q_k

We choose the coordinates shown in Fig. 1. The transverse displacements are x and z while s is measured along the particle's trajectory. The local radius of curvature, ρ, is dependent on the local magnetic field and is therefore a function of s. Using $\kappa = 1/\rho$, we can express the Hamiltonian in these coordinates as:

$$H(q_k,p_k,t) = c\left[\left(p_x - eA_x\right)^2 + \left(p_y - eA_y\right)^2 + \left\{\frac{p_s - eA_s}{1-\kappa\ x}\right\}^2 + m_0^2c^2\right]^{1/2} \tag{2}$$

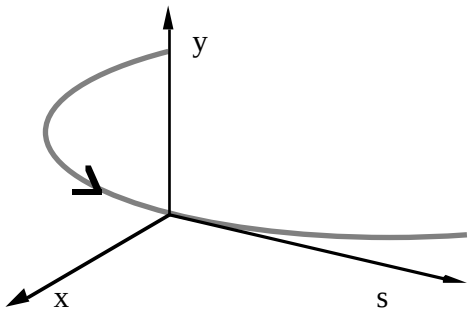


Fig. 1 Transverse coordinate system.

Here, and from now on, we shall drop ϕ which is assumed to be constant. We are only considering magnetic and not electric fields. The independent coordinate in (2) is t but it would be more convenient to use s since the machine turns out to be periodic in s. The reader may remember that since H is conjugate to t and P_S to s, one may turn the Hamiltonian (2) inside out by writing a new Hamiltonian, F, which is really $-p_S$ expressed as a function of the other canonical variables together with s, the new independent variable. We find this in Montague [2], again, in his Eq. (48) :

$$F\left(x,y,t,p_x,p_y,-H,s\right)=-p_s$$
$$=-eA_s-(1-\kappa\ x)\left[\frac{1}{c^2}\left(H^2-m_0^2c^4\right)-\left(p_x-eA_x\right)^2-\left(p_y-eA_y\right)^2\right]^{1/2} \tag{3}$$

unlike Montague we normalise by dividing all momenta including the Hamiltonian by the total momentum, P, defined in Montague's Eq. (49)

$$\frac{1}{c^2}\left(H^2-m_0^2c^4\right)=m_0^2c^2\left(\gamma^2-1\right)=m_0^2c^2\beta^2\gamma^2=\boldsymbol{P}\cdot\boldsymbol{P} \tag{4}$$

and apply the trivial transformation:

$$\bar{q}=q\ ,\quad \bar{s}=s\ ,\quad \bar{p}=\frac{p}{P}\ ,\quad \bar{F}=\frac{F}{P}\ , \tag{5}$$

This has the virtue that the new momenta, $\bar{p}_{x,z}$, are none other than the familiar transverse divergence angles of the trajectory, x',z'

$$\bar{F}=-\frac{eA_s}{P}-\left(1+\frac{x}{\rho}\right)\left\{1-\left(\bar{p}_x-\frac{eA_x}{P}\right)^2-\left(\bar{p}_z-\frac{eA_z}{P}\right)^2\right\}^{1/2}\ . \tag{6}$$

Finally, in our attempt to recast the Hamiltonian in a more tractable form, we shall ignore the vertical plane and approximate by assuming $\rho=\infty$. Furthermore, we shall assume the magnetic field is only transverse so that $A_x = A_z = 0$.

These drastic measures should not be forgotten. They are completely justified in order to study the horizontal motion in a large synchrotron comprising long magnets but would have to be re-examined in the case of a small machine in which the end fields of lattice elements are of importance. With these reservations we obtain a new Hamiltonian which we call H because we are running out of alphabet:

$$H\approx-\frac{eA_s}{P}-\left(1-\bar{p}_x^2\right)^{1/2}\ , \tag{7}$$

and if $p_x << 1$,

$$H\approx-\frac{eA_s}{p}+\frac{\bar{p}_x^2}{2}\ . \tag{8}$$

3. THE MAGNETIC VECTOR POTENTIAL FOR MULTIPOLES

It would be difficult to unravel the problem of nonlinear motion in a synchrotron if we were not able to analyse magnetic fields in to a series in which each term corresponds to a magnet with a certain number of poles.

We have already announced our intention to ignore the ends of magnets where there can be transverse components of the magnetic vector potential and restrict our analysis to the body of long magnets where only A_s is finite and there are only the transverse components of field, B_x and B_z. This has the virtue that the vector potential can be expressed by a series:

$$A_s = \sum_n A_n f_n(x,z) \tag{9}$$

$$= \sum_n \mathrm{A}_n (x+iz)^n \ . \tag{10}$$

In this expansion f_n corresponds to a multipole with $2n$ poles. The real terms form a series for "normal" multipoles for which the field is normal to a horizontal mid-plane and the imaginary terms correspond to skew multipoles.

We can derive the vertical field component for a $2n$ pole magnet:

$$B_z(z=0) = \frac{\partial A_s}{\partial x} = n \ \ A_n \ \ x^{(n-1)} \ , \tag{11}$$

and we can write this as a term in the Taylor series:

$$B_z(z=0) = \frac{1}{(n-1)!} \frac{\partial_{(n-1)} B_z}{\partial x^{(n-1)}} x^{(n-1)} \ . \tag{12}$$

We can also write the first term in H (Eq. 8) as:

$$\frac{eA_s}{p} = \frac{1}{B\rho} \sum \frac{1}{n!} \frac{\partial_{(n-1)B_z}}{\partial x_{(n-1)}} x^n \ , \tag{13}$$

and our Hamiltonian becomes:

$$H = \frac{p_x^2}{2} + \sum_{n=0}^{\infty} \frac{1}{(B\rho)} \frac{1}{n!} \frac{\partial_{(n-1)} B_z}{\partial x_{(n-1)}} x^n \ . \tag{14}$$

We immediately see that each order of multipole will contribute a term to the Hamiltonian. It is hardly a difficult step to see that if we were to allow both transverse degrees of freedom a normal multipole with $2n$ poles can contribute a set of terms x^n, $x^{n-2}z^2$, $x^{n-4}z^4$, while a skew multipole, corresponding to the imaginary terms in the expansion of Eq. (10), would introduce the missing terms $x^{n-1}z$, $x^{n-3}z^3$ ... in the homogeneous polynomial. It can be shown that each term corresponds to a line in the Q diagram, so this is helpful in identifying the kind of error which may cause resonant loss at a particular working point.

4. LINEAR DYNAMICS IN ACTION ANGLE VARIABLES

Readers who are already familiar with the theory of transverse dynamics will remember that linear motion is described locally by Hill's equation:

$$\frac{d^2x}{ds^2} + k(s)x = 0 \ . \tag{15}$$

This is a differential equation with a periodic coefficient. The focusing strength, $k(s)$, repeats every turn of the ring, or even every superperiod, if the lattice has a symmetry within a

turn. The periodic variation of $k(s)$ distinguishes the solution from simple harmonic motion giving rise to an amplitude function, $\sqrt{\beta}(s)$, that varies periodically with s reflecting the focusing pattern, $k(s)$. The motion can be converted into that of a harmonic oscillator with a simple sine like solution by a change of variables [5]:

$$\eta = \frac{x}{\sqrt{\beta}} \ . \tag{16}$$

We can see that if we take the linear Hamiltonian of Eq. (8) it has the same periodic coefficient:

$$H = \frac{p^2}{2} + \frac{k(s)x^2}{2} \ , \tag{17}$$

where we have dropped the bar in the notation and:

$$k = \frac{1}{(B\rho)} \frac{\partial B_z}{\partial x} \ . \tag{18}$$

Incidentally it is an excellent exercise for newcomers to apply Hamilton's equations to (17) and thus derive Hill's equation. (See for example Equations (52) of Ref. [2].)

Expressed in the jargon of Hamiltonian mechanics the difficulty with Eq. (17) is that H is not time-invariant. (Remember that earlier in Section 2 we used s instead of t as an independent variable). The difficulty is really the same as that with Hill's equation because if the Hamiltonian were to be time invariant it would generate the differential equation of a harmonic oscillator, and the trajectory of the particle in phase space would be the same closed ellipse independent of the observer's position, s. The trajectory could then be labelled with a numerical value, a constant of the motion. The momentum conjugate to any coordinate which is not in the expression for H is invariant, this applies equally to H itself which is conjugate to s.

Perhaps this is the first inkling for the diffident recipient of Hamilton's legacy that it might be of some practical value. He or she will be pleased to learn that there is a canonical transformation to action and angle variables which is Hamilton's solution to this difficulty.

If we were to derive Hill's equation from the Hamiltonian (17) and solve it we would find (using y as the general transverse coordinate rather than x which is specifically horizontal):

$$y = \varepsilon^{1/2} \beta^{1/2}(s) \cos\left[\varphi(s) + \delta\right] \tag{19}$$

$$\beta'' / 2 + k\beta - (1 + \alpha^2) / \beta = 0 \tag{20}$$

$$\varphi(s) = \int_0^s \frac{ds}{\beta(s)} \ , \tag{21}$$

where

y is the general transverse coordinate, and y' its derivative with s

$\beta(s)$ is an amplitude function periodic in azimuth, $\alpha(s) = -\beta'(s) / 2$

ε is an emittance defining an azimuthal invariant,

δ is an arbitrary phase.

We must find the momentum which is conjugate to y and for this we use Hamilton's Equation:

$$\frac{dy}{ds} = \frac{\partial H(y,s)}{\partial p} = p \ . \tag{22}$$

By differentiating Eq. (19) we find:

$$p = -\varepsilon^{1/2}\beta^{1/2}(s)\left\{\sin[\varphi(s)+\delta] - \frac{\beta'}{2}\cos[\varphi(s)+\delta]\right\} \ . \tag{23}$$

We are now in a position of having physical coordinates (p and y) which are conjugate according to the rules of Hamilton. Neither of them is a constant of the motion. In our desire to freeze out the linear terms it would be an advantage to transform (p and y) to new coordinates (J and ψ) in which J becomes independent of time. The next three equations recall how Montague (Section 6 of Reference [2]) achieved this for the case of a linear oscillator by using a canonical transformation from old coordinates which instead of his $G(q,Q)$ we write for our notation:

$$F_1(y,\psi) = \frac{\sqrt{k\ \ m}}{2} y^2 \cot\ \psi \tag{24}$$

into canonical coordinates (J,ψ) in which the Hamiltonian became cyclic in $\tilde{}\psi$

$$H(J,\psi) = \omega\ \ J. \tag{25}$$

from which we conclude that J is a constant of the motion and

$$\dot{\psi} = \frac{\partial H}{\partial J} = \omega \tag{26}$$

is simply the angular frequency of ψ. These new coordinates, (J,ψ), are called *action and angle.*

The search for an appropriate generating function to transform the transverse motion in a synchrotron into a simple action and angle requires considerable ingenuity and the student is best advised to simply accept that the generating function has a form bearing a certain similarity to Eq. (25):

$$F_1(y,\psi,s) = -\frac{y^2}{2\beta}\left[\tan\psi - \frac{\beta'}{2}\right] . \tag{27}$$

We remember that β is a function of s and note that, for purely historical reasons this author's generating functions are given the symbol F and not G as in Montague.

The rules for applying this function are:

$$p = \frac{\partial F_1}{\partial y} \ , \quad J = -\frac{\partial F_1}{\partial \psi} \ , \tag{28}$$

and the new Hamiltonian

$$K = H + \frac{\partial F_1}{\partial s} \ . \tag{29}$$

Applying these rules and remembering that $y' = p$ we find:

$$y' = -\frac{y}{\beta}\left(\tan\ \psi - \frac{\beta'}{2}\right) \tag{30}$$

$$\tan^2\psi = \left[y' - \left(\frac{\beta' y}{2\beta}\right)\frac{\beta}{y}\right]^2 \tag{31}$$

$$J = \frac{y^2}{2\beta}\sec^2\psi = \frac{y^2}{2\beta}\left[1 + \tan^2\psi\right] \tag{32}$$

$$= \frac{y^2}{2\beta}\left[1 + \left\{\left(y' - \frac{\beta' y}{2\beta}\right)\frac{\beta}{y}\right\}^2\right]^2 \tag{33}$$

which yields the invariant:

$$2J = \frac{1}{\beta}\left[y^2 + \left(\beta y' - \frac{\beta' y}{2}\right)^2\right] . \tag{34}$$

This will be recognized by anyone familiar with Courant and Snyder [5] as an invariant of particle motion around a synchrotron. Even though β is a function of s, J is constant. If J is the invariant for the largest amplitude particle in a beam $2J$ is just the emittance, ε, which includes the beam. $\pi\varepsilon$ or $2\pi J$ is the area of the phase space ellipse anywhere in the ring.

If we were to go through the full transformation procedure (see Ref. [2] again) we would find that the new Hamiltonian is:

$$K = J / \beta(s) \ , \tag{35}$$

Hamilton's Eqs. (17) and (18) of Ref. [2] gives:

$$\frac{d\psi}{ds} = \frac{\partial K}{\partial J} = \frac{1}{\beta(s)} \ . \tag{36}$$

ψ is none other than the familiar betatron phase advance and we are not surprised to find:

$$y = \sqrt{2J\beta}\,\cos\psi \tag{37}$$

$$y' = -\sqrt{2J\beta}\left[\sin\psi - (\beta'/2)\cos\psi\right] . \tag{38}$$

We have arrived at an invariant momentum, J, conjugate to ψ but the Hamiltonian K still varies like $1/\beta(s)$ around the ring. To finish the job of freezing out the linear motion, we need to transform into a new coordinate system (J_1, ψ_1). (Note we use suffix 1 to indicate new coordinates. We shall continue the policy of adding one to this suffix each time we transform). The new and old coordinates are related by a generating function of the second kind. This kind of generating function includes rotations in space and does not mix momenta and coordinates as does F_1:

$$F_2(\psi, J_1) = J_1\left[\frac{2\pi Qs}{C} - \int_0^s \frac{ds'}{\beta}\right] - \psi J_1 \tag{39}$$

$$\psi_1 = \frac{\partial F_2}{\partial J_1} = \psi + \frac{2\pi Qs}{C} - \int_0^s \frac{ds'}{\beta(s')} \tag{40}$$

$$J = \frac{\partial F_2}{\partial \psi} = J_1 \tag{41}$$

where

$$2\pi Q = \int_0^{2\pi R} \frac{ds}{\beta} \quad \text{and} \quad C = 2\pi R . \tag{42}$$

The new Hamiltonian is related to the old one of Eq. (35) by:

$$H_1 = K + \frac{\partial F_2}{\partial s} = \frac{2\pi Q}{2\pi R}\cdot J_1 = \frac{Q}{R}\cdot J = \text{constant} . \tag{43}$$

We have now reached our aim insofar as the linear dynamics is concerned. In the new coordinates the trajectory of the particle is a circle of radius J_1 which is numerically equal to half the emittance of the beam it includes. The variable ψ_1 goes from 0 to 2π in one betatron oscillation. The Hamiltonian is independent of s (or time). In the next section we shall see how this leaves the way free to express nonlinear motion as a perturbation of this circle. The circle will be distorted, its circumference will develop waves and an archipelago of islands will form.

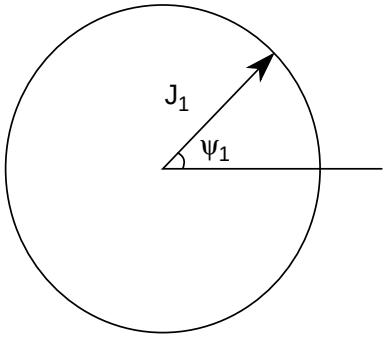


Fig 2 Phase space trajectory of a linear system in action angle coordinates

5. PERTURBATION THEORY

With the linear problem behind us we are ready to use perturbation theory to treat the nonlinear effects. The terms in the Hamiltonian (Eq. (14)) with a cubic or higher transverse dependence are nonlinear. They arise from sextupole and higher multipole fields distributed in azimuth, *s*, around the focusing lattice. Unlike the linear terms they retain their *s* dependence even when we apply the canonical transformation to action-angle variables, J_1 and ψ_1. In order to reveal the way they distort the perfect circle of Fig. 2 we must first apply a transformation which removes the *s* dependence of the Hamiltonian so that *H* becomes a new invariant of the motion. The transformation which does this will tell us how much the circle is distorted.

The prescription for finding the canonical transformation which irons out the *s* dependence follows. At the centre of canonical perturbation theory is a generating function of the second type which is an infinitesimal perturbation of the function which generates the identity transformation:

$$F_2 = (J_2, \psi_1, s) = J_2\psi_1 + \chi(J_2, \psi_1, s) \ , \tag{44}$$

where χ is small.

The new and old coordinates are simply related by the first derivatives of this small quantity, χ

$$J_1 = J_2 + \frac{\partial}{\partial \psi_1}\chi(\psi_1, J_2, s) \tag{45}$$

$$\psi_2 = \psi_1 + \frac{\partial}{\partial J_2}\chi(\psi_1, J_2, s) \tag{46}$$

$$H_2 = H_1 + \frac{\partial}{\partial s}\chi(\psi_1, J_2, s). \tag{47}$$

The first term in Eq. (47), H_1, will contain the unperturbed Hamiltonian (Eq. (17)) plus a small nonlinear perturbation due to one or more of the cubic or higher terms in Eq. (14). Of course, we must transform this nonlinear term into (J_1 ψ_1) coordinates using substitutions from Eqs. (35), (37), (38), (40), (41) and (43). The nonlinear terms will give it an *s* dependence.

Our plan of campaign is to choose the function χ in such a way that its differential, $\partial\chi / \partial s$, exactly cancels out the *s*-dependence making H_2 a constant of the motion.

But how do we choose the exact form of χ? At the moment it is arbitrary. We can first rewrite Eq. (47) to split H_1 into the linear part of the Hamiltonian, H_0, and the perturbation *U*. Both are functions of J_1 but we apply a trick and will write them as functions of $J_2 + \partial\chi / \partial\psi_1$ rather than J_1. This is to make it easier to identify first and second-order terms when we expand the terms:

$$H_2 = H_0\left(J_2 + \frac{\partial\chi}{\partial\psi_1}\right) + U\left(\psi_1, J_2 + \frac{\partial\chi}{\partial\psi_1}, s\right) + \frac{\partial\chi}{ds} \ . \tag{48}$$

We next must eliminate second-order terms and to do this we shall use another trick and introduce three new pairs of terms which cancel. This will help us group all the terms which are either of second order in *U* or are independent of *J*, ψ or *s* in rectangular brackets:

$$H_2 = H_0(J_2) + \left[H_0\left(J_2 + \frac{\partial\chi}{\partial\psi_1}\right) - H_0(J_2) - \frac{Q}{R}(J_2)\frac{\partial\chi}{\partial\psi_1}\right]$$

$$+\left[U\left(\psi_1, J_2 + \frac{\partial\chi}{\partial\psi_1}, s\right) - U(\psi_1, J_2, s)\right] + \frac{Q}{R}(J_2)\frac{\partial\chi}{\partial\psi_1} + \frac{\partial\chi}{\partial s} + U(\psi_1, J_2, s) \ . \tag{49}$$

The quantity *Q* is the familiar betatron wavenumber. If we examine the contents of the rectangular brackets very carefully and remember that (Eq. (43)):

$$\frac{Q}{R}(J_2) = \frac{\partial H_0}{\partial J_2} \tag{50}$$

we find that the brackets reduce to:

$$\frac{1}{2R}\frac{\partial Q}{\partial J_2}\left\{\frac{\partial\chi}{\partial\psi_1}\right\}^2 + \frac{\partial U}{\partial J_2}\frac{\partial\chi}{\partial\psi_1} \ , \tag{51}$$

which are purely second order and can be left out to give:

$$H_2 = H_0(J_2) + \frac{Q}{R}(J_2)\frac{\partial\chi}{\partial\psi_1} + \frac{\partial\chi}{\partial s} + U(\psi_1, J_2, s) \ . \tag{52}$$

At this point we shall digress to attend to a small detail. The independent variable is still *s* but we shall find it much more convenient as we move into Fourier analysis to use the azimuthal angle $\theta = s/R$. In Ref. [2] it was shown that one may scale a pair of conjugate variables. To preserve canonical variables we must multiply H_2 the conjugate to *s* by *R*:

$$H_3 = RH_2 \ , \quad \theta = s / R \tag{53}$$

$$\psi_3 = \psi_2 \tag{54}$$

$$J_3 = J_2 \ . \tag{55}$$

This gives the Hamiltonian of Eq. (53) as a function of θ:

$$H_3 = (\psi_3, J_3, \theta) = RH_0(J_3) + Q(J_3)\frac{\partial \chi}{\partial \psi_1} + R\frac{\partial \chi}{\partial s} + RU(\psi_1, J_3, \theta) \ . \tag{56}$$

To return to our aim of removing the *s* dependence in the last three terms of Eq. (56) we can simply impose the condition:

$$Q(J_3)\frac{\partial \chi}{\partial \psi_1} + \frac{\partial \chi}{\partial \theta} + RU(\psi_1, J_3, \theta) \ = 0 \ , \tag{57}$$

and $H_3 = RH_0(J_2)$ will be independent of *s*. We recall the suffix zero indicates the linear term only.

If we know the driving perturbation $U(\psi_1, J_3, \theta)$ we can solve the differential Eq. (57) to find χ.

To find this solution it is convenient, and physically revealing, to analyse U into a Fourier expansion, either a single series in ψ, or a double series in ψ and θ. The single series method is simpler to start with and is appropriate if one is far from a particular resonance condition. One assumes the function χ can be also expanded as a Fourier series and then one solves the differential equation for χ term by term, rather in the manner of solving electrical circuit problems.

6. EFFECT OF NONLINEARITIES FAR FROM A RESONANCE

To first flex our muscles in perturbation theory we solve the problem far from resonance. By far from resonance we mean that the quantity, *mQ-n*, is large compared with the width of the nearest resonance. We express the perturbation as a single Fourier series in ψ, the "betatron" variable. The index of the summation is the multipole number, 3 for sextupole, 4 for octupole, etc.

$$U = \sum_{n=0}^{\infty} U_n(J_3, \theta) e^{in\psi_1} \ . \tag{58}$$

The solution to (57), χ, will also be a Fourier series:

$$\chi = \sum_{n=0}^{\infty} \chi_n(J_3, \theta) e^{in\psi_1} \ . \tag{59}$$

and by substitution in (57) we find:

$$\left[inQ(J_3) + \frac{\partial}{\partial \theta} \right] x_n = -RU_n \ . \tag{60}$$

This has the solution:

$$\chi_n = \frac{i}{2 \sin \pi n Q} \int_{\theta}^{\theta+2\pi} e^{inQ(\theta'-\theta-\pi)} U_n(\theta') \, d\theta' \ . \tag{61}$$

The function χ formed by summation (Eq. (59)) will then remove the s dependence of the Hamiltonian and make the new Hamiltonian:

$$H_3 = H_0(J_3) + \sum_n R U_n(J_3, \theta) e^{in\psi_1} \ . \tag{62}$$

a constant of the motion.

How do we use this function χ which magically renders the Hamiltonian invariant in order to deduce the distortions it makes to the simple linear circle, $J_1(\psi_1)$? We can show that not only is the new H_3 independent of B but that J_3 is invariant to first order. Equation (50) reminds us that the rate of change of ψ_1 is just Q/R, a constant. It can be shown that the first-order effect of the perturbation is simply to add a small constant term to Q so that, to first order, $J_3 = J_2$ is an invariant. We must now look carefully at Equations (45) to (47). The new J is invariant and so the old J_1 is defined as a function of azimuth by subtracting the derivative of χ from a constant J_2. If we know χ from computation of Eq. (61) we can find this derivative as a function of χ and trace out the perturbed $J_1(\psi_1)$ circle.

7. RESONANCES

Close to a resonance we must expect to have to take into account particular harmonics of the error distribution which act coherently with betatron motion. As a first step we assume that the perturbation is periodic in both ψ and θ and expand as a double Fourier series:

$$U = \sum U_{nm}(J_2) e^{i(n\psi_2 - m\theta)} \tag{63}$$

in which the coefficients are:

$$U_{nm} = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_{\psi_2}^{2\pi+\psi_2} U(\psi, J_2, \theta) e^{i(n\psi - m\theta)} \, d\theta d\psi \ . \tag{64}$$

The treatment follows a parallel path to that of Section 6 and we arrive at the expression for χ and H:

$$\chi = i \sum_{m,n} \frac{R U_{nm}}{nQ - m} (J_2) e^{i(n\psi_2 - m\theta)} \tag{65}$$

$$H_3 = H_0(J_3) + R U_{nm} e^{i(n\psi - m\theta)} \ . \tag{66}$$

Again we may use χ to compute the nonlinear distortion of the circular trajectory but close to the resonant condition $nQ = m$ the dominator becomes small. Therefore, we begin to suspect

the validity of our first-order approximation and in particular our argument that H_3 was essentially independent of explicit J_3 dependence.

Fortunately, if we zoom in on the one resonant term (which we are justified in doing when we are close to $mQ = n$) there is a transformation we can apply which gives an H_4 which is the exact invariant of the motion and which yields frozen contours in phase space. We no longer need to apply Equations (45) to (47) assuming J_3 = constant since the equation of H_4 itself defines a contour or trajectory in phase space. This alternative technique of finding the trajectory only works if we drop all the Fourier terms except the resonant combination, n,m.

Starting from Eq. (66) we apply an F_2 type of generating function

$$F_2\left(\psi_3, J_4, \theta\right) = \left(\psi_3 - \frac{m}{n}\theta J_4\right) \tag{67}$$

to find

$$J_3 = J_4 \tag{68}$$

$$\psi_4 = \left(\psi_3 - \frac{m}{n}\right)\theta \tag{69}$$

$$H_4 = \left(Q - \frac{m}{n}\right)J_4 + RU_{00}\left(J_4\right) + RU_{mn}\ e^{in\psi_4}\ . \tag{70}$$

The new angle coordinate means that our reference frame rotates so that ψ_4 remains unchanged while θ advances n turns and at the same time the betatron phase, ψ_3, advances m oscillations. This has the effect of "freezing" the resonance in our new phase space. The result is a Hamiltonian which is independent of time and which defines a closed contour in phase space. We shall use this to explore the examples of third and fourth-order resonance before generalizing the theory and moving on to explain invariants and islands in phase space.

We have finished the bulk of the Hamiltonian formalism so a little recapitulation is perhaps in order. There are two approaches to finding out how nonlinear terms distort the simple circular trajectory due to linear focusing in $\left(J_1, \psi_1\right)$ space. The first, outlined in Section 5 uses first-order perturbation theory to find a generating function, χ, assumed to be a single Fourier series related to the azimuthal pattern of the perturbation. Once found, the generating function tells us the distortion of the circle provided it is small. The second approach is to use a double Fourier expansion and then change the coordinate so that it keeps pace with the resonant oscillations at a stopband: $m\psi = n\theta$. This gives an exact invariant Hamiltonian which itself defines the path in phase space but ignores all but the resonant perturbation.

8. THE THIRD-INTEGER RESONANCE

One of the confusing aspects of a generalized description involving a number of transformations is that we tend to lose track of the numerical relation between the final coordinates and the initial physical system. In this example we try to give the reader the link between these coordinate frames.

Let us now take a practical example of a sextupole-driven resonance. We suppose that the Hamiltonian expressed in "normal"" coordinates includes a sextupole field:

$$H = \frac{p^2}{p} + \frac{k(s)}{2}x^2 + \frac{1}{3!(B\rho)}\frac{\partial^2 B_z}{\partial x^2}x^3 \ . \tag{71}$$

First we must transform into action angle coordinates (J,ψ) applying Eqs. (35), (37) and (38) and then, to arrive in a coordinate system in which the Hamiltonian of a linear system would be a constant of the motion, we apply Eqs. (40), (41) and (43). The final coordinates are (J_1, ψ_1) and the Hamiltonian, H_1:

$$H_1 = \frac{Q}{R}J_1 + \frac{[2J_1\beta(s)]^{3/2}}{3!(B\rho)}\frac{\partial^2 B_z}{\partial x^2}\ \cos^3\psi_1 \ . \tag{72}$$

The new angle variable, ψ_1 is defined:

$$\psi_1 = \psi - \left[\int_0^s \frac{ds'}{\beta(s)} - \frac{Q}{R}s\right] \tag{73}$$

where

$$\psi = \int_0^s \frac{ds}{\beta(s)} \ . \tag{74}$$

We recall that the old coordinates are x and x' related to J and ψ via:

$$x = \sqrt{2J\beta(s)}\cos\ \psi$$
$$p = x' = -\sqrt{2J\beta(s)}\sin\ \psi \ . \tag{75}$$

We can now make use of the purely trigonometrical relation:

$$\cos^3\psi_1 = \frac{1}{2^{(3-1)}}(\cos\ 3\psi_1 + 3\ \cos\psi_1) \ .$$

Ignoring the second term which in general does not drive a one-third-integer resonance:

$$H_1 = \frac{Q}{R}J_1 + \left[\frac{2^{2/3}}{2^{3-1}}\cdot\frac{J^{3/2}\beta(s)^{3/2}}{3!(B\rho)}\cdot\frac{\partial^2 B_z(s)}{\partial x^2}\right]\cos 3\psi_1 \ . \tag{77}$$

The second term in this equation is the perturbation, U, in Eqs. (52) and (57) and, applying directly the result of perturbation theory at a resonance (Eq. (70)) we have:

$$H_4 = \left(Q - \frac{m}{3}\right)J_4 + RU_{3m}\cos 3\psi_4 \ . \tag{78}$$

The U_{OO} term in Eq. (71) is zero for m = odd multipoles. Here U_{3m} is the double Fourier coefficient (with respect to ψ and θ) of $U(\psi,\theta)$ and written explicitly is:

$$U_{3m} = \frac{J^{3/2}2^{3/2}}{3!(B\rho)2^{3-1}}\left\{\frac{1}{2\pi R}\int_0^{2\pi R}\beta^{3/2}B''e^{i\int_0^s 3[(1/\beta)-(Q/R)]ds'}e^{-ims/R}ds\right\}. \qquad (79)$$

In addition ψ_4 is given by Eq. (69).

The integral exponent merely takes care of the relative phase slip of betatron motion with respect to $Q\theta$ which occurs between quadrupoles of a regular lattice. In many cases it may be ignored.

Note too that the content of the curly brackets is none other than Guignard's [6] d_p. We have left the powers of 2 and factorials in their crude state from which one may generalise for any multipole with 2*n* poles:

$$U_{nm} = \frac{J_1^{n/2}}{n!(B\rho)2^{n-1}}\left\{\frac{1}{2\pi R}\int_0^{2\pi R}\beta^{n/2}B^{(n-1)}e^{i\int_0^s n[(1/\beta)-(Q/R)]ds'}e^{-ims/R}ds\right\}. \qquad (80)$$

Having computed U_{nm} numerically, we can substitute back in Eq. (71) to find the Hamiltonian for any desired order of resonance. It is worth remembering that the expression for U_{nm} becomes much simplified when one does the Fourier analysis for a single short sextupole or a random distribution.

Guignard [6] gives expressions for d_p, the curly brackets, for this latter case.

9. THE TRAJECTORY OF A THIRD-INTEGER RESONANCE

All this has become very mathematical and the reader may be forgiven for a little impatience to see some phase plots showing the perturbed trajectories. In order to find the shape of these phase plots we must take a hard look at Eq. (78). We shall drop the suffixes from the notation and examine the contour defined by the Hamiltonian of Eq. (78)

$$H_4 = \left(Q - \frac{m}{3}\right)J_4 + RU_{3m}\cos 3\psi_4 \ . \qquad (81)$$

Rewriting it to show the J dependence and defining constants $\delta = Q - m/3$ and $\varepsilon = RU_{3m} / J^{3/2}$ to obtain a streamlined:

$$H = \delta J + \varepsilon J^{3/2}\cos 3\psi \qquad (82)$$

If the coefficient of the second term is small, either because the amplitude is small or the driving strength is weak, the contour will be close to a circle (J = constant). The same will be true if the Q is distant from a third integer making δ, the coefficient of the first term, large.

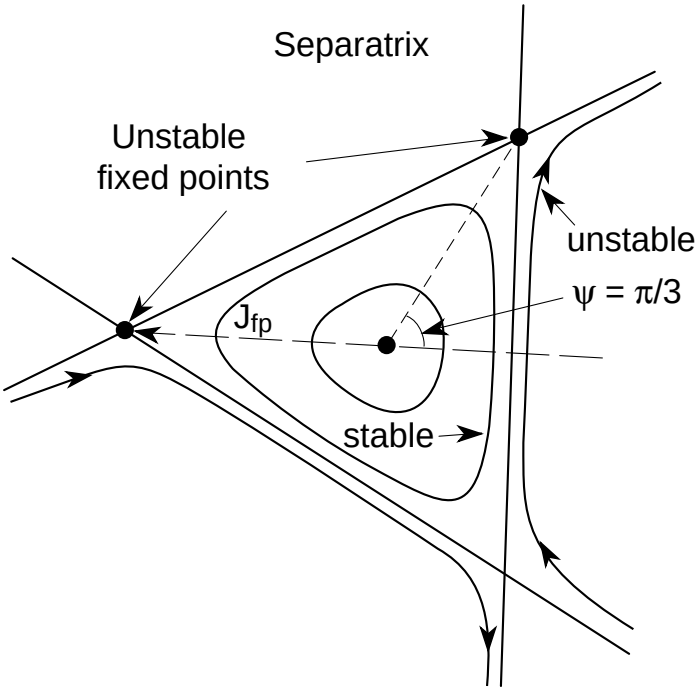


Fig. 3 Phase space plot in (J, ψ) coordinates of a third-integer resonance (adapted from Reference [3]).

If the converse is true the circle will be distorted inwards when $\psi_4 = 0, 2\pi/3, 4\pi/3$ and outwards inbetween to become somewhat triangular. We shall drop the suffix 4 from now on. Note that this triangle does not rotate but is frozen in (J, ψ) space – a consequence of the change to a rotating coordinate system. Figure 3 shows constant *H* contours in (J, ψ) space. We can see this kind of distortion growing with amplitude.

We remember from Hamilton's Equations that:

$$\frac{dJ}{ds} = -\frac{\partial H}{\partial \psi} = 3\varepsilon J^{3/2} \sin 3\psi \tag{83}$$

$$\frac{\mathrm{d}\psi}{\mathrm{ds}} = \frac{\partial H}{dJ} = \delta + \frac{3}{2}\varepsilon J^{1/2} \cos 3\psi \ . \tag{84}$$

When both these expressions are zero, a particle will stagnate in phase space lying on a "fixed point". For this to be the case Eq. (83) demands:

$$\sin 3\psi = 0 \ , \tag{85}$$

so that $dJ/ds = 0$. Then we can write $\cos 3\psi = \pm 1$ in Eq. (84) and ask how this can be zero. Above resonance, the sign of δ is positive and provided ε is positive $d\psi/ds$ can only be zero if:

$$\cos 3\psi = -1 \ . \tag{86}$$

These conditions on ψ define three fixed points at

$$\psi = \pi/3\ , \quad 3\pi/3 \quad \text{and} \quad 5\pi/3\ . \tag{87}$$

From Eq. (84) we can also find their amplitude:

$$J_{fp} = \left(2\delta/3\varepsilon\right)^2\ . \tag{88}$$

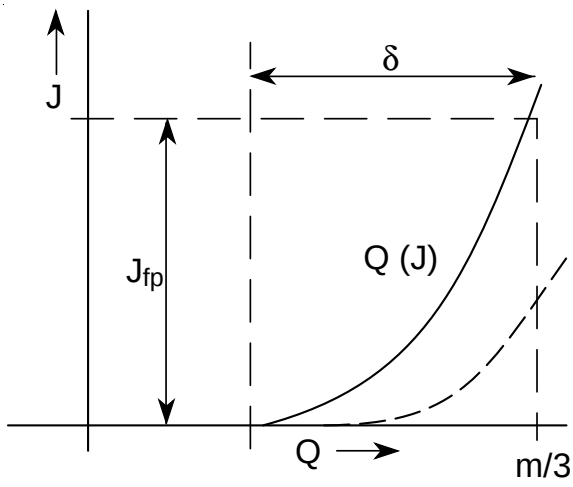


Fig. 4 The variation of Q as a function of amplitude close to a third-integer resonance

Inside this amplitude the triangular trajectory becomes less and less distorted and more circular as one approaches $J = 0$. Outside, the trajectories diverge in unbounded motion towards infinite amplitude.

The value of Q in Eq. (81) is the unperturbed Q for zero amplitude particles which corresponds to the centre of the diagram. The betatron wave number increases with J and becomes exactly one-third integer at the fixed points. This difference in Q

$$\delta = Q - (m/3) \tag{89}$$

is just the stopband width for a particle of amplitude J. Any particle with an unperturbed Q closer to the third integer than this will be already unstable if its amplitude is J. Note that if we tune the Q closer to $m/3$ or increase the strength of ε the stable triangle where motion is bounded will shrink and expel particles. This is the principle of third-integer ejection.

10. THE EFFECT OF AN OCTUPOLE

This section is not merely a repetition of the previous one for a different multipole. An octupole leads to an extra term in the invariant Hamiltonian which corresponds to a nonlinear variation of frequency with amplitude. This is characteristic of even m multipoles and defines quite a different topology of phase space.

First we must write down the Hamiltonian in "physical" coordinates.We have done this for the horizontal plane where, to be precise for small radius machines, the term $1/\rho$ should be included.

$$H = \frac{p_x^{\,2}}{2} + \frac{1}{2}\left(\frac{1}{\rho} - k(s)\right) - k(s)x^2 + \frac{1}{4!(B\rho)}\frac{\partial^3 B_z}{\partial x^3}x^4\ . \tag{90}$$

where $k(s)$ is defined in Eq. (18).

After transforming to action-angle coordinates:

$$H_1 = \frac{Q_0}{R} J_1 + \frac{(2J\beta)^{4/2}}{4!(B\rho)} B^{(3)} \cos^4 \psi_1 \ . \tag{91}$$

where

$$B^{(3)} = \left. \frac{\partial^3 B_z}{\partial x^3} \right|_{\substack{x=0 \\ z=0}} \ . \tag{92}$$

Again we expand

$$\cos^4 \psi_1 = \frac{1}{2^4} \left(\cos 4\psi_1 + 4 \cos 2\psi_1 + \frac{4!}{[(4/2)!]^2} \right) . \tag{93}$$

The first term in Eq. (93) will resonate at quarter-integer Q values, the second at half-integer values and the third adds a U_{00} term to the Fourier analysis of the perturbation. It is this that gives the amplitude variation of Q.

We ignore the half-integer term for the purpose of this discussion, but we are interested in the U_{00} term. This was missing in the case of the third- (and all odd-) order resonances. We skip the transformations which take us to (J_4, ψ_4) space and are reminded by Equation (71) that the general form of the Hamiltonian for a resonance

$$H_4 = \left(Q - \frac{m}{n} \right) J_4 + R U_{00}(J_4) + R U_{nm} \cos n\psi_4 \ , \tag{94}$$

may be differentiated to find an average shift in tune:

$$\frac{d\psi_4}{d\theta} = \frac{\partial H_4}{\partial J_4} = \left(Q - \frac{m}{n} \right) + \frac{R \partial U_{00}'(J_4)}{\partial J_4} \ . \tag{95}$$

For the fourth-integer case we find ΔQ_{NL}, the perturbed second term, is proportional to J:

$$\Delta Q_{NL} = \frac{4! R}{2^4 [(\psi_4 / 2)!]^2} \cdot \frac{J_4 \beta}{3! B\rho} B^{(3)} \tag{96}$$

and for other even orders of n we can find this nonlinear time shift Q_{NL} by differentiating:

$$U_{00} = \frac{J_4{}^{n/2} n!}{2^n [(n/2)!]^2} \cdot \frac{1}{\pi R} \int_0^{2\pi R} \beta^{n/2} B^{(n-1)} ds \ . \tag{97}$$

11. PHASE-SPACE TOPOLOGY WITH AMPLITUDE FREQUENCY VARIATION

The trajectories discussed for the third-integer case had no term equivalent to $RU_{OO}(J)$ in Eq. (14). Let us explore the topology of phase space trajectories with a generalised invariant Hamiltonian of the same form as Eq. (82) but including the nonlinear shift $\alpha(J)$ – note that from now on we have dropped the suffix 4 :

$$H = \delta J + \alpha(J) + \varepsilon J^{n/2}\cos m\psi \ . \tag{98}$$

We remember that the perturbed Q is:

$$Q = \frac{\partial H}{\partial J} = \delta + \alpha'(J) + \frac{n}{2}\varepsilon J^{(n/2-1)}\cos m\psi \ . \tag{99}$$

The first two terms cancel on average when J has a resonant value J_r defined by the condition:

$$\alpha'(J_r) = -\delta \ . \tag{100}$$

If we look at small changes in J about J_r by a second differentiation we find:

$$(J - J_r) = -\frac{n}{2}\frac{\varepsilon}{\alpha''} J_r^{(n/2-1)}\cos m\psi \ . \tag{101}$$

We now have stable and unstable fixed points at cos $m\psi$ = -1 and +1 respectively and separated in J by:

$$\frac{n\varepsilon}{\alpha''} J_r^{(n/2-1)} \ . \tag{102}$$

The existence of stable fixed points in the real plane is new and is only the case for even-order resonances. We find that they are the centres of an archipelago of stable islands. Trajectories circulate around these points within the islands. The unstable fixed points are at the junction of the islands and the diagram is identical to a series of RF buckets plotted in polar coordinates with a harmonic number equal to m.

The width of these islands may be calculated and is found to be:

$$\Delta J_W = 2\sqrt{\frac{\varepsilon(J_u)^{n/2}}{\alpha''}} \ . \tag{103}$$

where J_u is that of an unstable fixed point.

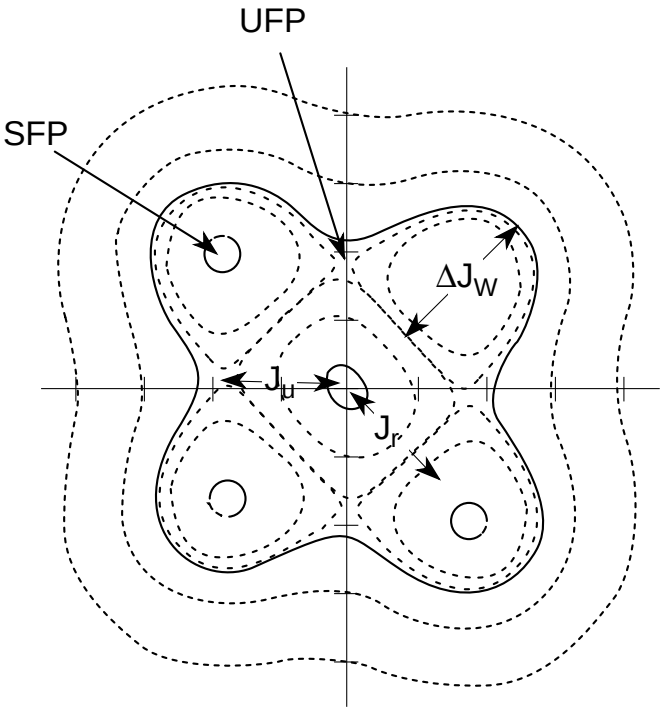


Fig. 5 The effect of an octupole near a fourth-order resonance (adapted from Reference [7]).

Motion will remain contained within the islands unless there is some way in which particles may "leak" out. One mechanism which may cause leakage is the presence of another archipelago which overlaps. This is said to cause chaotic behaviour. The next set of islands of comparable order will correspond to a change of Q of the order of $1/n$ where n is the order of the resonance. Since α'' controls $\partial Q / \partial J$ the spacing in phase space is:

$$\Delta J_S = \frac{1}{\alpha''} \cdot \frac{1}{n} \ . \tag{104}$$

We can expect these archipelagos to overlap and chaos to ensue when:

$$\Delta J_W = \Delta J_S \tag{105}$$

$$4\varepsilon \mathrm{J}^{\mathrm{n}} \geq \frac{1}{n^2} \cdot \frac{1}{\alpha''} \ . \tag{106}$$

This is the Chirikov limit and normally does not happen in any machine built with reasonable care. Unfortunately, even in a clean machine, synchrotron motion or magnet ripple can produce sidebands to the main resonance and increases the number of resonance lines each of which produces its own set of islands. The chaotic condition is much easier to reach as phase space becomes much more crowded.

12. AMPLITUDE GROWTH ON CROSSING A RESONANCE

So far we have described the dynamics of a machine in which neither the magnetic field nor the momentum of the particles change with time. Many of the less welcome effects of nonlinear fields only become apparent when either the magnetic field changes because of ripple in the power supply or when the particle's momentum changes sinusoidally as it executes synchrotron oscillations in an RF bucket. It is a relatively common experience to find that an

injected beam survives until the accelerating cavities are switched on and then is extracted by a nonlinear resonance. The explanation usually given is that oscillations in radial position together with the chromatic properties of the lattice cause the particle to cross and recross the resonance.

One way to see what happens in the time domain when a crossing occurs is derived by Guignard [6]. We will not attempt to give his derivation here but only his result. Suppose we cross a resonance (order, m) at a speed of ΔQ_t per turn. The equation of growth is:

$$\int_0^x \frac{dx}{x^{n-1}} = \Delta Q_e \frac{2\pi}{\left(n|\Delta Q_t|\right)^{1/2}} \ , \tag{107}$$

where:

n = order of the resonance,
ΔQ_e = half width of the resonance.

On integration this gives (for one crossing):

$$\left(1 - \frac{x_0}{x}\right)^{n-2} = \frac{2\pi Q_e}{\left(n|\Delta Q_t|\right)^{1/2}} \ . \tag{108}$$

One may go on and calculate what happens if the time is modulated due to synchrotron oscillations:

$$Q = Q_0 + \hat{Q} \ \sin\left(Q_s \Omega t\right) \ , \tag{109}$$

where Ω is the revolution frequency.

However, it is more direct at this point to move into the frequency domain and analyse synchrobetatron resonances.

13. SYNCHROTRON RESONANCES

It is a fact well known to electron accelerator physicists that each nonlinear resonance has a series of satellite lines, parallel to it and spaced by Q_S/n on either side of the resonance, where Q_S is the synchrotron wave number or number of synchrotron oscillations per turn.

If we return to the Hamiltonian for a nonlinear resonance we find the perturbation term is always proportional to:

$$\cos\left(\mathrm{n}\psi - \mathrm{m}\theta\right) \ , \tag{110}$$

where:

$$d\psi / d\theta = Q \ ,$$
$$\psi = \int Q d\theta \ .$$

Now if Q is modulated by synchrotron motion:

$$Q = Q_0 + \hat{Q}\ \sin Q_s\theta \tag{111}$$

then

$$\psi = \psi_0 - \frac{\hat{Q}}{Q_s}\cos Q_s\theta\ , \tag{112}$$

and the perturbation term in the Hamiltonian will change:

$$\cos m\psi \rightarrow \cos\left[\mathrm{n}\psi_0 - \left(\frac{n\hat{Q}}{Q_s}\right)\cos Q_s\theta\right] = \sum_K J_K\left(\frac{n\hat{Q}}{Q_s}\right)\cos\left[\mathrm{n}\psi_0 + kQ_s\theta\right]\ . \tag{113}$$

Each value of k corresponds to a different satellite resonance line. The resonant condition becomes:

$$Q = \frac{m}{n} + \frac{kQ_s}{n}\ , \quad k = 0 \pm 1,\ 2,\ 3,\ \ldots \tag{114}$$

The strength of each line is determined by:

$$J_k\left(n\hat{Q}/Q_s\right)\ . \tag{115}$$

Figure 6 shows how the strength will vary with the order $k = n\Delta Q/Q_S$ where k is the distance from the resonance in units of Q_S/n and with the depth of modulation $n\hat{Q}_s / Q_s$ expressed in the same units. In general the effect will be worst when $\Delta\hat{Q} = Q_s\ ,\ 2Q_s\ ,\ 3Q_s$ etc. and will drop to 30% when either ΔQ or $\Delta\hat{Q}$ becomes larger than 10 Q_S/n. This is nevertheless a rather considerable extension of the influence of a single nonlinear resonance. In the worst case, when the depth of modulation is about 10 Q_S. The single line is split into about ten other lines of comparable strength.

Each satellite line generates its own set of islands in phase space, so that spacing can be reduced by an order of magnitude. Each resonance is narrower by $J_k\left(n\hat{Q}/Q_s\right)$ but as Fig. 6 shows this is a factor of the order of 1/3. The consequence is that the Chirikov threshold is reduced by a factor of 3 or so.

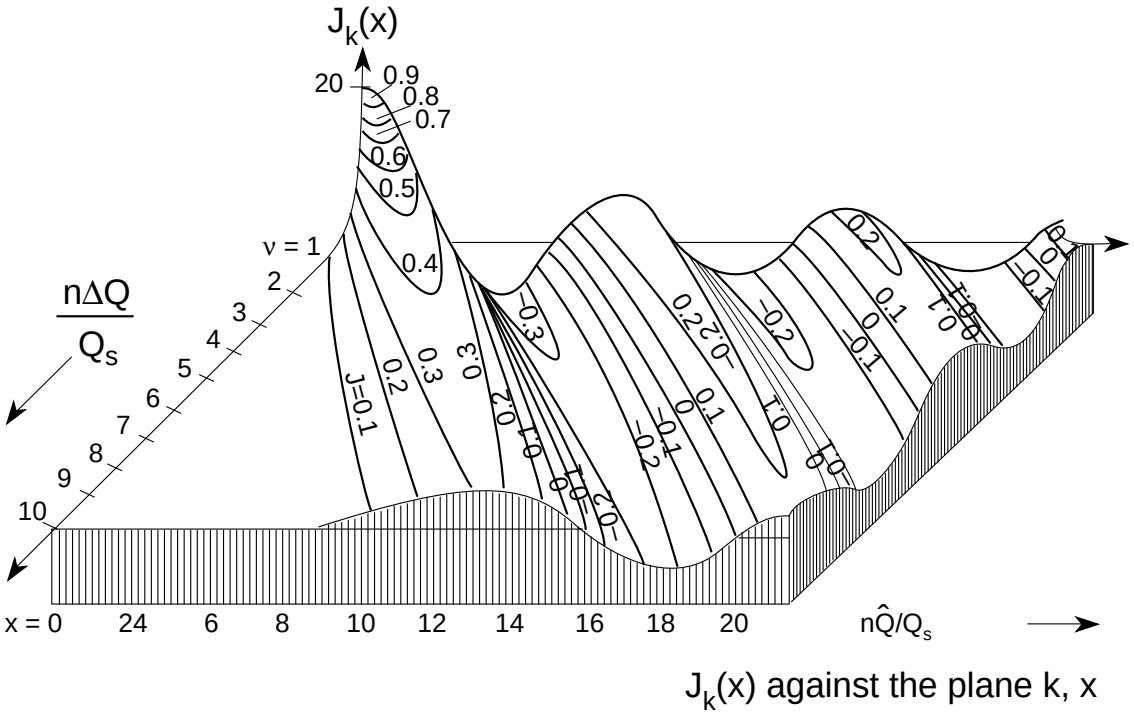


Fig. 6 The variation of the Bessel function with distance from a resonance and with depth of modulation (with acknowledgement to Jahnke, Emde and Lösch, Tables of Higher Functions, pub. B.G. Teubner, Stuttgart)

14. BEAM LIFETIME DUE TO MAGNET IMPERFECTIONS

We are at the end of a long road and in the last few miles have abandoned some of the luggage containing the rigour of mathematics to speed to our goal. Our conclusion is not to present a neat analytic expression into which one may plug field imperfections and obtain beam lifetime but to outline a method to arrive at this result. What is offered is a means to bypass the perhaps prohibitively expensive computer simulations needed to prove stability.

Previous attempts to predict lifetime on the basis of nonlinear field content have often foundered by not including either the concept of overlapping islands or the catalytic effect of ripple or synchrotron motion. When both of these ingredients are included it turns out that the nonlinear detuning, α, plays a crucial role in determining machine performance. This detuning may be due to effects other than the nonlinearity studied like beam-beam or ion. The scenario then is as follows:

i) Choose an order of resonance too high to be avoided by tuning Q.

ii) Calculate the nonlinear detuning due to ALL forms of nonlinearities.

iii) Invoke Q_S to find the spacing and strength of the resonance lines.

iv) Calculate the Chirikov limit, and if it is exceeded go on to calculate diffusion rates.

v) Calculate the growth per crossing.

vi) Assume crossings are uncorrelated because of Q.

vii) Sum the growth rate in random walk fashion to get a diffusion rate.

If this method proves to be effective one would expect a lifetime of 24 hours for a beam-beam effect in the CERN Sp$\bar{p}$S storage ring), ten minutes for ion-induced resonances in the CERN Antiproton Accumulator (Ref. [8]), and perhaps only a few tens of seconds for a large ring with inadequate power supply filtering). The experience found in the Refs. [7 to 9] suggests this may not be far from the truth.

Those who are already expert in the field may scoff at an attempt to describe beam diffusion ignoring the second transverse degree of freedom which changes the topology of the problem. Nevertheless, the numbers seem to fit and anyone with ambitions to include both transverse planes will find the next Section contains the prescription for this.

15. THE EFFECT OF TWO DIMENSIONS OF TRANSVERSE MOTION

The linear motion in one transverse degree of freedom is a circle in (J_1, ψ_1) space. The original Hamiltonian only included terms proportional to p^2 and to x^2 coming from the transverse distribution of A_S for a multipole (Eq. (10)) remembering that $m = 2$ for a quadrupole and a normal orientation gives only the imaginary terms. We find:

$$H = \frac{p_x^2}{2} + \left(\frac{1}{\rho^2} - k\right)\frac{x^2}{2} + \frac{p_z^2}{2} + k\frac{z^2}{2} \ . \tag{116}$$

We may transform these coordinates into action angle coordinates and remove the s dependence of H, J_x and J_z just as in the one-dimensional case. In the case of one degree of freedom, plus s the azimuth, the particle was confined to a torus. Slicing the torus at any azimuth gave a circle. With two degrees of freedom the "slice" will produce a four dimensional surface which becomes a circle when projected on either (J_x, ψ_x) plane or (J_z, ψ_z). Either J_x or J_z is a horizontal plane when plotted against $(\psi_x \psi_z)$ in a "hill and dale" fashion.

Now we move on to include other multipoles. Equation (10) tells us we should include in the Hamiltonian all the terms in the homogeneous polynomial expansion of:

$$(x + iz)^n \ . \tag{117}$$

We should select only real terms if the $2n$ poled magnet is in normal orientation and only imaginary terms if the magnet is skew. Thus a normal sextupole will give two terms:

$$x^3 - 3xz^2 \tag{118}$$

while a skew sextupole will give:

$$3x^2z - z^3 \ . \tag{119}$$

Now we can apply all the tricks of perturbation theory to be found in Sections 5 to 7. We simply must write two element vectors for Q, ψ, J and n taking dot products to form terms like $\boldsymbol{n}.\boldsymbol{\psi}$. One element is for the x plane, the other for the z plane.

When we come to Fourier analysis there will be two indices, n_1 for the phase angle ψ_x and n_2 for ψ_z. The denominator of Eq. (65) which defines a resonant condition becomes:

$$n_1 Q_x + n_2 Q_z - m = 0 \ . \tag{120}$$

The resonances determined by this condition are the familiar mesh of lines in the working diagram (Fig. 7). If m_1 and m_2 have the same sign they are sum resonances and if they are different, they are difference resonances.

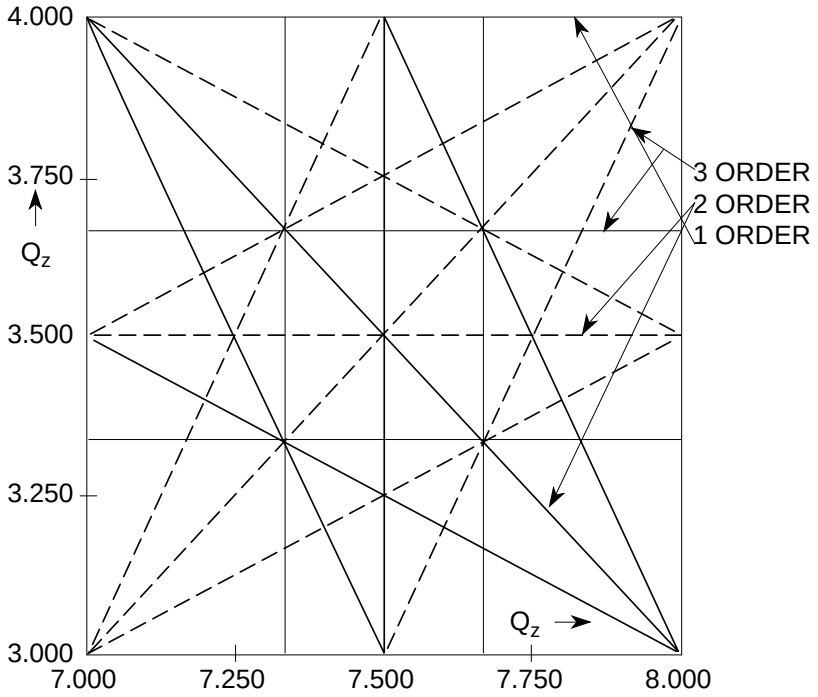


Fig. 7 Typical working diagram with sum and difference resonances to third order (Ref. [3])

The Hamiltonian, truncated to include just the resonant term (Eq. (67)) will become:

$$H_3 = H_0(\mathbf{J}_3) + RU_{n,m} e^{i(n_1\psi_x + n_2\psi_z - m\theta)} \ . \tag{121}$$

Defining:

$$n_1\varphi_x = n_1\psi_x + n_2\psi_z - m\theta \tag{122}$$

we can freeze the motion with a generating function reminiscent of Eq. (68) but with an extra term.

$$F_2(\boldsymbol{\psi}, J_4, \theta) = (\mathrm{n}\cdot\boldsymbol{\psi} - m\theta)J_{4,x} + \psi_z J_{4,z} \ . \tag{123}$$

We now find a new Hamiltonian:

$$H = (\boldsymbol{n}.\boldsymbol{Q} - m)J_x + Q_z J_z + RU_{00}(J) + RU_{\boldsymbol{n}m}\cos m\varphi_x \ . \tag{124}$$

We also find on making the transformation from (J_3, ψ_3) to (J_4, ψ_4) that:

$$\varphi_z = \Psi_z \tag{125}$$

$$J_{4,x} = n_1 J_{4,x} \tag{126}$$

$$J_{3,z} = n_2 J_{4,z} + J_{4,z} \ . \tag{127}$$

The invariants of the motion whose conjugates are absent from the Hamiltonian are:

$$H_4 = \text{constant} \tag{128}$$

$$J_{4z} = J_{3z} - \left(n_2 / n_1\right) J_{3x} \ . \tag{129}$$

Only opposite signs of n_1 and n_2 will ensure stability. Thus difference resonances are stable but can couple the two planes while sum resonances are unstable.

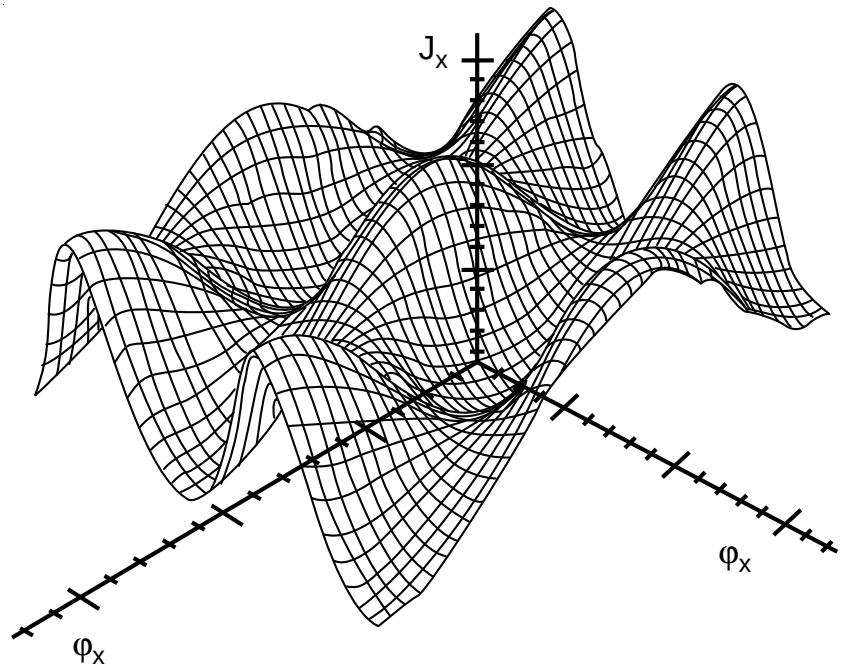


Fig. 8 A plot showing an invariant surface (H = constant) as a function of phase variables in other transverse planes (Ref. [3]).

The motion is difficult to plot or imagine. The exchange of J components ensures that particle simulation plotted on a single phase plane will not have a continuous path but appear as a halo of dots as J breathes. We no longer see undulations around a closed circle. However, Ruth [3] has suggested that if one plots either J_x or J_Z as a function of ϕ_x and ϕ_z to form a mountain range plot, one reveals the modes of distortion characteristic of the indices m_1 and m_2 (see Fig. 8).

16. CONCLUSIONS

We have covered all the basic nonlinear theory related to single-particle behaviour and in so doing identified a link between the specification of magnetic field quality and beam lifetime. There are also nonlinear forces on the beam due to collisions with an oncoming "target beam" in a collider or, if the particles of the beam are negative, with the space charge field of the neutralizing ion cloud. The potentials due to these effects are usually richer in high-order multipole fields at the beam radius. However, I shall now leave other authors to take up the baton in this relay race from alternating gradient focusing to Chirikov diffusion.

REFERENCES

[1] E.J.N. Wilson, Nonlinearities and Resonances, Proc. of fifth General CERN Accelerator School, Jyväskyla CERN 94-01, pp. 131-158 (1994).

[2] B. Montague, Basic Hamiltonian Dynamics, Proc. of this School.

[3] R.D. Ruth, Nonlinear Dynamics in Circular Accelerators and Storage Rings, Proc. of the 5th US Summer School High Energy Particle Accelerators, SLAC (1985).

[4] E.D. Courant, R.D. Ruth and W.T. Weng, Stability in Dynamical Systems I, SLAC-PUB-3415.

[5] E.D. Courant and H.S. Snyder, Theory of the Alternating-Gradient Synchrotron, Ann. Phys. 3 (1958).

[6] G. Guignard, Effets des champs magnetiques perturbateurs d'un synchrotron, CERN 70–24, Chapter 4 (1970).

[7] L.R. Evans, The Beam-Beam Interaction, Proc. of CERN Accelerator School on Antiprotons for Colliding Beam Facilities, Geneva, CERN 84-15, p. 319 (1984).

[8] E. Jones, F. Pedersen, A. Poncet, S. van der Meer and E.J.N. Wilson, Transverse Instabilities due to Beam-Trapped Ions and Charged Matter in the CERN Antiproton Accumulator, Proc. Particle Accel. Conf., Vancouver (1985).

[9] E.D. Courant, A Computer Study of Widening of Stopbands by Ripple, CERN Lab.II-DI-PA/Int. 74-14 (1974).

[10 R. Stiening and E.J.N. Wilson, Transverse Single Particle Instability in the NAL 500 GeV Accelerator, Nucl. Instrum. Methods, Vol.I 121 (1974).

BETATRON COUPLING AND RELATED IMPACT OF RADIATION

G. Guignard
CERN, Geneva, Switzerland

Abstract
Linear coupling of vertical and horizontal oscillations in a circular accelerator or collider is analysed in the presence of solenoidal and tilted-quadrupole fields. An analytical treatment of the coupled motions based on a Hamiltonian formalism is proposed. Solutions are first given explicitly in absence of synchrotron radiation and then extended to the case of loss of energy and photon emission by the particles. Applications of the derived formulae to a generic accelerator are presented. Pertinent information related to the large Electron-Positron collider of CERN is conveyed in the last part, in which strategies for dealing with a very large, low harmonic, tilted quadrupole are detailed, and in which a particular scheme for local compensation of solenoids is discussed.

1. INTRODUCTION AND SUMMARY

In circular accelerators and colliders of particles, a correct compensation of the betatron coupling is crucial for the ring operation and for the performance. The presence of too strong coupling may indeed induce optics distortions, transfer of oscillations from one plane to the other, tilt of the normal directions in which the betatron motion is again decoupled, beating of the betatron motion and shifts of the wave numbers or tunes. All these effects perturb the checking and the control of the designed optics and may even confuse diagnostics such as tune measurements when running conditions are particularly sensitive to coupling effects. Also critical is the implication of linear coupling on the performance of an accelerator and the luminosity of a collider, because of its impact on the transverse emittances. When the collider conveys electrons and positrons, like the Large Electron-Positron (LEP) storage ring at CERN, the equilibrium emittances are critically dependent on the combined influence of betatron coupling and synchrotron radiation in the bending magnets. This phenomenon is all the more important in LEP as a good luminosity requires a flat beam, i.e. a vertical emittance many times smaller than the horizontal one. This article therefore presents the observations made on linear coupling during LEP commissioning and the practical means used to control it. It also describes the analytical tools which have been developed to help in understanding the physics of the mechanisms involved, analysing the measurements done and designing correction schemes or strategies. The first part gives the general basic formulae governing coupling in absence of radiation and the second shows how to use them for diagnostic purposes. The next one describes how coupled betatron oscillations are in turn affected by photon emission and energy loss in lepton rings, while the last section emphasises the application to LEP of the treatment presented and the strategy of compensation applied in such a practical case.

2. HAMILTONIAN TREATMENT OF LINEAR COUPLING

2.1 Sources of coupling and form of the perturbed Hamiltonian [1,2]

Since the perturbation theory gives the possibility to find the exact equations of perturbed betatron motion (see Appendix), it is convenient to consider linear coupling as a perturbation of the transverse particle oscillations. These oscillations can be coupled by three-dimensional magnetic fields of components B_x, B_z and $B\theta$ (longitudinal). Basically, there are two sources

of coupling, i.e. skew-quadrupole and solenoidal field components, keeping in mind that finite vertical orbit in sextupoles generates a skew-quadrupole field. Specific examples of magnetic elements which can generate coupling are:

1) Tilted-quadrupole lenses for which one has

$$K(\theta) = \frac{R^2}{2B\rho}\left(\frac{\partial B_x}{\partial x} - \frac{\partial B_z}{\partial z}\right) \tag{1}$$

2) Solenoidal fields for which we define

$$S(\theta) = \frac{R}{2B\rho} B_\theta \tag{2}$$

3) End effects of large solenoids, described by

$$K(\theta) = (2a_s - 1)\dot{S} \tag{3}$$

where a_S characterises the geometry of the end plates. For horizontal slots, $a_S = 1$, and for open ends, $a_S = 1/2$.

4) Sextupoles with vertical orbit deviations z_O,

$$K(\theta) = -\frac{R^2}{B\rho}\frac{\partial^2 B_z}{\partial x^2}\cdot z_o \tag{4}$$

where the second derivative of B_Z characterizes the sextupole strength.

A three-dimensional field couples the equations of motion in the following way [3]

$$\begin{aligned} \ddot{x} + K_1(\theta)x &= \frac{R^2}{B\rho}\frac{\partial B_z}{\partial z} z - \frac{R}{B\rho} B_\theta \dot{z} \\ \ddot{z} + K_2(\theta)z &= \frac{R^2}{B\rho}\frac{\partial B_x}{\partial x} x + \frac{R}{B\rho} B_\theta \dot{x} \ . \end{aligned} \tag{5}$$

Equations (5) can be rewritten using the definitions (1) and (2) as well as the property div $\boldsymbol{B} = 0$, that links together the transverse and the longitudinal derivatives.

$$\begin{aligned} \ddot{x} + K_1(\theta)x &= -\left(K + \dot{S}\right)z - 2S\dot{z} \\ \ddot{z} + K_2(\theta)z &= -\left(K - \dot{S}\right)x + 2S\dot{x} \end{aligned} \tag{6}$$

all the functions depending on the variable θ, in general.

The form of the Hamiltonian H associated with Eqs. (6) can be shown to be

$$H = \frac{1}{2}\left[K_1 x^2 + K_2 z^2 + 2Kxz + \left(p_x - Sz\right)^2 + \left(p_z + Sx\right)^2\right] \tag{7}$$

The proof of this comes from writing the subsequent canonical equations, according to (A.3)

$$
\begin{aligned}
\dot{x} &= \frac{\partial H}{\partial p_x} = p_x - Sz \\
\dot{p}_x &= -\frac{\partial H}{\partial x} = -K_1 x - Kz - (p_z + Sx)S \\
\dot{z} &= \frac{\partial H}{\partial p_z} = p_z + Sx \\
\dot{p}_z &= -\frac{\partial H}{\partial z} = -K_2 z - Kx - (p_x - Sz)S \ .
\end{aligned}
\tag{8}
$$

The first and third equations can be solved for p_ρ and then differentiated with respect to θ. These two successive results can then be included in the second and fourth equations. Rearranging the terms gives finally the two equations (6), proving the validity of our pre-supposed form (7).

Knowing H_0 (A.12) for betatron motion, the perturbed Hamiltonian for linear coupling is obtained by subtracting H_0 from H (7).

$$
H_1(y, p_y) = \left[Kxz + Sxp_z - Szp_x + \frac{1}{2}S^2(x^2 + z^2) \right] \tag{9}
$$

which is a quadratic form in the y, p_y coordinates.

Perturbation theory imposes then obtaining an explicit form of $U = H_1$, as function of the constants a_1 and a_2 of the betatron motion. Note that this form is subordinated to the following properties:

a) H_1 is a quadratic function (9) of the coordinates and momentum conjugates.

b) The solutions of the unperturbed betatron motion (A.19) are linear functions of the four constants a_1, $\bar{a}_1$, a_2, and $\bar{a}_2$ and contain oscillatory terms with frequencies Q_x and Q_z.

c) For circular accelerators and storage rings, the functions K and S , characterizing the linear-coupling perturbation (9), are obviously periodic in θ with period 2π.

Introducing solutions (A.19) into the expression (9) is a somewhat tedious calculation, that the reader interested in the derivation of U is encouraged to do and that provides [1,2]

$$
U = \sum_{j,k,l,m=0}^{2} h^{(2)}_{jklm} \; a_1^j \; \bar{a}_1^k \; a_2^l \; \bar{a}_2^m \exp\left\{ i\left[(j-k)Q_x + (l-m)Q_z \right]\theta \right\} \tag{10}
$$

with the rules coming from the above properties that the sum $j+k+l+m$ is equal to 2 and any index only takes one of the values 0, 1 or 2. The coefficients h [2] obviously depend on the Floquet functions u and v (A.18) and on the coupling forces K (1) and S (2). They are given explicitly hereafter [2]

$$
\begin{aligned}
h^{(2)}_{2000} &= \frac{1}{2}S^2u^2 \qquad h^{(2)}_{0200} = \frac{1}{2}S^2\bar{u}^2 \qquad h^{(2)}_{1100} = S^2u\bar{u} \\
h^{(2)}_{0020} &= \frac{1}{2}S^2v^2 \qquad h^{(2)}_{0002} = \frac{1}{2}S^2\bar{v}^2 \qquad h^{(2)}_{0011} = S^2v\bar{v} \\
h^{(2)}_{1010} &= Kuv + S\left[u\left(\dot{v} + iQ_zv\right) - v\left(\dot{u} + iQ_xu\right)\right] \\
h^{(2)}_{1001} &= Ku\bar{v} + S\left[u\left(\dot{\bar{v}} - iQ_z\bar{v}\right) - \bar{v}\left(\dot{u} + iQ_xu\right)\right] \\
h^{(2)}_{0101} &= \text{c.c. of } h^{(2)}_{1010} \qquad h^{(2)}_{0110} = \text{c.c. of } h^{(2)}_{1001}
\end{aligned}
\tag{11}
$$

where c.c. stands for complex conjugate.

The property c) mentioned above makes it possible and judicious to develop the coefficients h [2] in Fourier's series,

$$
h^{(2)}_{jklm}(\theta) = \sum_{q=-\infty}^{+\infty} h^{(2)}_{jklmq} e^{iq\theta} \tag{12}
$$

with

$$
h^{(2)}_{jklmq} = \frac{1}{2\pi}\int_{o}^{2\pi} h^{(2)}_{jklm}(\theta) e^{-iq\theta} d\theta \ .
$$

Putting Eq. (12) into (10) adds another sum over the harmonics q and modifies slightly the phase term or argument of the exponential function,

$$
\text{Argument } = i\left[(j-k)Q_x + (l-m)Q_y + q\right]\theta \ . \tag{13}
$$

Let us introduce now our *first approximation*: we assume, with other authors [4, 5], that the low-frequency part of the Hamiltonian (10) gives the important variations of the constants, a_1, $\bar{a}_1$, a_2, $\bar{a}_2$.

The low frequency part corresponds to the special choices of the indices that are cancelling the argument (13). The corresponding conditions are:

1) $j = k \quad l = m \quad q = 0$
with j and l taking either the value 0 or 1.

2) $(j - k)Q_X + (l - m)Q_Z + q = 0$, as a whole.
Redefining $n_1 = j - k$, $n_2 = l - m$ and $p = -q$, this is equivalent to the well-known resonance condition

$$
n_1Q_X + n_2Q_Z - p = 0
$$

where n_1 can take the values 0, 1 or 2 and n_2 the values 0, ±1 or 2 according to selection rules.

With this approximation to the low-frequency part, the Hamiltonian U can be explicitly written [2]

$$
\begin{aligned}
U &= h^{(2)}_{11000}a_1\bar{a}_1 + h^{(2)}_{00110}a_2\bar{a}_2 && \text{Freq. Shift} \\
&+ h^{(2)}_{1010-p}a_1a_2e^{i(Q_x+Q_z-p)\theta} + c.c. && \text{Sum Resonance} \\
&+ h^{(2)}_{1001-p}a_1\bar{a}_2e^{i(Q_x-Q_z-p)\theta} + c.c. && \text{Diff. Resonance} \\
&+ h^{(2)}_{2000-p}a_1^2e^{i(2Q_x-p)\theta} + c.c. && \text{Hor. Resonance} \\
&+ h^{(2)}_{0020-p}a_2^2e^{i(2Q_z-p)\theta} + c.c. && \text{Vert. Resonance}
\end{aligned}
\tag{14}
$$

where c.c. stands for complex conjugate again. Equations (14) indicate that coupling perturbation implies frequency (or tune) shifts with amplitude, sum and difference coupling resonances, but also one-dimensional quadrupole resonances. However, only sum and difference resonance driving terms are to first order in the perturbation functions (1) to (4), the others being to second order. This can be seen in the expressions of these driving terms, deduced from Eqs. (11) to (13) and Eqs. (A.17) and (A.18),

$$
\begin{aligned}
h^{(2)}_{\substack{11000\\00110}} &= \frac{1}{8\pi R}\int_o^{2\pi} S^2\beta_y d\theta \\
h^{(2)}_{\substack{2000\\0020-p}} &= \frac{1}{8\pi R}\int_o^{2\pi} S^2\beta_y e^{i(2\mu_y-2Q_y+p)\theta} d\theta \\
h^{(2)}_{\substack{1001\\1010-p}} &= \frac{1}{4\pi R}\int_o^{2\pi}\sqrt{\beta_x\beta_z}\left[K + RS\left(\frac{\alpha_x}{\beta_x}-\frac{\alpha_z}{\beta_z}\right) - iRS\left(\frac{1}{\beta_x}\pm\frac{1}{\beta_z}\right)\right]\times \\
&\quad \times \exp\left\{i\left[(\mu_x \mp \mu_z) - (Q_x \mp Q_z - p)\theta\right]\right\} d\theta
\end{aligned}
\tag{15}
$$

where the coordinate y (x or z) and the sign in the last expression have to be chosen according to the indices of $h^{(2)}$ (x and top sign with the upper indices, and vice versa).

For the "first-order" (in the perturbation function) resonances, coupling coefficients proportional to the driving terms can be written by definition [2, 6],

$$
C^+ = 2h^{(2)}_{1010-p} \qquad C^- = 2h^{(2)}_{1001-p} \tag{16}
$$

with, in addition

$$
\begin{aligned}
\Delta^{\mp} &= Q_x \mp Q_z - p \\
\lambda_x &= h^{(2)}_{11000} \quad \lambda_z = h^{(2)}_{00110} \ .
\end{aligned}
\tag{17}
$$

Effects of coupling are then frequency shifts and excitation of several resonances. Equations of motion of the type (A.22) can be written and explicit solutions have been calculated [2] for each resonance separately, including the "second-order" (in the perturbation function) tune shifts (characterized by λ_x and λ_z).

2.2 Solution of coupled motion near a single resonance [2]

It is possible to find the explicit solutions for every resonance excited by linear-coupling sources, i.e. $Q_x - Q_z$, $Q_x + Q_z$, $2Q_x$ and $2Q_z$, which are all of second order in the wave numbers (this order being defined by the absolute-value sum of n_1 and n_2, the coefficients of the tunes). Taking separately each resonance and including the terms λ_x and λ_z of Eqs. (17),

let us first deal with the two more important sum- and difference-resonances, that are to first order in the perturbation $K(\theta)$ and $S(\theta)$.

Introducing the expression (14) of U for one single resonance at the time in the equations (A.22) and using the definitions (16) and (17), the explicit equations of motion can be written in either case of a dominant sum resonance or difference resonance.

$$\begin{aligned} \frac{da_1}{d\theta} &= i\lambda_x a_1 + if_a \ e^{-i\theta\Delta} \\ \frac{da_2}{d\theta} &= i\lambda_z a_2 + ig_a \ e^{\pm i\theta\Delta} \end{aligned} \tag{18}$$

where one has introduced

1) $f_a = \frac{1}{2}\overline{C}^- a_2$, $g_a = \frac{1}{2}C^- a_1$ for difference resonance

2) $f_a = \frac{1}{2}\overline{C}^+ \overline{a}_2$, $g_a = \frac{1}{2}\overline{C}^+ \overline{a}_1$ for sum resonance .

The sign + or - of the exponent in the second equation (18) concerns case 1) or 2), and Δ is equal to Δ^- or Δ^+, respectively.

Equations (18) are the manifestation of our *second approximation*: we assume that the working point $\{Q_x, Q_z\}$ is close enough to one resonance due to coupling and remote from the others, so that one can keep only the corresponding term of U. To be able to solve explicitly the equations of motion, this approximation is required, but to compensate coupling effects, both resonance driving-terms (16) can simultaneously be cancelled.

Since λ_x and λ_z are real quantities, we can make the following change in the variables in order to solve (18):

$$a_1 = b_1 \ e^{i\lambda_x\theta} \qquad a_2 = b_2 \ e^{i\lambda_z\theta} \ . \tag{19}$$

Putting (19), which is the solution of (18) for $C^- = 0$ or $C^+ = 0$, in (18) gives the equations for b_1 and b_2

$$\begin{aligned} \frac{db_1}{d\theta} &= if_b \ e^{-iD\theta} \\ \frac{db_2}{d\theta} &= ig_b \ e^{\pm iD\theta} \end{aligned} \tag{20}$$

where f_b and g_b are defined as f_a and g_a in the items 1) and 2) above, after replacing a_1, a_2 by b_1, b_2, and the appropriate sign is chosen according to the resonance considered. The coefficient D has also a definition that depends on the resonance and the whole thing can be summarized as follows,

1) Difference resonance

$$f_b = \frac{1}{2}\overline{C}^- b_2 \ ; \quad g_b = \frac{1}{2}C^- b_1 \ ; \quad D = \Delta^- + \lambda_x - \lambda_z$$

2) Sum resonance

$$f_b = \frac{1}{2}\overline{C}^{+}\overline{b}_2 \;; \quad g_b = \frac{1}{2}\overline{C}^{+}\overline{b}_1 \;; \quad D = \Delta^{+} + \lambda_x + \lambda_z \;.$$

The set of equations (20) can be relieved of the exponential functions by defining a new variable $\tilde{b_2}$

$$\tilde{b_2} = b_2 \; e^{\mp iD\theta} \tag{21}$$

as is shown in the following results,

$$\begin{aligned}
&\frac{\mathrm{d}b_1}{\mathrm{d}\theta} = if_{b^{\sim}} \qquad \text{with } f_{b^{\sim}} = \frac{1}{2}\overline{C}^{-}\tilde{b_2} \text{ or } \frac{1}{2}\overline{C}^{+}\tilde{\overline{b}_2} \\
&\frac{\mathrm{d}\tilde{b_2}}{\mathrm{d}\theta} = \frac{1}{2}iC^{-}b_1 - iD\tilde{b_2} \qquad \text{for difference resonance} \\
&\frac{\mathrm{d}\tilde{\overline{b}_2}}{\mathrm{d}\theta} = -\frac{1}{2}iC^{+}b_1 - iD\tilde{\overline{b}_2} \quad \text{for sum resonance .}
\end{aligned} \tag{22}$$

Combining the equation for b_1 with either the first (difference resonance) or the second (sum resonance) equation for $\tilde{b_2}$, it is possible to write a second-order equation in either case,

$$\begin{aligned}
&\frac{\mathrm{d}^2\tilde{b_2}}{\mathrm{d}\theta^2} + iD\frac{\mathrm{d}\tilde{b_2}}{\mathrm{d}\theta} + \frac{C^{-}\overline{C}^{-}}{4}\tilde{b_2} = 0 \quad \text{(Difference)} \\
&\frac{\mathrm{d}^2\tilde{\overline{b}}_2}{\mathrm{d}\theta^2} + iD\frac{\mathrm{d}\tilde{\overline{b}}_2}{\mathrm{d}\theta} - \frac{C^{+}\overline{C}^{+}}{4}\tilde{\overline{b}}_2 = 0 \quad \text{(Sum) .}
\end{aligned} \tag{23}$$

The solutions of these two equations solve the perturbation problem defined in Section 2.1 and are very similar, since the equations are themselves similar. One obtains,

$$\begin{aligned}
b_1 &= \frac{1}{2}\overline{C}\left(\frac{B_1}{\omega_2}e^{i\omega_2\theta} + \frac{B_2}{\omega_1}e^{i\omega_1\theta}\right) \\
\tilde{b_2} \text{ or } \tilde{\overline{b}}_2 &= B_1 e^{-i\omega_1\theta} + B_2 e^{-i\omega_2\theta}
\end{aligned} \tag{24}$$

with the following definition of the ω's,

$$\begin{aligned}
\omega_{1,2} &= \frac{1}{2}\left[-D \pm \sqrt{D^2 + \left|C^{-}\right|^2}\right] \quad \text{(Difference)} \\
\omega_{1,2} &= \frac{1}{2}\left[-D \pm \sqrt{D^2 - \left|C^{+}\right|^2}\right] \quad \text{(Sum)}
\end{aligned} \tag{25}$$

where the indices 1, 2 are associated with the double sign in front of the square root, and D is defined in the paragraph below Eq. (20). At times, for convenience, one introduces by definition $\eta^{\mp} = \sqrt{D^2 \pm \left|C^{\mp}\right|^2}$. The quantities B_1 and B_2 are now the actual complex constants of the perturbed motion.

Moving backwards through Eqs. (24), (21) and (19), one can write the variation with θ of the "constants" a_1 and a_2 of the unperturbed motion, in the case of linear betatron-coupling. Near a single *difference resonance* one has

$$\begin{aligned} a_1 &= \frac{1}{2}\overline{C}^{-}\left(\frac{B_1}{\omega_2}e^{i\omega_{x2}\theta} + \frac{B_2}{\omega_1}e^{i\omega_{x1}\theta}\right) \\ a_2 &= B_1\ e^{-i\omega_{z2}\theta} + B_2\ e^{-i\omega_{z1}\theta} \end{aligned} \tag{26}$$

where the wave numbers ω depend now not only on the sign in (25) but also on the plane considered, by virtue of (19) that implies $\lambda_X \neq \lambda_Z$, $\lambda_X \neq 0$ and $\lambda_Z \neq 0$,

$$\begin{aligned} \omega_{x1,2} &= \frac{1}{2}\left[-\left(\Delta^{-} - \lambda_x - \lambda_z\right) \pm \sqrt{\left(\Delta^{-} + \lambda_x - \lambda_z\right)^2 + \left|C^{-}\right|^2}\right] \\ \omega_{z1,2} &= \frac{1}{2}\left[-\left(\Delta^{-} + \lambda_x + \lambda_z\right) \mp \sqrt{\left(\Delta^{-} + \lambda_x - \lambda_z\right)^2 + \left|C^{-}\right|^2}\right]. \end{aligned} \tag{27}$$

These equations indicate that in this case the motion is always stable with amplitude beating and exchange between the two transverse directions. Similarly, near a single *sum resonance* one can write

$$\begin{aligned} a_1 &= \frac{1}{2}\overline{C}^{+}\left(\frac{B_1}{\omega_2}e^{i\omega_{x2}\theta} + \frac{B_2}{\omega_1}e^{i\omega_{x1}\theta}\right) \\ a_2 &= \overline{B}_1\ e^{i\omega_{z1}\theta} + \overline{B}_2\ e^{i\omega_{z2}\theta} \end{aligned} \tag{28}$$

and the wave numbers are slightly different from (27),

$$\begin{aligned} \omega_{x1,2} &= \frac{1}{2}\left[-\left(\Delta^{+} - \lambda_x + \lambda_z\right) \pm \sqrt{\left(\Delta^{+} + \lambda_x + \lambda_z\right)^2 - \left|C^{+}\right|^2}\right] \\ \omega_{z1,2} &= \frac{1}{2}\left[-\left(\Delta^{+} + \lambda_x + \lambda_z\right) \pm \sqrt{\left(\Delta^{+} + \lambda_x + \lambda_z\right)^2 - \left|C^{+}\right|^2}\right]. \end{aligned} \tag{29}$$

These last expressions indicate that the motion near a sum resonance can be stable or unstable, depending on the amplitude of $|C^{+}|$. It is stable if ω_X and ω_Z are real. In other words, this means that

if $|C^{+}| \leq |\Delta^{+} + \lambda_X + \lambda_Z|$, the motion is stable,
if $|C^{+}| > |\Delta^{+} + \lambda_X + \lambda_Z|$, the motion is unstable.

When the motion is unstable, amplitudes may grow to infinity in both transverse directions.

The forms of the equations (26) to (29) show that the perturbed betatron motion is made of two modes associated with two different frequencies ω_1 and ω_2. Results (27) and (29) show moreover that the inclusion of the S^2 terms from the Hamiltonian, i.e. λ_X and λ_Z, only slightly modifies the frequencies of the two modes and that this effect is different from the horizontal and vertical mode. In the special case where the terms λ_X and λ_Z are neglected, Eqs. (27) and (29) simplify to

$$\omega_{1,2} = \frac{1}{2}\left[-\Delta^- \pm \sqrt{\Delta^{-2} + \left|C^-\right|^2}\right] = \frac{1}{2}\left[-\Delta^- \pm \eta^-\right]$$
$$\omega_{1,2} = \frac{1}{2}\left[-\Delta^+ \pm \sqrt{\Delta^{+2} - \left|C^+\right|^2}\right] = \frac{1}{2}\left[-\Delta^+ \pm \eta^+\right] \tag{30}$$

and become independent of the plane considered.

Considering now only a difference resonance and neglecting the tune shifts with amplitude λ_X and λ_Z according to the first equation (30), the complete solution can be deduced from Eqs. (A.19) and the definitions (A.17) and (A.18). Dealing only with the amplitudes, and ignoring for simplicity the conjugate momenta, the solution is,

$$x(\theta) = \frac{1}{2}\bar{C}^- \left(\frac{B_1}{\omega_2} e^{i\omega_2\theta} + \frac{B_2}{\omega_1} e^{i\omega_1\theta}\right)\sqrt{\frac{\beta_x}{2R}} e^{i\mu_x} + c.c.$$
$$z(\theta) = \left(B_1\, e^{-i\omega_1\theta} + B_2\ e^{-i\omega_2\theta}\right)\sqrt{\frac{\beta_z}{2R}} e^{i\mu_z} + c.c. \tag{31}$$

With these equations, we went as far as possible in dealing with linear coupling by using perturbation theory with the Hamiltonian formalism.

Before closing this section, let us make two more remarks. The first one concerns the effects related to λ_X and λ_Z. Considering as an example the situation where C^- is vanishing after compensation, the solution of the motion is given by (19) together with (A.19) and therefore Eq. (31) is replaced by

$$x(\theta) = b_1\sqrt{\frac{\beta_x}{2R}} e^{i(\mu_x + \lambda_x\theta)} + c.c.$$
$$z(\theta) = b_2\sqrt{\frac{\beta_z}{2R}} e^{i(\mu_z + \lambda_z\theta)} + c.c. \tag{32}$$

The effects referred to are different tune shifts in the two transverse directions, the amplitudes of which are given by first Eq. (15), hence proportional to the square of $S(\theta)$ and to betatron amplitude. Therefore, one has to keep in mind that it might be necessary to adjust the tunes after compensation of solenoidal fields with tilted quadrupoles.

The second remark concerns the one-dimensional resonances $2Q_y$-p, that are considered as less important because they are to second order in the perturbation $S(\theta)$. In fact, large stopbands are opened near each integer or half-integer by the strong quadrupoles of the nominal lattice, responsible for $K_{1,2}$ (A.13). The effect of $S(\theta)$ in this respect is just to slightly modify the forbidden bandwidths around these particular values, i.e. the quadrupole driving terms, according to

$$\kappa = \frac{1}{8\pi R}\int_o^{2\pi}\left(K_{1,2} + S^2\right)\beta_y\ e^{i\left(2\mu_y - 2Q_y + p\right)} d\theta\ . \tag{33}$$

Usually S^2 is small with respect to the nominal $K_{1,2}$ and therefore negligible. But a judicious compensation of the tune shifts (32) mentioned might also be able to cancel the effects on κ if the corresponding change $\Delta K_{1,2}$ balances the S^2 term in (33).

3. MEASURING THE DIFFERENCE COUPLING COEFFICIENT

The solutions of the linearly coupled motions (31) give a clue to possible methods for measuring the coupling coefficient C^- of the difference resonance. Using the form of the coupled motions following a dipole kick in one transverse direction, one can show that the subsequent coherent oscillations can be analysed for measuring the real and imaginary part of this complex coefficient. More generally, considering the two frequency modes (30) that characterize the motions (31) and mapping out their separation as a function of the tunes can be used for measuring the module $|C^-|$. These two methods are described below.

3.1 Coupling measurement from coherent oscillations [7-9]

Let us summarize the analysis of the transverse oscillations after kicking the beam horizontally, the details of which can be found in Ref. [7]. The equations of motion (31) are functions of two complex constants (four real constants) that can be defined from the initial conditions of the motion. According to our assumption, these conditions are

$$\begin{aligned} x_o = z_o = \dot{z}_o = 0 \\ \dot{x}_o \neq 0 \ . \end{aligned} \tag{34}$$

Now, it is useful to rewrite (31) by using the sinusoidal functions and bringing into evidence the two modes characterized by the frequencies ω_1 and ω_2.

$$\begin{aligned} x = \left(\frac{4R^2}{\beta_{x_o}\beta_{z_o}} - C_1^2\right)^{-1} \frac{\sqrt{\beta_x}}{\eta} [d_1 \cos(\mu_x + \omega_1\theta) - e_1 \sin(\mu_x + \omega_1\theta) \\ - d_2 \cos(\mu_x + \omega_2\theta) + e_2 \sin(\mu_x + \omega_2\theta)] \\ z = \left(\frac{4R^2}{\beta_{x_o}\beta_{z_o}} - C_1^2\right)^{-1} \frac{\sqrt{\beta_z}}{\eta} [f_1 \cos(\mu_z - \omega_2\theta) - g_1 \sin(\mu_z - \omega_2\theta) \\ - f_2 \cos(\mu_z - \omega_1\theta) + g_2 \sin(\mu_z - \omega_1\theta)] \end{aligned} \tag{35}$$

where β_{xo}, β_{zo} are the betatron amplitudes at the kicker position, C_1 (and C_2) the real (and imaginary) parts of C^-, d, e, f and g new constants of motion and η is equal to η^- defined after Eq. (25) or in (30). After some cumbersome algebra [7], it is possible to write the constants as functions of the initial conditions (34),

$$\begin{aligned} d_{1,2} &= \frac{C_1 C_2}{\sqrt{\beta_{x_o}}} \dot{x}_o \\ e_{1,2} &= -\left(\frac{R|C^-|^2}{\omega_{1,2}\beta_{z_o}} - C_1^2\right) \frac{\dot{x}_o}{\sqrt{\beta_{x_o}}} \\ f_{1,2} &= 2\frac{RC_2}{\beta_{z_o}\sqrt{\beta_{x_o}}} \dot{x}_o \\ g_{1,2} &= -2C_1\left(\frac{R}{\beta_{z,o}} - \omega_{1,2}\right) \frac{\dot{x}_o}{\sqrt{\beta_{x_o}}} \ . \end{aligned} \tag{36}$$

One observes from (35) that the single-particle motion contains fast and slow oscillations associated with the phase μ_y and $\omega_{1,2}\theta$, respectively. It is therefore possible, using the combination law of sinusoidal functions, to factorise the signal into a slowly oscillating envelope and a fast oscillating component [7]. The result is:

$$\begin{aligned} x(\theta) &= \frac{\sqrt{\beta_x}}{\eta} E_x \cos\left(\mu_x - \frac{\Delta}{2}\theta - \phi_x\right) \\ z(\theta) &= \frac{\sqrt{\beta_z}}{\eta} E_z \cos\left(\mu_z + \frac{\Delta}{2}\theta - \phi_z\right) \end{aligned} \tag{37}$$

where both the phases ϕ_X, ϕ_Z and the envelopes E_X, E_Z are functions of the coefficients d, e, f and g, and Δ stands for the distance (17) from the difference resonance "measured". In the particular case of a horizontal kick, with the particular coefficients (36), the envelopes become, after neglecting the terms containing $|C|\beta_{Z0}/R$ (<< 1),

$$\begin{aligned} E_x^2 &= \frac{\beta_{x_o}}{R^2}\left[\eta^2 - \left|C^-\right|^2 \sin^2\frac{\eta}{2}\theta\right]\dot{x}_o^2 \\ E_z^2 &= \frac{\beta_{x_o}}{R^2}\left|C^-\right|^2 \sin^2\frac{\eta}{2}\theta \;\; \dot{x}_o^2 \; . \end{aligned} \tag{38}$$

It comes out from Eq. (38) that the fraction F of the energy (taken here as the square of the envelope amplitudes) interchanged between the two signals, the ratio G of the minimum to the maximum of the horizontal envelope and the period T of the envelope oscillations (involving the revolution frequency f_{rev}) are

$$\begin{aligned} F &= \frac{\left|C^-\right|^2}{\eta^2} = \frac{\left|C^-\right|^2}{\Delta^2 + \left|C^-\right|^2} \\ G &= \frac{|\Delta|}{\eta} = \frac{|\Delta|}{\sqrt{\Delta^2 + \left|C^-\right|^2}} \\ T &= \frac{1}{\eta \; f_{rev}} \end{aligned} \tag{39}$$

in agreement with Refs. [7] and [8]. It is interesting to note that the maxima and minima of the envelopes (38) appear for $\eta\theta = n\pi$, n being an integer, and that these envelopes are independent of the phase of C^- in the complex plane, if the assumption $R/\beta_{Z0} >> |C^-|$ is verified. Consequently, the knowledge of C_1 and C_2 requires complementary measurements, like the response to an inclined kick, such as

$$\begin{aligned} & x_o = z_o = 0 \\ & \dot{x}_o \neq 0 \quad \dot{z}_o \neq 0 \quad \dot{x}_o / \dot{z}_o = \sqrt{\beta_{z_o} / \beta_{x_o}} \; . \end{aligned} \tag{40}$$

More precisely, the coherent oscillations following a horizontal kick have the characteristics shown in Fig. 1 [9], accordingly with Eqs. (38). In the plane parallel to the kick, the initial amplitude of the envelope is finite ($\dot{x}_o \neq 0$) while it is zero in the orthogonal plane. With the time, the envelopes beat with the period T and the horizontal one has minima

when the vertical one has maxima. This illustrates the interchange of energy already mentioned, in the presence of second-order (in tune) difference resonance. Easily measured are the period T and the ratio G of the minimum to the maximum of the horizontal amplitude. Solving Eqs. (39), such measurements give the distance from, and the driving term of, the resonance

$$|\Delta| = \frac{G}{Tf_{rev}} \qquad f_{rev} = \text{revolution frequency}$$
$$|C^-| = \frac{1}{Tf_{rev}}\sqrt{1-G^2} \tag{41}$$

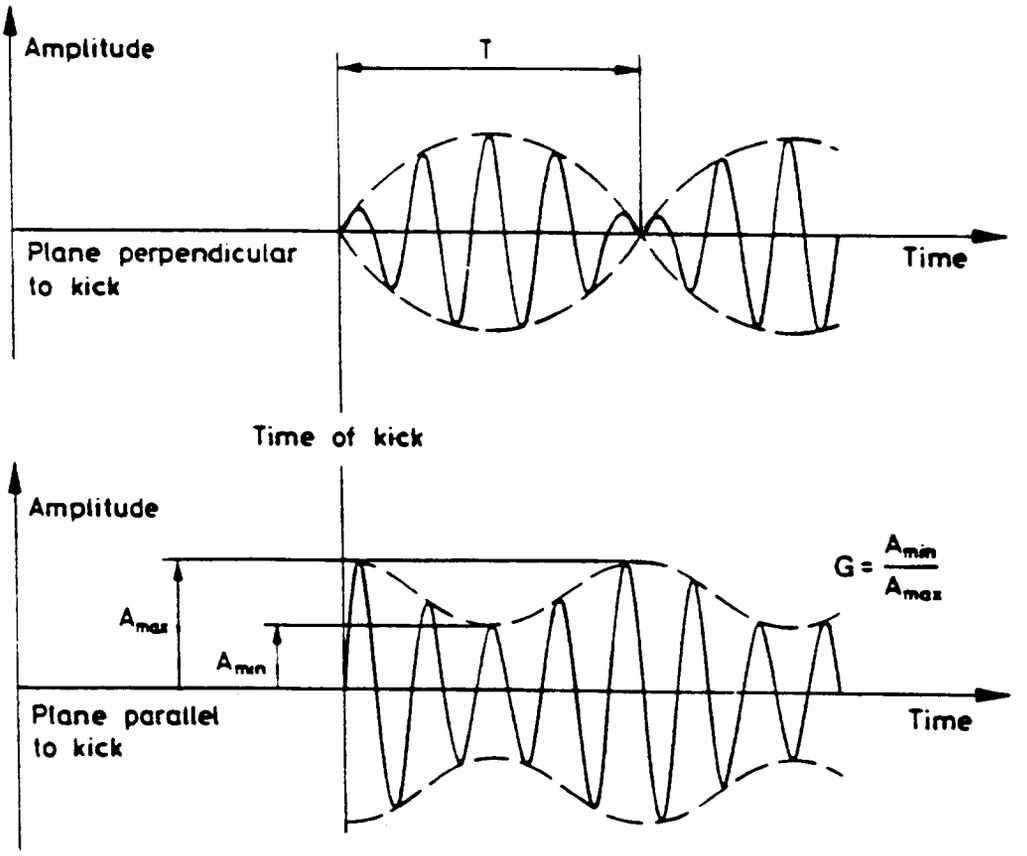


Fig. 1 Coherent oscillations following a horizontal kick

As an illustration of a practical use of this method, Fig. 2 shows one of the first signals obtained in the now dismantled Intersecting Storage Rings (at CERN) from the filter output of the device that was used to measure the tunes (kicker and large band pick-up). It is clear from this picture that the coherency of these signals (proton beam) was more than sufficient to measure T and G with a very good accuracy. This question of coherency has to be considered when designing a coupling-meter, in particular for an electron beam, and the kicker has to be strong enough to generate oscillations of amplitude which are large with respect to the pick-up resolution. The revolution frequency does not limit the precision of this method since it is

usually known to a high accuracy. In the case of Fig. 2, the measured value was $|C^-| = 1.2\ 10^{-2}$ and corresponded to the residual coupling of the ring plus the contribution of one experimental solenoid.

In the ISR, the carrier frequency was 30-300 kHz and modulation frequency 0.5-10 kHz. The pick-up signal was passing through a rectifier followed by a sharp-edge, low-pass filter, and differentiated. The zero crossings were used for measuring T and triggering units which stored maxima and minima of the signal. The precision was ± 3% for the maximum and minimum values and ± 1% for T, giving about 4% on $|C^-|$ and $|\Delta|$.

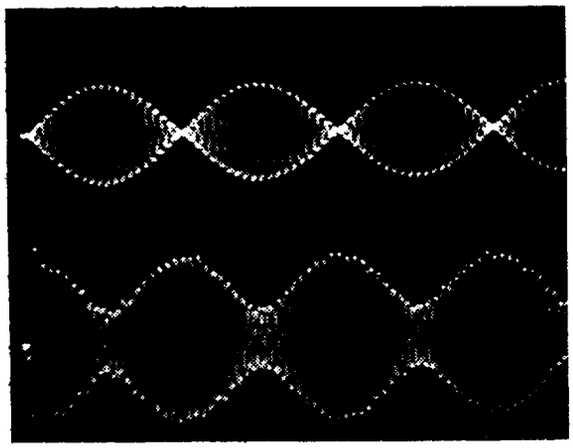


Fig. 2 Vertical (above) and horizontal (below) coherent oscillations measured in the Intersecting Storage Rings (CERN) after a horizontal kick

3.2 Coupling measurements from mode frequencies

Solving the equations of motion for a difference resonance revealed the existence of two modes at frequencies ω_1 and ω_2 (35). The positions x and z are given by a mixture of these modes, but it is possible to rotate the initial axes until the modes are decoupled. These inclined modes are termed normal modes. According to the first equation (30) that gives the frequencies of these modes when λ_x and λ_z are neglected, their wave numbers are separated by $\eta/2$, i.e.

$$Q_{n,m,1} = Q_x - \frac{\Delta}{2} + \frac{1}{2}\sqrt{\Delta^2 + |C^-|^2}$$
$$Q_{n,m,2} = Q_z + \frac{\Delta}{2} - \frac{1}{2}\sqrt{\Delta^2 + |C^-|^2} \qquad (42)$$

where only the fractional parts of the tunes are included. When $|C^-| = 0$, the normal mode tunes are equal to the unperturbed betatron tunes Q_x and Q_z . When $|C^-| \neq 0$, the effects on the normal mode tunes depend on Δ (= Q_x - Q_z). If Δ is large with respect to $|C^-|$, there is little change of the tunes, but when Δ is small the impact of coupling increases, to reach a maximum at $\Delta = 0$ (tune split equal to $|C^-|$). This frequency split of the normal modes is associated with the above mentioned rotation of the normal axes and sometime an elliptical polarization if $C_2 \neq 0$.

These effects can be observed experimentally. Increasing Q_x and reducing Q_z in the vicinity of the coupling resonance, while also measuring the tunes with horizontal and vertical

kickers and pick-ups will result in curves similar to those of Fig. 3. The pick-ups see indeed the frequencies of the normal modes 1 and 2 and one can distinguish three typical situations:

a) $|C^-| < \Delta$; the modes are nearly vertical and horizontal and the pick-up measurements are reliable.

b) $|C^-| \approx \Delta$; difficulties in measuring the tunes begin to appear and tune readings start to jump back and forth from mode 1 to mode 2. Two lines are visible on each (horizontal or vertical) pick-up.

c) $|C^-| > \Delta$; tune readings are spanning the stopband visible in Fig. 3. At $\Delta \approx 0$, the pick-ups respond equally well to both modes, that are now at 45°.

This shows that the difference in normal mode frequencies is exactly equal to $|C^-|$ when $\Delta = 0$ and that measuring the tunes while scanning the resonance vicinity provides another way of measuring $|C^-|$.

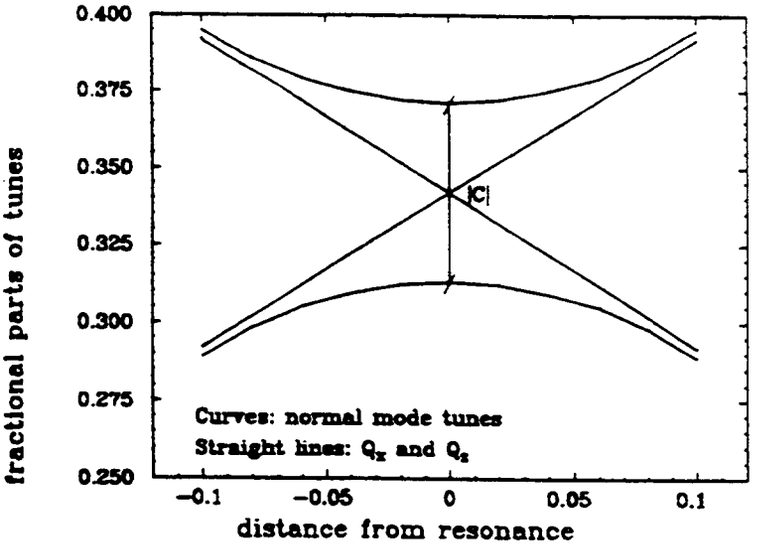


Fig. 3 Coupled tunes as functions of the distance from the difference resonance

4. AMPLITUDE VARIATION DUE TO RADIATION AND ACCELERATION

4.1 General expression of the amplitude variation

Considering the expression (31) for a difference resonance and including now the conjugate momenta $p_x = \dot{x} + Sz/2$ and $p_z = \dot{z} - Sx/2$, the entire solution for the vector $Y = (x, p_x, z, p_z)$ can be written as follows in presence of linear coupling

$$Y_j = \sum_{k=1}^{4} w_{jk}(\theta) A_k \qquad j = 1,\ldots 4 \tag{43}$$

where j is numbering the four components of Y,

$$A_1 = B_1 \ , \quad A_2 = \overline{B}_1 \ , \quad A_3 = B_2 \ \text{ and } \ A_4 = \overline{B}_2 \ ,$$

and the functions w are,

$$
\begin{aligned}
w_{\substack{11\\13}} &= \frac{\overline{C}^{-}}{2\omega_{\substack{2\\1}}}\sqrt{\frac{\beta_x}{2R}}\; e^{i\left(\mu_x+\omega_{\substack{2\\1}}\theta\right)} \\
w_{\substack{21\\23}} &= \frac{\overline{C}^{-}}{2\omega_{\substack{2\\1}}}\sqrt{\frac{R}{2\beta_x}}(i-\alpha_x)\; e^{i\left(\mu_x+\omega_{\substack{2\\1}}\theta\right)} \\
w_{\substack{31\\33}} &= \sqrt{\frac{\beta_z}{2R}}\; e^{i\left(\mu_z-\omega_{\substack{1\\2}}\theta\right)} \\
w_{\substack{41\\43}} &= \sqrt{\frac{R}{2\beta_z}}(i-\alpha_z)\; e^{i\left(\mu_z-\omega_{\substack{1\\2}}\theta\right)} \\
w_{j2} &= \overline{w}_{j1}\ , \quad w_{j4} = \overline{w}_{j3}
\end{aligned}
\tag{44}
$$

The two different subscripts of the w's are associated with the two indices of ω and the form of the functions (44) implies that the following relations hold

$$w_{jk}(\theta+2\pi) = w_{jk}(\theta)\; e^{i2\pi\lambda_{jk}} \tag{45}$$

with

$$\lambda_{j2} = -\lambda_{j1} \quad \text{and} \quad \lambda_{j4} = -\lambda_{j3}\ .$$

These solutions are strictly valid in the absence of radiation, but coupled betatron oscillations of leptons are in turn enhanced by photon emission and damped by the longitudinal acceleration as well as by the average energy loss in the presence of the focusing component of the magnetic field.

Starting from the constants A_k of the coupled motion, we now look for their variations due to these effects [10]. *Assuming* that these constants change slowly with respect to the quantum fluctuations and the period of the coupled betatron oscillations, they will reach an equilibrium between excitation and damping after a few damping times. It is precisely these equilibrium values of A_k which we want to derive in the following sections.

It follows from Eq. (44) that the quantity F defined below is an invariant of the motion:

$$F\left(w_{jl}, w_{jk}\right) = w_{1l}\ w_{2k} - w_{2l}\ w_{1k} + w_{3l}\ w_{4k} - w_{4l}\ w_{3k}\ . \tag{46}$$

As a consequence of the invariance of F and of the properties (45), we can write:

$$F\left[w_{jl}(\theta), w_{jk}(\theta)\right] = F\left[w_{jl}(\theta+2\pi), w_{jk}(\theta+2\pi)\right] = \exp\left[i2\pi\left(\lambda_{jl}+\lambda_{jk}\right)\right]F\left[w_{jl}(\theta), w_{jk}(\theta)\right] \tag{47}$$

and the equality between the first and last term of Eq. (47) (which can only be satisfied if the exponential is equal to 1) induces the following orthonormality :

$$F\left(w_{jl},w_{jk}\right)=1 \quad \text{if} \quad w_{jl}=\overline{w}_{jk}$$
$$F\left(w_{jl},w_{jk}\right)=0 \quad \text{if} \quad w_{jl}\neq\overline{w}_{jk} \ . \tag{48}$$

This property of orthonormality can now be used to solve Eq. (43) for the A_k's. Let us write indeed:

$$F\left(Y_j,w_{jk}\right)=\sum_{l=1}^{4}A_l F\left(w_{jl},w_{jk}\right)=\overline{A}_k \ F\left(\overline{w}_{jk},w_{jk}\right) \ , \tag{49}$$

making use of Eqs. (43) and (48). Then we derive from the last equality :

$$A_k=\frac{F\left(Y_j,\overline{w}_{jk}\right)}{F\left(w_{jk},\overline{w}_{jk}\right)} \ . \tag{50}$$

Since we want to look for the variations of the A_k's due to changes in the canonical variables (vector Y), Eq. (50) is the key equation. In general, any variation of $|A_k|$ due to trajectory changes δY_j can be expressed as follows :

$$\delta\left|A_k\right|^2=\overline{A}_k \ \ \delta A_k+A_k\delta\overline{A}_k+\left|\delta A_k\right|^2=\frac{1}{F\left(w_{jk},\overline{w}_{jk}\right)}$$
$$\left\{\overline{A}_k F\left(\delta Y_j,\overline{w}_{jk}\right)-A_k F\left(\delta Y_j,w_{jk}\right)-\frac{1}{F\left(w_{jk},\overline{w}_{jk}\right)}F\left(\delta Y_j,\overline{w}_{jk}\right)F\left(\delta Y_j,w_{jk}\right)\right\} \ . \tag{51}$$

4.2 Application to photon emission and acceleration

Considering the photon emission, it is well known [11] that the equilibrium orbit and the betatron variables are changed by a quantity which is proportional to the photon energy ε and to the dispersion $\boldsymbol{D}$, whose components are $\left(D_x,\dot{D}_x,D_z,\dot{D}_z\right)$:

$$\delta\boldsymbol{Y}=\frac{\varepsilon}{E_0}\boldsymbol{D} \ . \tag{52}$$

In the presence of longitudinal acceleration δE in a cavity, only the transverse momenta are changed while the transverse coordinates remain constant:

$$\delta\boldsymbol{Y}=\left(\delta x,\delta p_x,\delta z,\delta p_z\right)=\left(0,-p_x,0,-p_z\right)\frac{\delta E}{E_0} \ , \tag{53}$$

E_0 being the nominal energy.

Let us first deal with the orbit change due to photon emission. Since $\delta\boldsymbol{Y}$ is proportional to ε (Eq. (52)) and since δY_j appears linearly as well as quadratically in Eq. (51), it is necessary to evaluate the average $<\varepsilon>$ and the mean square $<\varepsilon^2>$ of the quantum emission, over a time interval Δt, and to multiply both by the mean emission rate N. We will now consider successively these two terms:

a) If P_γ stands for the rate of loss of energy by radiation, we have

$$N\ \Delta t < \varepsilon >= \frac{1}{c} P_\gamma \Delta \ell \ , \tag{54}$$

where $\Delta\ell$ is the path length interval and

$$\begin{aligned} P_\gamma &= \frac{2}{3} r_e c\ \ \gamma^3\ \ E \frac{1}{\rho^2} \\ \Delta \ell &= \Delta s \left(1 + \frac{x}{\rho_x} + \frac{z}{\rho_z} \right) . \end{aligned} \tag{55}$$

The quantity $1/\rho^2$ in P_γ is proportional to the square of the field B^2. Taking into account the possible presence of field gradients, we must develop B^2 in a series for x and z. Keeping only first order terms,

$$\frac{B^2}{B_0^2} = 1 + G_x x + G_z z \ , \tag{56}$$

with

$$G_x = \frac{2}{B_0^2} \left(B_z \frac{\partial B_z}{\partial x} - B_x \frac{\partial B_z}{\partial z} \right) \ \text{ and } \ G_z = \frac{2}{B_0^2} \left(B_x \frac{\partial B_z}{\partial x} + B_z \frac{\partial B_z}{\partial z} \right) .$$

Regrouping all the first order terms in x and z which appear in the product $P_\gamma \Delta\ell$, we obtain:

$$\frac{1}{c} P_\gamma \Delta \ell = \frac{1}{c} P_{\gamma 0} \Delta s \left(1 + C_x x + C_z z \right) \ , \tag{57}$$

with $C_x = 1/\rho_x + G_x$ and $C_z = 1/\rho_z + G_z$. On the right hand side, $P_{\gamma 0}$ is calculated on the central trajectory $x = z = 0$ with nominal field B_0.

Putting all these results together and replacing x and z by their development in eigenfunctions (Eq. (43)), the expression we are looking for is:

$$N \Delta t < \varepsilon >= \Delta t\ \ P_{\gamma 0} \left[1 + \sum_{k=1}^{4} A_k \left(C_x w_{1k} + C_z w_{3k} \right) \right] . \tag{58}$$

b) If Q_ε stands for the mean value of the product $N<\varepsilon^2>$, we have simply:

$$N \Delta t < \varepsilon^2 >= \Delta t Q_\varepsilon \tag{59}$$

with

$$Q_\varepsilon = \frac{55}{24\sqrt{3}} r_e h c^2 \gamma^6 E \frac{1}{\rho^3} \ . \tag{60}$$

In the expression (51) for the amplitude variation, we still need to evaluate the function $F(\delta Y_j, w_{jk}) = \varepsilon F(D_j / E_0, w_{jk})$, and the similar one $F(\delta Y_j, \overline{w}_{jk})$. This is simple if we make use of the definition (46):

$$F(D_j / E_0, w_{jk}) = \frac{1}{E_0}\left(D_x w_{2k} - \dot{D}_x w_{1k} + D_z w_{4k} - \dot{D}_z w_{3k}\right) . \tag{61}$$

Let us now turn to the question of longitudinal acceleration δE (Eq. (53)). For the same reasons previously evoked, linear and quadratic terms in δE will be present in Eq. (51). If the quadratic term δE^2 becomes negligible towards the limit $\Delta t \to 0$, the linear term averages to (using eigenfunctions again):

$$\delta p_y = -\frac{<\delta E>}{E_0} p_y = -\frac{<\delta E>}{E_0} \sum_{k=1}^{4} w_{\substack{2k \\ 4k}}(\theta) A_k \ . \tag{62}$$

Introducing Eq. (62) in the linear terms of Eq. (51) and assuming logically that the average $<\delta E>$ must exactly compensate for the radiation loss, gives:

$$\delta |A_k|^2_{\text{accel.}} = -\frac{P_{\gamma 0} \Delta t}{E_0} \left| A_{k^2} \right| . \tag{63}$$

Putting together Eqs. (58), (59) and (61) into Eq. (51) for the photon emission effect and adding the contribution (63) of the acceleration, we can derive the following expression for the amplitude variation:

$$\begin{aligned} <\delta |A_k|^2> &= -\frac{P_\gamma \Delta t}{E_0} \left| A_k^2 \right| \\ &-\frac{P_\gamma \Delta t}{E_0} \frac{\left| A_k^2 \right|}{F(w_{jk}, \overline{w}_{jk})} 2i < \text{Im}\left[\left(C_x \overline{w}_{1k} + C_z \overline{w}_{3k} \right) \left(D_x w_{2k} - \dot{D}_x w_{1k} + D_z w_{4k} - \dot{D}_z w_{3k} \right) \right] > \\ &-\frac{Q_\varepsilon \Delta t}{E_0^2 F^2 (w_{jk}, \overline{w}_{jk})} < \left| D_x w_{2k} - \dot{D}_x w_{1k} + D_z w_{4k} - \dot{D}_z w_{3k} \right|^2 > \ . \end{aligned} \tag{64}$$

The equality $\overline{A} F(\delta Y, \overline{w}) - A F(\delta Y, w) = -2i \ \text{Im}\left[\overline{A} F(\delta Y, \overline{w}) \right]$ has been applied and the subscript 0 of P_γ abandonned for simplicity. The three terms of (64) give respectively the amplitude variation associated with the acceleration, radiation damping and quantum excitation.

4.3 Equilibrium amplitudes with coupling and radiation

The finite amplitude variations with radiation for a finite time interval Δt are written explicitly in Eq. (64). On the limit of infinitesimal interval ($\Delta t \to dt$), Eq. (64) becomes with the usual notation [11]:

$$\frac{d|A_k|^2}{dt} = -2\alpha_k \left| A_k^2 \right| + Q_k \ , \tag{65}$$

where α_k are the damping coefficients

$$\alpha_k = < \frac{P_\gamma}{2E_0} > J_k \ , \tag{66}$$

which are proportional to the damping partition numbers J_k

$$J_k = 1 + < \frac{\mathrm{Im}\left[\left(C_x \overline{w}_{1k} + C_z \overline{w}_{3k} \right) \left(D_x w_{2k} - \dot{D}_x w_{1k} + D_z w_{4k} - \dot{D}_z w_{3k} \right) \right]}{\mathrm{Im}\left(w_{1k} \overline{w}_{2k} + w_{3k} \overline{w}_{4k} \right)} > \ , \tag{67}$$

and where Q_k are the transverse beam amplitude coefficients

$$Q_k = < \frac{1}{4} \frac{Q_\varepsilon}{E_0^2} >< \frac{\left| D_x w_{2k} - \dot{D}_x w_{1k} + D_z w_{4k} - \dot{D}_z w_{3k} \right|^2}{\mathrm{Im}^2 \left(w_{1k} \overline{w}_{2k} + w_{3k} \overline{w}_{4k} \right)} > \ . \tag{68}$$

These relations use the identity $F\left(w_{jk}, \overline{w}_{jk} \right) = 2i \ \mathrm{Im}\left(w_{1k} \overline{w}_{2k} + w_{3k} \overline{w}_{4k} \right)$.

Within the assumption made in Section 4.1, a stationary state will occur after a few damping times and it corresponds to the condition $d\left| A_k \right|^2 / dt = 0$. Hence, the equilibrium amplitudes are (Eq. (65)):

$$\left| A_k^2 \right| = \frac{Q_k}{2\alpha_k} \ , \tag{69}$$

and this is the important result to be used in Eq. (43). Betatron coupling is present through the eigenfunctions (Eq. (44)), the fact that both vertical and horizontal dispersions have an influence and the need of four coefficients (k = 1 to 4) in order to describe the whole motion.

4.4 Equilibrium emittances with coupling and radiation

Betatron oscillations are characterized by the transverse invariants of the motion, which define the commonly used emittances E_y. If by definition E_y represents the invariant mean-square amplitudes of the transverse oscillations, we must have:

$$E_y = \frac{< y^2 >}{\beta_y} \ . \tag{70a}$$

Starting from the solution (43) and using the eigenfunctions (44) as well as the equalities between complex conjugates, we can rewrite Eq. (70a) in the following manner:

$$E_y = \frac{< y^2 >}{\beta_y} = \frac{2}{\beta_y} \left(\left| A_1^2 \right| \underset{31}{\left| w_{11} \right|^2} + \left| A_3^2 \right| \underset{33}{\left| w_{13} \right|^2} \right) , \tag{70b}$$

the two subscripts of the w's being associated with the horizontal and vertical coordinates, respectively.

In order to simplify the next calculations, let us now *assume* that the accelerator or storage ring of interest is large and has separated functions. This means that the radius of curvature ρ is large and there is no gradient in the dipoles ($C_x \equiv C_z \equiv 0$).

Consequently, $<D/\rho>$ is small with respect to 1 and Eq. (66) becomes simply:

$$\alpha_k = < \frac{P_\gamma}{2E_0} > \quad (\text{with } J_k = 1) \ . \tag{71}$$

Hence, putting together Eqs. (68) to (71) and introducing the explicit forms (44) of the eigenfunctions make it possible to write the emittances as follows:

$$\begin{aligned} E_x &= < \frac{Q_\varepsilon}{RE_0 P_\gamma} > \left[\frac{4\omega_2^2 \left|C^-\right|^2}{\left(4\omega_2^2 + \left|C^-\right|^2\right)^2} < H_1 > + \frac{4\omega_1^2 \left|C^-\right|^2}{\left(4\omega_1^2 + \left|C^-\right|^2\right)^2} < H_3 > \right] \\ E_z &= < \frac{Q_\varepsilon}{RE_0 P_\gamma} > \left[\frac{16\omega_2^4}{\left(4\omega_2^2 + \left|C^-\right|^2\right)^2} < H_1 > + \frac{16\omega_1^4}{\left(4\omega_1^2 + \left|C^-\right|^2\right)^2} < H_3 > \right] , \end{aligned} \tag{72}$$

where P_γ and Q_ε are given by Eqs. (55) and (60), respectively. It remains to define the functions H_1 and H_3 which are simply the numerator of the second bracket in Eq. (68) of Q_1 and Q_3:

$$\begin{aligned} H_{\substack{1\\3}} &= \frac{\left|C^-\right|^2}{\omega_{\substack{2\\1}}^2} \frac{R}{8\beta_x} \left[D_x^2 + \frac{1}{R^2} \left(\beta_x \dot{D}_x + R\alpha_x D_x\right)^2 \right] + \frac{R}{2\beta_z} \left[D_z^2 + \frac{1}{R^2} \left(\beta_z \dot{D}_z + R\alpha_z D_z\right)^2 \right] \\ &+ \frac{R}{2\omega_{\substack{2\\1}} \sqrt{\beta_x \beta_z}} \mathrm{Re}\left[\overline{C}^- (\alpha_x - i)(\alpha_z + i) \exp(i\phi) \right] D_x D_z + \frac{\sqrt{\beta_x \beta_z}}{2\omega_{\substack{2\\1}} R} \mathrm{Re}[\overline{C}^- \exp(i\phi) \dot{D}_x \dot{D}_z + \\ &+ \frac{1}{2\omega_{\substack{2\\1}}} \sqrt{\frac{\beta_z}{\beta_x}} \mathrm{Re}\left[\overline{C}^- (\alpha_x - i) \exp(i\phi) \right] D_x \dot{D}_z + \frac{1}{2\omega_{\substack{2\\1}}} \sqrt{\frac{\beta_x}{\beta_z}} \mathrm{Re}[\overline{C}^- (\alpha_z - i) \exp(i\phi)] \dot{D}_x D_z \ . \end{aligned} \tag{73}$$

The subscripts of H are associated with the two indices of ω (see Eq. (31)), and the phase ϕ is written for:

$$\phi = \phi_x - \phi_z - \theta\Delta \ . \tag{74}$$

The two first terms of (73) are directly proportional to the dispersion invariants (named I_X and I_Z), which appear naturally in E_X and E_Z respectively [11], in the absence of coupling. The other terms are obviously coupled terms for the dispersion.

Turning back to the expressions of the equilibrium emittances (72) and (73), which appear to be fairly complicated, let us look at two borderline cases:

a) For vanishing linear coupling $\left(C^- \to 0\right)$, the four terms in the square brackets of Eq. (72) have finite limits. Two of them are equal to zero, while the two remaining ones become equivalent to:

$$\frac{4\omega_2^2\left|C^-\right|^2}{\left(4\omega_2^2+\left|C^-\right|^2\right)^2}<H_1>=\frac{R}{2}<I_x>\ ;\quad \frac{16\omega_1^4}{\left(4\omega_1^2+\left|C^-\right|^2\right)^2}<H_3>=\frac{R}{2}<I_z>\ . \tag{75}$$

In this case, the transverse equilibrium emittances and their ratio $g = E_Z/E_X$ are simply given by:

$$\begin{aligned} E_y &=<\frac{Q_\varepsilon}{2E_0P_\gamma}><I_y> \\ g &=\frac{<I_z>}{<I_x>}\ . \end{aligned} \tag{76}$$

As expected, the vertical emittance is nul, if the vertical dispersion vanishes in addition to C^-.

b) For vanishing vertical dispersion (but $C \neq 0$), the functions (73) can be written as follows:

$$H_{\substack{1\\3}}=\frac{\left|C^-\right|^2}{\omega_{\substack{2\\1}}^2}\frac{R}{8}<I_x>\ . \tag{77}$$

Introducing Eq. (77) into Eq. (72) gives for the emittances and their ratio g:

$$\begin{aligned} E_x &=<\frac{Q_\varepsilon I_x}{2E_0P_\gamma}>\frac{\frac{1}{2}\left(\left|C^-\right|/\Delta\right)^2+1}{\left(\left|C^-\right|/\Delta\right)^2+1} \\ E_z &=<\frac{Q_\varepsilon I_x}{2E_0P_\gamma}>\frac{\frac{1}{2}\left(\left|C^-\right|/\Delta\right)^2}{\left(\left|C^-\right|/\Delta\right)^2+1} \\ g &=\frac{\left(\left|C^-\right|/\Delta\right)^2}{\left(\left|C^-\right|/\Delta\right)^2+2} \end{aligned} \tag{78}$$

The corresponding curves for E_X and g are plotted as functions of the ratio $\left|C^-\right|/2\Delta$ in Fig. 4. Both Eq. (78) and Fig. 4 show that on the limit $\left|C^-\right| >> \Delta$ (sometimes called full coupling) the transverse emittances are equal and take half the value of the horizontal emittance at $C^- = 0$.

In general, coupling and vertical dispersion are not vanishing, so that not only I_X and I_Z are contributing but also the products D_xD_z, $\dot{D}_x\dot{D}_z$, $D_x\dot{D}_z$ and $\dot{D}_xD_z$ in agreement with expression (73).

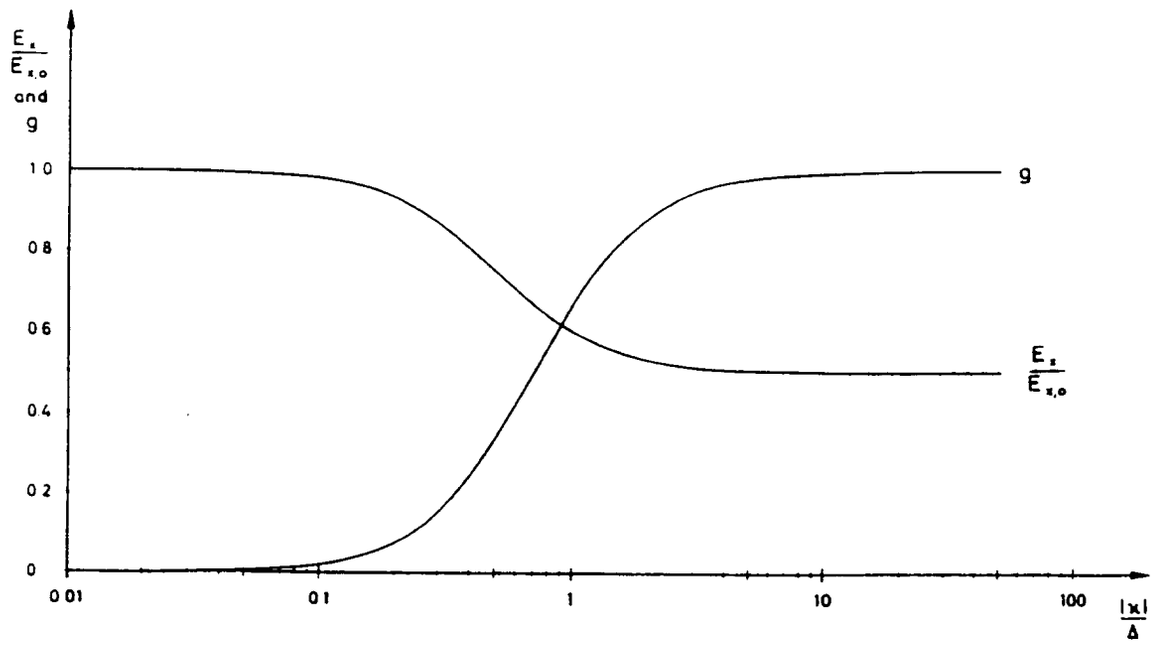


Fig. 4 Horizontal emittance and emittance ratio as function of $|C^-|/2\Delta$, with vanishing vertical dispersion

5. APPLICATION TO LEP BETATRON COUPLING

In a circular collider like LEP [12] with experiments installed around each interaction point, it is important to compensate for the linear coupling due to experimental solenoids [13] on the one side and the unavoidable sources of residual imperfections all around the ring on the other side. Coupling would indeed not only generate beating of the β-function and tilting of the beam at crossing points, but also modify the nominal equilibrium emittances according to the mechanism described above, which in turn would influence the performance. Moreover, performance optimization requires an adjustment of the emittance ratio $g = E_z/E_x$, which can be achieved through variations of the coupling coefficient C^- (Section 4.4) using a few specific tilted quadrupoles.

5.1 Emittance control with linear coupling

The performance of LEP at given beam current (limited by other mechanisms) is optimum when the beam-beam forces are such that the horizontal and the vertical beam-beam tune shifts are equal (and below a threshold of the order of 0.03 at collision). This implies the following relations between the emittances, the β*-functions and beam sizes σ* at the crossing point

$$\frac{E_x}{E_z} = \frac{\beta_x^*}{\beta_z^*} = \frac{\sigma_x^*}{\sigma_z^*} \ . \tag{79}$$

The β^*-ratio associated with the design of the experimental insertions (nominal value of 25 in LEP) implies the use of flat beams and small values of g (4%). The expressions for the emittances in the presence of betatron coupling and radiation (Section 4.4) allow an estimation of the range in which g can be controlled. The assumptions made in Section 4 are fully valid for LEP which has tune values near a difference resonance (initial nominal tunes were Q_x = 70.4 and Q_z = 78.3, with a difference of 8 in the integers selected for beam-beam effect

optimization), a very large radius of curvature and a lattice with separated functions (i.e. $C_x = C_z = 0$). If the horizontal dispersion invariant $<I_x>$ has been estimated [13] to be equal to 1.75 10^{-2} m, the vertical one $<I_z>$ is of the order of 1.8 10^{-4} m for a residual dispersion of 10 cm approximately (as expected in the best conditions), 11.2 10^{-4} m for D_z of 25 cm and 28.8 10^{-4} m for D_z of 40 cm.

In these conditions, the theory developed above allows to estimate the minimum emittance ratio [13] which can be reached for different amplitudes of the vertical dispersion. The best one can do consists of compensating exactly the linear coupling such as $C^- = 0$. Equations (76) then apply directly and give minimum emittance ratios g_{min} of 1%, 6.25% and 16%, for vertical dispersions of 10 cm, 25 cm and 40 cm, respectively. The nominal conditions ($g = 4\%$) imply that D_z must remain below 20 cm approximately. To get $C^- \cong 0$ in LEP, there exist four tilted-quadrupole schemes located in sections where $D_x = 0$, designed for the compensation of the fields of the experimental solenoids present at the collision-points and for the control of the residual machine coupling [14].

In the other extreme case, the coupling coefficient can be so large that the contribution of the vertical dispersion to the emittance becomes negligible and the equations (78) apply. Using these equations and/or the curves of Fig. 4, one finds out that, for a distance to the difference resonance $\Delta = Q_x - Q_z + 8$ of 0.1, values of 0.1, 0.2, 0.3 and 0.5 for C^- correspond to emittance ratios g of 33%, 66%, 82% and 92% (all well above g_{min} associated with D_z only). The capacity of the LEP tilted-quadrupole schemes makes it possible to provide such high values of C^-, allowing either for full coupling if necessary or for compensation of contingently large C^- if required.

Looking at the numbers, the flat-beam configuration required for the LEP performance implies reducing the value of C^- below 0.01 approximately while keeping the residual amplitude of the vertical dispersion lower than ~ 10 cm. If the nominal value of $g = 4\%$ is exceeded at constant beam current and insertion optics, the loss of luminosity will rise with $\sqrt{g}$, about. Independently, a high value of C^- (> 0.1 say) prevents accurately running of the ring and its injection system, since the optics is perturbed (in particular, for tunes close to a difference resonance) and the diagnostics are confused.

Among the possible sources of coupling to be compensated, most evident was an abnormally large betatron coupling discovered during early LEP commissioning. It manifested itself by coupling the first-turn trajectories, tilting and blowing-up the beam and confusing tune measurements. All the necessary corrections were based on the formalism described above and can be used to illustrate its application.

5.2 Compensation of the ring linear coupling

Among the expected sources of linear coupling in LEP, there are three that have been considered as important and estimated in the design phase. Indeed, the random tilts of all quadrupoles of r.m.s. value $<\theta>$ and finite amplitudes of the vertical orbit (r.m.s. $<z>$) in the systematic sextupoles required for chromaticity correction generate coupling, estimated to be

$$\begin{aligned} |C^-| &\cong 0.009 \ , \quad \text{for} \quad <\theta> = 0.24 \text{ mrad} \\ |C^-| &\cong 0.012 \ , \quad \text{for} \quad <z> = 1.0 \text{ mm} \end{aligned} \tag{80}$$

on the nominal optics (tunes separated by 8). The third source was of course the experimental solenoids, of which the strongest field integral corresponds to $|C^-| \cong 0.06$. This result justified

the introduction of tilted quadrupoles near each interaction point (total of 8 per point) to reduce every solenoid contribution to below ~ 0.003.

There remain however sources due to field imperfections, initially considered as negligible, like field asymmetries in the magnets, earth field, induced current in a dissymmetric vacuum chamber and presence of magnetic material perturbing the field lines. One of those had to be responsible for the observed $|C^-|$ that was an order of magnitude higher than expected ($|C^-| \approx 0.3$) and corresponded to a systematic skew-gradient of ~ 2 Gm^{-1} in all arcs. Measurements ruled out the earth field as being the main contributor since the consequent skew-gradient estimated by including the shielding effect of the dipole is only about 0.15 Gm^{-1}. The importance of asymmetries has not been precisely quantified, but the presence of ferromagnetic nickel in the contact layer between the aluminium chamber and the lead shield was identified as the main source of the unexpected perturbations [15]. The remanent field of the nickel can be strong and its component in the horizontal faces of the chamber creates an undesirable skew-field (Fig. 5).

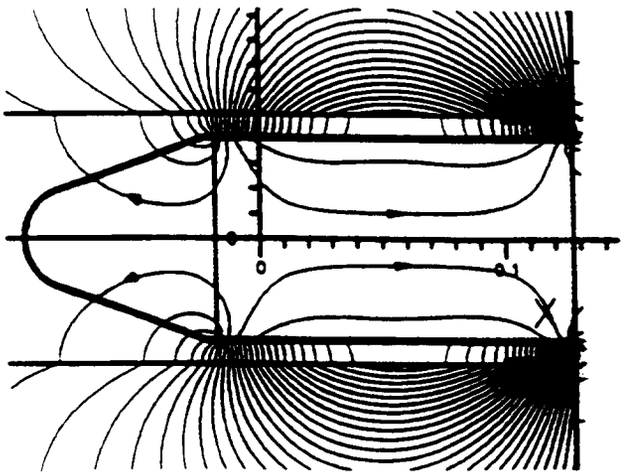


Fig. 5 Calculated field map in the LEP dipoles, due to nickel remanent field

In view of coupling compensation [16], it was first appropriate to re-optimize the linear optics of LEP in order to modify the tune-integer separation (nominally equal to 8), for the collider was difficult to control and the source of imperfections had the periodicity 8 of the arcs. With tune integers separated by 6 (Q_x = 71.4, Q_z= 77.3), the driving term (15) of the difference resonance was reduced by 5 approximately, but was still large

$$\left|C^-\right| = 0.058 \ , \quad \text{for} \quad \left|Q_x - Q_z\right| = 6 \ . \tag{81}$$

The next step consisted of trying to compensate linear coupling by using the tilted quadrupoles already installed near each of the four experimental areas. For the necessary solenoid compensation, there are four pairs (or families) around every even crossing point able to entirely decouple betatron motions outside the experimental sections and at the interaction position. The corresponding pairs of magnets are termed QT1 to QT4, the elements of a pair being symmetrically located with respect to the solenoid centre. Antisymmetrically powered, these elements generate an imaginary component C^-, while they create a real component when powered with the same sign. The two pairs QT2 and QT3 are antisymmetrically powered to

mainly compensate for the solenoids (imaginary C^-, see Section 5.3) while the magnets of QT1 and QT4 are independent in order to give means of compensating ring imperfections (mainly real C^-). The margin in their strengths, foreseen for running at ~ 100 GeV, made it possible to use them for correcting the strong systematic coupling due to the vacuum chamber at injection.

Since the working point of LEP is close to a difference resonance, the efforts for correcting the ring imperfections were focused on a reduction of C^-. Measurements basically concerned with the normal mode frequencies (Section 3.2), and predictions of corrections were based on estimates of C^- using Eqs. (15) and (16), as well as on numerical simulations.

The first successful compensation was obtained using the optics with $|Q_x - Q_z| = 6$ and the QT1 tilted quadrupoles [17]. It was based on the observation supported by numerical simulations that a second harmonic of skew-gradient correction had strong effects on coupling with this optics, for the source of the imperfections was mainly an harmonic 8. Subsidiarily, two arcs enclosing one crossing point and having approximately symmetrical errors generate a real component C^- at this point, as can be seen from formula (15) using the appropriate phases. Both arguments incited us to excite the QT1 according to the following pattern (e.g. L2 stands for left of point 2 and R2 for right of point 2),

Position	L2	R2	L4	R4	L6	R6	L8	R8
QT1 polarity	-	-	+	+	-	-	+	+

QT1 absolute normalised strength: 0.006 m^{-2} at 20 GeV.

This first "historical" compensation [16, 17] succeeded in decreasing $|C^-|$ by more than an order of magnitude, down to 0.001, as shown in Fig. 6 (with experimental solenoids switched off). The corresponding change in the beam aspect ratio, at positions with and without horizontal dispersion (top and bottom), can be seen in Fig. 7. This correction made the control of the machine much easier and physics runs successful. Later, compensation was achieved by using the QT4 quadrupoles, that are almost equivalent but the tuning of which do not depend on the betatron amplitudes at the interaction points. The compensation was also distributed in the arcs by adding pairs of small tilted quadrupoles near the centre of each arc. All these schemes were based on the treatment recalled in Section 2 [2].

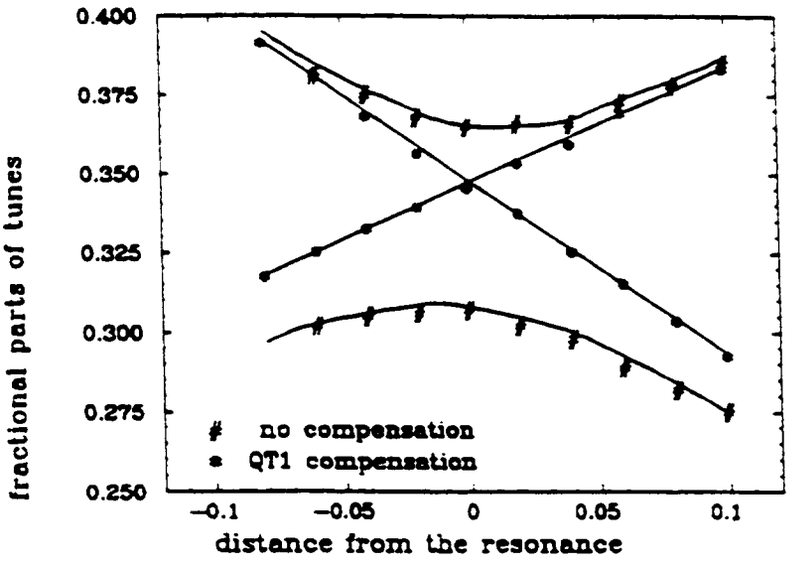


Fig. 6 Tune separation on the difference resonance, before and after the first compensation with the tilted quadrupoles QT1 of LEP

Given the limited resolution of the information conveyed by the luminescent screens (Fig. 7), the careful comparison of the two pictures at $D_x = 0$ and at finite D_x (assumed to be close to the nominal value) allowed a rough estimate of the actual emittance ratio. Hence, what seems to be full coupling on the bottom left picture before compensation, corresponds to an emittance ratio g of ~ 0.3. Similarly, the bottom right picture taken after compensation conceals a ratio about two times smaller (in agreement with the observed vertical dispersion of about 40 cm, as pointed out in Section 5.1). The ring compensation done as explained made possible further optical adjustments and correction of solenoid effects (Section 5.3), leading to a rise of luminosity (an order of magnitude, say). After reducing D_z by careful orbit corrections to about 20-25 cm, the emittance ratio approached the nominal 4% (Section 5.1) and the expected luminosity was within reach.

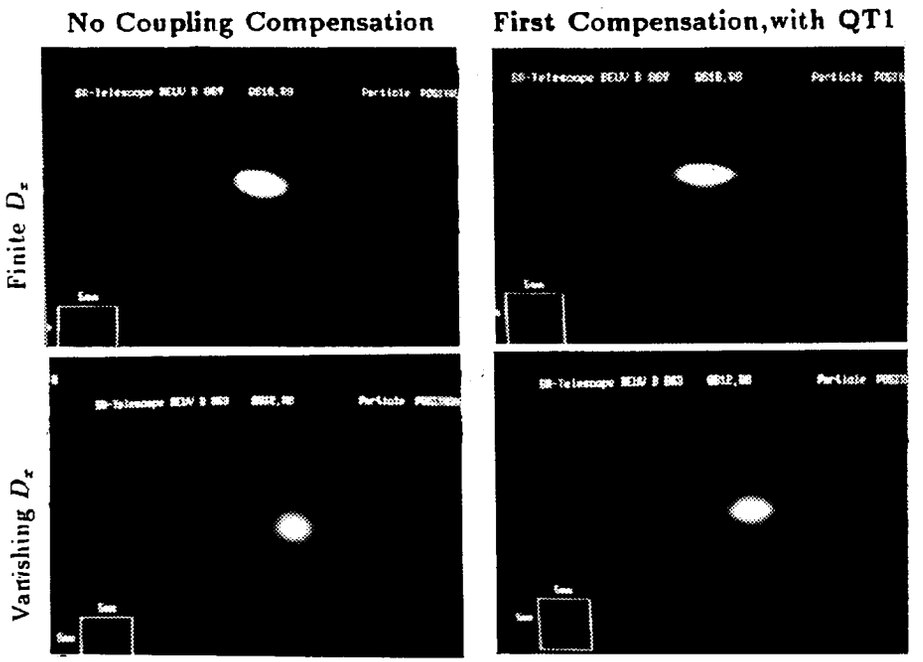


Fig. 7 Beam aspect from the light monitors placed at finite or vanishing dispersion, before and after the first coupling compensation in LEP

Even though the betatron coupling due to the magnetization of the nickel layer of the LEP dipole vacuum chamber has been much weakened as described, the source itself has been suppressed recently [15] by demagnetizing the nickel layer of the chamber, reducing the bare coupling by more than a factor five.

5.3 Compensation of solenoid linear coupling

Let us here differentiate between the coefficients arising from tilted quadrupoles and from solenoids by $C^{\pm}_{SQ}$ and $C^{\pm}_{SOL}$. Now, the requirement of full compensation of solenoidal effects with tilted quadrupoles may be formulated easily,

$$C^{\pm}_{SOL} + \sum C^{\pm}_{SQ} = 0 \ . \tag{82}$$

The summation has to be made over all tilted quadrupoles that are excited to compensate the solenoids. Hence, there are four linear equations since C^+ and C^- are complex and, in a thin-lens approximation for the tilted quadrupoles only, the system (82) may be written as follows [14, 18] by virtue of (16) and the last equation (15) (with $S = 0$),

$$\sum_j \frac{1}{2\pi R^2}\left[\sqrt{\beta_x\beta_z}K\ell\right]_j \cos\left(\mu_x - \mu_z - \frac{s}{R}\Delta^{\mp}\right)_j + C^{\mp}_{SOL,1} = 0$$
$$\sum_j \frac{1}{2\pi R^2}\left[\sqrt{\beta_x\beta_z}K\ell\right]_j \sin\left(\mu_x - \mu_z - \frac{s}{R}\Delta^{\mp}\right)_j + C^{\mp}_{SOL,2} = 0 \qquad (83)$$

where s is the quadrupole position and ℓ the quadrupole length. K is defined by (1). The two equations (83) correspond to the real and imaginary components of $C^{\mp}$ (index 1 and 2, respectively) and are valid for either the difference or the sum resonance. In the case of a solenoid with a pure longitudinal field and centred at the minimum of the betatron functions, the coupling coefficients can be derived analytically from Eq. (15) (with $K = 0$ and $S \neq 0$). The variations of the "Twiss functions" α_y, β_y, and μ_y in a drift space are known to be

$$\beta_y(s) = \beta_y^* + \frac{s^2}{\beta_y^*}, \quad \alpha_y = -\frac{s}{\beta_y^*}$$
$$\mu_y = \text{arctg}\,\frac{s}{\beta_y^*} \qquad (84)$$

where the star denotes the parameter values at the minimum of the betatron amplitude, chosen as the origin of s. Inserting these expressions in the integral (15) gives the simple solution [18]

$$C^{*\mp}_{SOL} = -i\frac{SL}{\pi R}\left[\sqrt{\frac{\beta_z^*}{\beta_x^*}} \mp \sqrt{\frac{\beta_x^*}{\beta_z^*}}\right]. \qquad (85)$$

where L is the length of the solenoid and S is defined by (2). Both vectors are purely imaginary in the chosen coordinate system. Their cancellation then imposes the use of pairs of skew-quadrupoles symmetrically placed with respect to the interaction point and having opposite currents, because C_{SQ} (left) is the complex conjugate of C_{SQ} (right). The presence of a real component due to ring imperfections or overlap of one solenoidal field with the next focusing quadrupoles (for example the L3 experiment in LEP) requires, however, the presence of two pairs of symmetrical quadrupoles with independent power supplies to allow currents of the same or opposite sign. All these considerations, the four conditions (83) linked to the expressions (85) and the particular values of the phases μ_y and functions β_y, induced us to choose the scheme shown in Fig. 8 for half an insertion [18]. It can be used to compensate either the solenoids or the ring, as explained in Section 5.1.

Let us underline here that this scheme makes it possible to decouple the motions at the solenoid centres, in order to avoid any distortion of the four-dimensional beam ellipsoid, that could reduce performance and luminosity. Moreover, the simultaneous compensation of C^+ and C^- outside the insertion of Fig. 8 and at the interaction point prevents the existence of tilted normal modes and β_y-beating in these positions and in the arcs [18]. Neglecting C^+ would indeed let a perturbation develop according to Eqs. (28), and it can be interpreted as a modulation of the β_y-function, considering the form of the complete solution that is similar to (31) by virtue of (A.19). This modulation is the more important the stronger is the tilt of the normal modes, i.e. the modifications (29) or (30) of the wave numbers. The effect is therefore a function of the ratio $|C^+|/\Delta^+$ and decreases with increasing Δ^+. It is however never completely negligible even when the working point is roughly centred between two sum resonance lines.

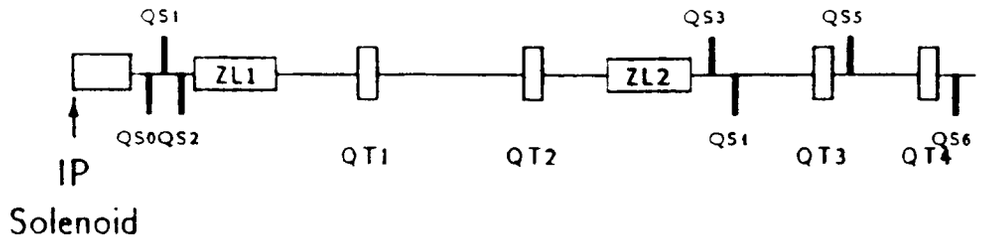


Fig. 8 Half LEP insertion with the experimental solenoid, the tilted quadrupoles (QT) for compensation, the focusing quadrupoles (QS) and the vertical electrostatic separators (ZL)

Only, the compensation of both C^+ and C^- avoid this imperfection and completely decoupled the transfer matrix across an experimental insertion. This fact was numerically tested [18] for one particular LEP solenoid, using the programs TRANSPORT and PETROS. TRANSPORT simulates the propagation of the beam ellipsoid throughout an insertion and PETROS simulates the beam dynamics of the whole machine with a compensated solenoid. TRANSPORT confirmed that the beam sizes at the crossing were not perturbed by more than 0.5% after compensation and PETROS showed that the tilt of normal modes was below 0.005° at the crossing and 0.04° in the arcs.

The four solenoids to be compensated in LEP are those of the experiments L3, ALEPH, OPAL and DELPHI, with difference coupling coefficients ranging from 0.015 to 0.06. Compensation with the scheme described was precalculated according to the present theory and initially set up. These corrections were successful in general, but some experimental adjustments were required some time to get a residual $|C|$ between 0.003 and 0.006 [14]. These might however correspond to a general minimization of the actual coupling around the ring, rather than to a local correction. The outcome was checked by measuring $|C^-|$ with the method of Section 3.2 and individual results obtained in the early commissioning phase are displayed in Fig. 9.

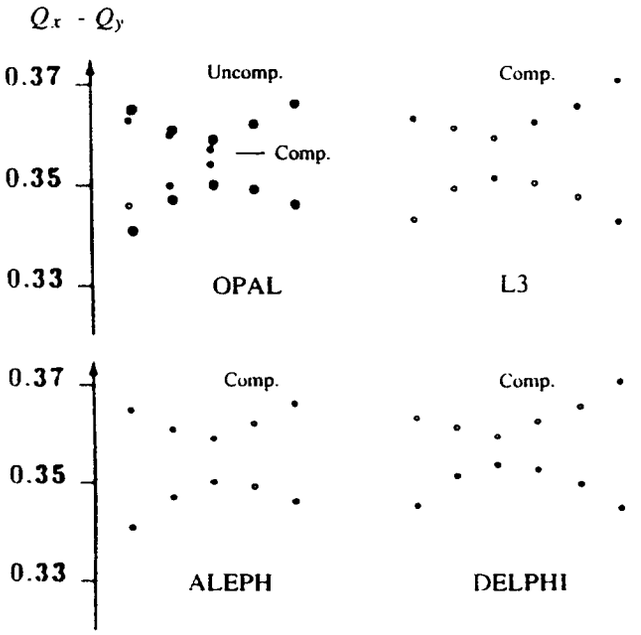


Fig. 9 Residual tune separations measured after initial compensation of each experimental magnet in LEP

APPENDIX

PERTURBATION THEORY

A.1 Treatment of the perturbation in Hamiltonian formalism [2]

The motion is defined by the total Hamiltonian H obtained by the sum of the Hamiltonian H_O of the unperturbed motion and of the perturbing Hamiltonian H_1, i.e.

$$H = H_o(p,q,\theta) + H_1(p,q,\theta) \tag{A.1}$$

where θ is the independent variable and p,q are the $2N$ canonical variables in an N-dimensional space,

$$(p,q) = \{p_\rho, q_\rho\} \quad \rho = 1,2,\ldots,N \ . \tag{A.2}$$

The canonical equations of the motion are then,

$$\begin{aligned} \dot{q}_\rho &= \frac{\partial H}{\partial p_\rho} = \frac{\partial H_o}{\partial p_\rho} + \frac{\partial H_1}{\partial p_\rho} \\ \dot{p}_\rho &= -\frac{\partial H}{\partial q_\rho} = -\frac{\partial H_o}{\partial q_\rho} - \frac{\partial H_1}{\partial q_\rho} \ . \end{aligned} \tag{A.3}$$

The perturbation problem consists in setting up a technique to find the motion for the total H, knowing the general solution of unperturbed motion,

$$\begin{aligned} q_\rho^{(0)} &= q_\rho^{(0)}(a_j,\theta) \\ & \qquad\qquad j = 1,2,\ldots,2N \ . \\ p_\rho^{(0)} &= p_\rho^{(0)}(a_j,\theta) \end{aligned} \tag{A.4}$$

and without making any approximations based on the smallness of H_1. The $2N$ arbitrary constants a_j are constant along each of the unperturbed trajectories. Solving Eqs. (A.4) gives

$$a_j = a_j(p_\rho, q_\rho, \theta) \tag{A.5}$$

these functions being determined by the form of H_O. They form a system of coordinates in the phase space at $\theta = 0$ and the unperturbed trajectories (A.4) a system of projection lines Γ_O from this initial-condition space into the general q-θ-p space, that includes the θ-dimension. Consequently, the perturbed motion in q-θ-p can be followed by a moving projection line Γ_O, which in turn corresponds to a set of a_j-values that change with θ. Therefore, the method is called the method of the variation of constants and is based on the principle that the general solution of the perturbed motion keeps the form (A.4) with time-dependent a_j however, i.e.

$$\begin{aligned} q_\rho &= q_\rho^{(0)}[a_j(\theta),\theta] \\ p_\rho &= p_\rho^{(0)}[a_j(\theta),\theta] \end{aligned} \tag{A.6}$$

Using the definition of the Poisson bracket

$$[u,v]=\sum_{\rho}\left[\frac{\partial u}{\partial q_\rho}\frac{\partial v}{\partial p_\rho}-\frac{\partial v}{\partial q_\rho}\frac{\partial u}{\partial p_\rho}\right] \tag{A.7}$$

the rate of change of any function $f(q_\rho,p_\rho,\theta)$, when the representative point moves along a trajectory, is given by

$$\frac{\mathrm{d}f}{\mathrm{d}\theta}=\frac{\partial f}{\partial\theta}+\sum_{\rho}\left[\frac{\partial f}{\partial q_\rho}\dot{q}_\rho+\frac{\partial f}{\partial p_\rho}\dot{p}_\rho\right]=\frac{\partial f}{\partial\theta}+[f,H] \tag{A.8}$$

by virtue of the form of the canonical equations (A.3). Applying Eq. (A.8) to the coefficients a_j provides

$$\begin{aligned}\dot{a}_j&=\frac{\partial a_j}{\partial\theta}+\left[a_j,H_o+H_1\right]\\ 0&=\frac{\partial a_j}{\partial\theta}+\left[a_j,H_o\right]\end{aligned} \tag{A.9}$$

for the perturbed and unperturbed motions, respectively.

By subtraction, one gets in the presence of perturbation

$$\dot{a}_j=\left[a_j,H_1\right]. \tag{A.10}$$

The right-hand side of Eq. (A.10) is a function of a_j and θ, because of (A.4) and so we have a set of $2N$-equations to determine the functions $a_j=f_j(\theta)$.

Expressing H_1 as a function of a_j and θ, termed U in what follows, it is possible to rewrite the equations (A.10) as:

$$\begin{aligned}H_1\left(p_\rho,q_\rho,\theta\right)&=U\left(a_j,\theta\right)\\ \dot{a}_j&=\sum_m\left[a_j,a_m\right]\frac{\partial U}{\partial a_m}\end{aligned} \tag{A.11}$$

making use of the definition (A.7).

So far *everything is exact.* Hence, to solve the perturbed motion knowing the solution of the unperturbed motion, it is necessary to find H, i.e. H_1, to express it as a function of the constants a_j, and to derive the $2N$ first-order equations (A.11).

A.2 Specific case of betatron motion [1,2]

In this case, the unperturbed motion is the motion of the particles, which are in the presence of magnetic fields linearly varying with x and z, and which are therefore stable. The well-known form of the corresponding Hamiltonian H_o is recalled here,

$$H_o=\frac{1}{2}\left[K_1(\theta)x^2+K_2(\theta)z^2+p_x^2+p_z^2\right] \tag{A.12}$$

where y denotes either x or z (for two-dimensional motion) and $K1, K2$ are the transverse focusing forces depending on θ

$$K_{1,2} = \frac{R^2}{B\rho} G_{1,2}(\theta) \ . \tag{A.13}$$

The functions $G_{1,2}$ are field gradients and the magnetic rigidity $B\rho$ is equal to the momentum divided by the charge, p/e. Canonical equations in either transverse direction are,

$$\dot{y} = p_y \quad \text{and} \quad \dot{p}_y + K_\rho(\theta) y = 0 \qquad \rho = 1,2 \ . \tag{A.14}$$

These are Hill's equations which always have two particular integrals of the following form, according to the theorem of Floquet [19],

$$\begin{aligned} y(\theta) &= a_\rho \ e^{iQ_y\theta} w(\theta) \\ y^*(\theta) &= \bar{a}_\rho \ e^{-iQ_y\theta} \bar{w}(\theta) \end{aligned} \tag{A.15}$$

using the complex notation for oscillatory motion. On the other hand, it has been shown by Courant and Snyder [20] that solutions of (A.14) may be written as follows:

$$\begin{aligned} y(\theta) &= a_\rho \ \sqrt{\frac{\beta_y(\theta)}{2R}} \exp\left[i \ \mu_y(\theta)\right] \\ \bar{y}(\theta) &= \text{complex conjugate} \end{aligned} \tag{A.16}$$

where $\beta_y(\theta)$ is the familiar transverse betatron amplitude. The integrals of the reciprocal of β_y functions define the phase advances μ_y and the wave numbers (tunes) Q_y,

$$\mu_y(\theta) = \int_o^\theta \frac{R}{\beta_y(\xi)} d\xi \ , \quad Q_y = \frac{1}{2\pi} \int_o^{2\pi} \frac{R}{\beta_y(\xi)} d\xi \ . \tag{A.17}$$

The direct comparison of Eqs. (A.15) and (A.16) provides Floquet's function w (and its complex conjugate)

$$w = \sqrt{\frac{\beta_y(\theta)}{2R}} \exp\left[i \ \left(\mu_y - Q_y\theta\right)\right] \ . \tag{A.18}$$

Let us denote by u the horizontal Floquet function and by v the vertical Floquet function, which are both given by (A.18) with y equal to x or z, respectively. Thus, the general solutions of the unperturbed motion equations (A.14) are

$$\begin{array}{lll} x = a_1 u & e^{iQ_x\theta} + \bar{a}_1 \bar{u} & e^{-iQ_x\theta} \\ p_x = a_1\left(\dot{u} + iQ_x u\right) & e^{iQ_x\theta} + \bar{a}_1\left(\dot{\bar{u}} - iQ_x \bar{u}\right) & e^{-iQ_x\theta} \\ z = a_2 v & e^{iQ_z\theta} + \bar{a}_2 \bar{v} & e^{-iQ_z\theta} \\ p_z = a_2\left(\dot{v} + iQ_z v\right) & e^{iQ_z\theta} + \bar{a}_2\left(\dot{\bar{v}} - iQ_z \bar{v}\right) & e^{-iQ_z\theta} \end{array} \ . \tag{A.19}$$

The expressions for x and z come directly from (A.15) and (A.18), while p_x and p_z were obtained by differentiating those expressions, keeping in mind that $\dot{y} = p_y$. The constants of the motion are here the complex quantities a_1 and a_2.

According to the development of Section A.1, the Eqs. (A.19) must be solved with respect to the four constants a_1, $\bar{a}_1$ a_2 and $\bar{a}_2$ of the unperturbed motion. Leaving this task to the reader as an exercise, we just write the solution [2]

$$\begin{aligned} a_1 &= \frac{1}{W(u)}\left[\left(\dot{\bar{u}} - iQ_x\bar{u}\right)x - \bar{u}p_x\right]e^{-iQ_x\theta} \\ \bar{a}_1 &= \frac{-1}{W(u)}\left[\left(\dot{u} + iQ_x u\right)x - up_x\right]e^{iQ_x\theta} \\ a_2 &= \frac{1}{W(v)}\left[\left(\dot{\bar{v}} - iQ_z\bar{v}\right)z - \bar{v}p_z\right]e^{-iQ_z\theta} \\ \bar{a}_2 &= \frac{-1}{W(v)}\left[\left(\dot{v} + iQ_z v\right)z - vp_z\right]e^{iQ_z\theta} \end{aligned} \tag{A.20}$$

where $W(u)$ and $W(v)$ are the Wronskians associated with Floquet's functions, reducing to

$$W(u) = u\left(\dot{\bar{u}} - iQ_x\bar{u}\right) - \bar{u}\left(\dot{u} + iQ_x u\right) = W(v) \equiv -i \tag{A.21}$$

Let us now introduce a perturbation H_1 to the betatron motion and assume that we know explicitly the corresponding function U, according to Eqs. (A.11). Perturbation theory and second relation (A.11) make it possible to write the four first-order equations for the quantities a, now functions of θ. As a second exercise, the reader may compute the Poisson brackets $[a_j, a_m]$ in the particular case of the betatron motion [Eqs. (A.20) and (A.21)]. It will be found that they confine themselves to simple complex numbers that combine with the Wronskians to give [2],

$$\begin{aligned} \frac{\mathrm{d}\bar{a}_1}{\mathrm{d}\theta} &= -i\frac{\partial U}{\partial a_1} & \frac{\mathrm{d}a_1}{\mathrm{d}\theta} &= i\frac{\partial U}{\partial \bar{a}_1} \\ \frac{\mathrm{d}\bar{a}_2}{\mathrm{d}\theta} &= -i\frac{\partial U}{\partial a_2} & \frac{\mathrm{d}a_2}{\mathrm{d}\theta} &= i\frac{\partial U}{\partial \bar{a}_2} \end{aligned} . \tag{A.22}$$

At this point, it is important to stress that Eqs. (A.22) are identical to canonical equations, if we introduce new canonical variables q_ρ^n and p_ρ^n and their associated Hamiltonian U^n (equal to U), according to

$$\begin{aligned} q_1^n &= \bar{a}_1 & p_1^n &= ia_1 \\ q_2^n &= \bar{a}_2 & p_2^n &= ia_2 \\ U^n\left(q_1^n, p_1^n, q_2^n, p_2^n, \theta\right) &= U\left(a_1, \bar{a}_1, a_2, \bar{a}_2, \theta\right) . \end{aligned} \tag{A.23}$$

So far *everything is exact* again and we are sure that *symplecticity* has been *conserved* through the whole treatment, since the equations (A.22) of the perturbed motion are still canonical. At this point, to solve the perturbed betatron-motion, it is necessary to find the solution of Eqs. (A.22).

REFERENCES

[1] G. Guignard, Nonlinear Dynamics, in "Physics of Particle Accelerators", AIP Conference Proceedings 184, Vol. 1, pp. 822-890 (1989).

[2] G. Guignard, The general theory of all sum and difference resonances in a 3-D magnetic field, CERN 76-06 (1976).

[3] G. Ripken, Untersuchungen zur Strahlführung und Stabilität der Teilchenbewegung unter Berücksichtigung einer Kopplung der Betatronschwingungen, DESY R1-70/4 (1970).

[4] A. Schoch, Theory of linear and nonlinear perturbations of betatron oscillations in AG synchrotron, CERN 57-21 (1957).

[5] R. Hagedorn, Stability and amplitude ranges of 2D nonlinear oscillations with periodical Hamiltonian, CERN 57-1 (1957).

[6] P.J. Bryant, A simple theory for betatron coupling, CERN-ISR-MA/75-28 (1975).

[7] G. Guignard, Beam blowup and luminosity reduction due to linear coupling, CERN ISR-BOM/77-43 (1977).

[8] K. Takikawa, A simple and precise method for measuring the coupling coefficient of the difference resonance, CERN ISR-MA/75-34 (1975).

[9] P.J. Bryant and G. Guignard, Methods for measuring the complex coupling coefficient for the second-order difference resonance, CERN ISR-MA/75-42 (1975).

[10] A. Piwinski, private communication.

[11] M. Sands, The physics of electron storage rings, An introduction, SLAC-121 (1970).

[12] LEP Design Report, Vol. II, The LEP main ring, CERN-LEP/84-01 (1984).

[13] G. Guignard, Adjustment of emittance ratio by coupling control in electron-positron storage rings, Proc. 11th Int. Conf. on High-Energy Accelerators, CERN, Geneva, 1980 (Birkhauser, Basle, 1980), p. 682.

[14] G. Guignard et al., Measurement and compensation of the solenoid effects in LEP, Proc. of EPAC 90, 2, 1432 (1990).

[15] J. Billan et al., Suppression of the Main LEP Coupling Source, Proc. of PAC 93, Washington D.C., May 1993.

[16] J.P. Gourber, G. Guignard, A. Hofmann, J.P. Koutchouk and H. Moshammer, Compensation of linear coupling in LEP, Proc. of EPAC 90, 2, 1429 (1990).

[17] G. Guignard, LEP imperfection diagnosis and correction, Proc. on "Effect of errors in Accelerators", 5th ICFA Beam Dynamics Workshop, Corpus Christi, USA, Oct. 1991.

[18] G. Guignard, Skew quadrupole schemes for the latest LEP design, Private Communication, 1979 (LEP Note 199).

[19] H. Bruck, Accélérateurs circulaires de particules; Introduction à la théorie, Presses Universitaires de France, Paris (1966).

[20] E.D. Courant and H.S. Snyder, Theory of the alternating-gradient synchrotron, Ann. Phys. USA 3, 1 (1958).

CHROMATICITY

A. Verdier
CERN, Geneva, Switzerland

Abstract
The meaning of 'Chromaticity' is taken here as : change of the linear parameters of the transverse motion of a single particle related with a change of the beam energy, for a machine in a fixed status. We give here an overview on general theoretical aspects with a special emphasis on the case of low-β insertions.

1 INTRODUCTION

Etymologically, chromaticity comes from the Greek word $\chi\rho\omega\mu\alpha$ which means colour. Its use in the accelerator field comes from the fact that, like electro-magnetic waves, the charged particles are deflected according to their energy. However the expression 'Chromaticity correction' has a different meaning depending on the person who uses it. In the old times of accelerators, it meant only correction of the first derivative of the tunes with respect to the momentum. The word chromaticity itself was used for the expression Q'/Q because it has a value of the order of -1 for a simple lattice. Nowadays the word chromaticity is used for 'change of linear optics parameters with beam energy' and will be used below with this meaning. In terms of transfer maps or aberrations, this means considering contributions linear in the transverse coordinates and of arbitrary order in relative momentum deviation. This new meaning of 'chromaticity' arises from the increasing complexity of the machine lattices and correction procedures.

As the chromaticity correction will use non-linear fields, it must be treated with consideration of the non-linear dynamics. Therefore in practice the chromaticity correction will involve non-linear correction too. In this respect, a good method, which can be implemented easily only in large machines, is to construct multipole schemes which are a priori free from low-order transverse non-linearities for any status of the chromaticity correction.

At this point it is worth questioning whether effects which are non-linear both in momentum and coordinates should be considered. From experience, it appears that, if the chromaticity correction is done with multipoles of order below octupole, these effects can be neglected. This means that, as long as the betatron tunes are far from a linear instability condition, the non-linear dynamics are little affected by a momentum change. If the chromaticity correction is done with higher-order multipoles, it may happen that unacceptable non-linearities are created on the off-momentum closed orbit. The most simple example is the creation of off-momentum octupoles from decapoles, which makes an off-momentum anharmonicity. Such effects will not be considered here.

In this course we will give the basic principle which can be used to perform a chromaticity correction in a wide range of machines. The methods differ because all machines do not have the same needs. For instance the correction with sextupole families

is very efficient for a large machine like LEP, with low-β insertions which make strong non-linear chromatic perturbations. This sort of correction is mandatory to preserve the quantum lifetime of the beam. However, for a proton machine like LHC, which has the same size as, and similar optics to LEP, it is useless because the useful momentum range is far smaller. In a small machine with a small dispersion such as a synchrotron light source, the chromaticity correction with chromatic sextupole families cannot be done, and it is necessary to compensate for the geometric aberrations by means of sextupoles which do not act on the chromaticity.

In what follows, we will first examine why the chromaticity correction has to be made. We will give basic concepts on a simple example. Then, we will give examples of tools used to perform a chromaticity correction. Finally, we will treat in more detail the problem of a machine with low-β insertions. For chromaticity correction involving corrections of non-linear transverse motion, the reader should refer to the courses on transverse dynamics or to reference [1].

2 WHY CORRECT THE CHROMATICITY ?

The first reason why the derivatives of the tunes with respect to momentum have to be cancelled is to eliminate the dipole mode of the head-tail instability above transition energy. This phenomenon was first observed in Frascati in the storage ring ADONE where a feed-back system was used to cure this instability, as there was no sextupole available. An analysis can be found in Ref. [2].

As this instability is a potential limit to the beam intensity, it is preferable to make the first-order tune derivatives slightly positive in any machine design. This is extremely easy to do, two sextupole families are enough to adjust the two first derivatives of the tunes with respect to momentum. The exact values to which these derivatives have to be set depend on the transverse impedance of the machine. It may happen that, as in LEP, they must not be made too largely positive, otherwise the transverse quadrupole mode m=-1 become unstable.

On top of the cure of this head-tail instability, the first-order chromaticity correction is mandatory in large machines, because the high value of the derivatives results in a betatron instability for a small momentum deviation. For instance a tune derivative of -250 leads surely to a betatron instability for a relative momentum deviation in a one per-mill range. This is indeed what occurs for LEP as shown on Fig. 1.

Once the first derivatives of the tunes are corrected, the variations of the tunes with momentum may still be unacceptable. In order to examine this, we must have a good computer program which calculates these variations. This is not trivial, we will give examples below which illustrate mistakes made in the past. It may then happen that variations with momentum of non-linear parameters like anharmonicity are unacceptable. Clearly chromaticity correction may take us very far! Before going to such complications, we first look at the most simple chromaticity correction we can think of.

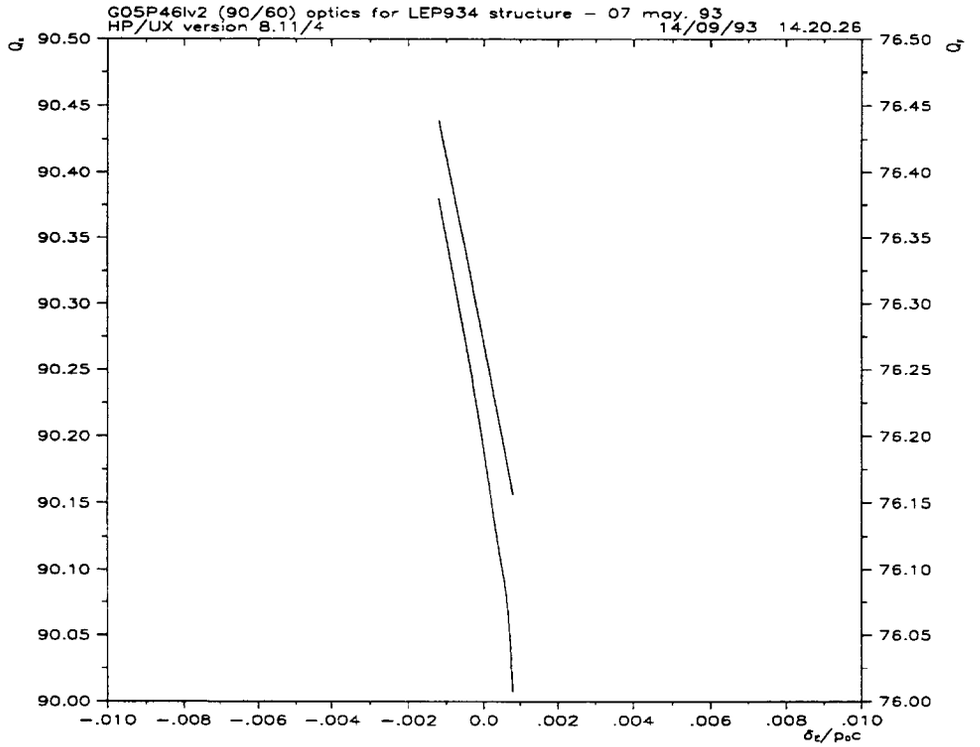


Figure 1: Tunes versus momentum for the LEP 1993 physics optics with no sextupole, G05P46lv2 optics, Q_h=90.27, Q_v=76.19, β^* of 5cm and 1.25m.

3 CHROMATICITY CORRECTION OF A FODO CELL

3.1 Natural tune derivatives

The most simple FODO-cell model we can think of is made from thin quadrupoles and thin dipoles. Note immediately that using thin dipoles is questionable because the two ends of the dipoles, which produce second-order aberrations, are at the same place and cancel each other. Therefore the simple model is then restricted to the case where the radius of curvature of the dipoles is larger than the length of the dipole in a half cell of length L by at least one order of magnitude. The phase advance μ in a FODO cell is given by [3]:

$$\sin(\mu/2) = LKl/4. \tag{1}$$

Kl is the integrated normalised gradient $B'l$, which is the same in the F and D quadrupoles, divided by $B\rho$, the magnetic rigidity of the beam. In Eq. (1), the only quantities which depend on δ are μ and K. The explicit dependence of K is:

$$K = \frac{k}{1+\delta} = k(1 - \delta + \delta^2 - \delta^3 + ...) \tag{2}$$

where k is the value of the normalised gradient for the energy of the reference orbit. From now on, we will use a capital letter for a quantity which depend on δ, as in (2), and small

ones for the value of the same quantity for δ=0. Differentiating Eq. (1) with respect to the relative momentum deviation δ, we obtain :

$$\mu' = -\frac{2}{1+\delta} tg(\mu/2)$$

Note that the ratio μ'/μ has values close to -1 for phase advances below 90°. Therefore the ratio Q'/Q is of the order of -1 for a machine made of FODO cells. This was the case of the first alternating-gradient machines and this ratio was called 'chromaticity ' (and given the symbol ξ).

Differentiating Eq. (1) twice, we obtain:

$$\mu'' = \frac{2\tan(\mu/2)}{(1+\delta)^2}\left[1 + \frac{1}{1+tg^2\mu/2}\right].$$

This formula provides only an order of magnitude because the model is already approximate to first-order. However, we see already that the natural second-order derivative of the tunes is numerically of the same order as the natural first-order derivative, but with an opposite sign, for a machine made of FODO cells, neglecting the effect of the dipoles.

It is easy to check that the derivatives of higher order are still of the same order of magnitude. Consequently, the variation of the tunes of a machine made of FODO cells with equal gradients is surely smooth in a range of momentum deviation of some percents.

3.2 Cancelling the first derivatives of the tunes

We exploit the fact that off-momentum particles oscillate around an off-centred closed orbit which passes off-centre in the machine magnets. The position of the off-momentum closed orbit is given to first-order in δ by the product $D\delta$, where D is the dispersion function.

Introducing in the FODO structure a sextupole, which has a quadratic field variation, produces a gradient depending linearly on δ, thanks to the existence of the dispersion function. This can be explained as follows. The normalised components of the sextupole field are $K'(x^2 - y^2)/2$ in the horizontal plane and $-K'xy$ in the vertical one. Here K' is the normalised second derivative of the magnetic field : $B''/B\rho$. In order to look for a betatron oscillation around the off-momentum closed orbit, we substitute for x : $D\delta + x_\delta$ (position of the off-momentum closed orbit to first-order in δ, plus betatron oscillation) in the above expressions of the sextupole field. Keeping the terms linear in x_δ in this expansion, we observe that a gradient equal to $k'D\delta$ in the horizontal plane and $-k'D\delta$ in the vertical plane appears. k' is now the value of K' for $\delta = 0.0$. Choosing $k'D = k$, we compensate exactly the decrease of the quadrupole gradient to first-order in δ. Doing this exercise for the two quadrupoles of the FODO cell makes an exact compensation of the decrease of the quadrupole gradient with the momentum deviation to first-order in δ.

The reader can check for himself that this argument can be extended to higher order: a field varying like x^n can compensate a dependence of the gradient on δ^{n-1}. Doing this compensation, variations of order higher than δ^n are introduced. Therefore a deeper analysis has to be done. A first approach is given at the end of the course. On top of

these higher-order chromatic aberrations, non-linear terms are introduced in the equation of motion : their effect has been analysed in courses on non-linear dynamics (see those in this course or in [1] for instance).

4 GENERAL CHROMATICITY CORRECTION

After this introduction concerning basic concepts, we tackle the chromaticity problem for more complicated cases and with more accuracy. In the problem hierarchy, the first one is the computation of the effects. Indeed it is useless to compute beautiful corrections if we are not able to check their efficiency.

The best way to compute chromatic effects is:

- calculate the position of the off-momentum closed orbit by means of an accurate mapping per element
- determine the associated tunes and β-functions. To this end we must have a correct computation of the transfer matrices around the computed off-momentum closed orbit or an efficient algorithm which can extract the tune values from tracking.

This will provide all variations of the optics parameters with momentum with the maximum accuracy. Many computer codes have been made to solve this problem. An exhaustive review can be found in Ref. [4].

In the case where we are confronted with a non-standard chromaticity problem, the best procedure is to make computations with several different codes. As shown below, this can reveal a mistake. But of course the best thing is to go back to the exact equation of motion in order to make sure of the physics involved in our problem.

Before going into the details of the computation of chromatic effects, it is of prime importance to stress that the numerical computation of the δ-dependence can be performed exactly for machines without synchrotron radiation. This is a consequence of the fact that a mere scaling of the canonical momenta is a canonical transformation. If all momenta are divided by the actual momentum of the particle, the electro-magnetic potential is also divided by this actual momentum p_s so that eventually the momentum dependence disappears from the Hamiltonian of the motion. For a motion with synchrotron losses a similar procedure can be applied, except that p_s is now a sort of momentum averaged over the circumference. Then, in order to describe the synchrotron motion, it is necessary to expand the equations of motion with respect to a variable p_t given by:

$$p_t = \frac{E}{p_s c} - \frac{1}{\beta_s}$$

where E is the energy and β_s the relativistic factor associated with p_s. As mentioned above, this normalisation to p_s is only applicable to numerical computations, when the explicit dependence with respect to the momentum deviation is not wanted. In what follows we want to describe explicitly the δ-dependence, therefore we will go back to the equation of motion for the computation of chromatic effects.

5 EQUATION OF MOTION

Starting from the exact relativistic equation of motion with the Lorentz force:

$$d\vec{p}/dt = e\vec{v} \times \vec{B}$$

we expand it in a curvilinear coordinate system with a curvature h in the horizontal plane and project the motion on the three axes. This has been done by several authors [5, 6]. The exact equation of motion projected on the three axes gives a set of three equations:

$$x'' - h(1+hx) - \frac{x'}{T'^2}[x'x'' + y'y'' + (1+hx)(hx)'] = \frac{T'}{1+\delta}(y'b_s - (1+hx)b_y)$$

$$y'' - \frac{y'}{T'^2}[x'x'' + y'y'' + (1+hx)(hx)'] = \frac{T'}{1+\delta}(-x'b_s + (1+hx)b_x)$$

$$2(hx)' - \frac{(1+hx)}{T'^2}[x'x'' + y'y'' + (1+hx)(hx)'] = \frac{T'}{1+\delta}(x'b_y - y'b_x)$$

The prime denotes here a derivation with respect to the longitudinal coordinate s, and T is obtained from :

$$T'^2 = (1+hx)^2 + x'^2 + y'^2.$$

There is clearly a notation problem associated with the prime which means either derivative with respect to the longitudinal coordinate s or with respect to the relative momentum deviation. In fact, for all variables used in this course except the dispersion function, the differentiation is unambiguous and will be specified. For the dispersion function the derivatives with respect to momentum will be noted with indices like D_2. The b's with indices are the normalised components of the magnetic field: $b = \mathcal{B}/B\rho$, where $\mathcal{B}$ is the magnetic field itself and $B\rho$ the magnetic rigidity of a particle on the reference orbit. Setting x, y and their derivatives equal to zero, we have :

$$h = b_y(0,0,s)$$

This fixes the sign convention in the field components : a positive b_y means focussing in the horizontal plane. This convention should be applied in any good optics code, it is done in the code MAD [7] that we use at CERN. For right multipoles (i.e. such that the field in the plane of the machine is parallel to the main dipole field), the expansion of $b_y(x,y,s)$ is, up to decapole :

$$b_y(x,y,s) = h + K(s)x + \frac{K'(s)}{2!}[x^2 - y^2] + \frac{K''(s)}{3!}[x^3 - 3xy^2] + \frac{K'''(s)}{4!}[x^4 - 6x^2y^2 + y^4]$$

In order to obtain the chromatic effects, we have to keep all monomials linear in the transverse coordinates and their derivatives. For instance, we show the classical example of the computation of the derivatives of the tunes with respect to momentum. To this end we must expand the equations to second-order in the two transverse planes, because we want to have monomials in $x\delta, x'\delta, y\delta$ and $y'\delta$. This expansion gives (see for instance Ref. [5]) :

$$x'' - h(1+hx) - x'(hx' + h'x) = \frac{T'}{1+\delta}[y'b_s - (1+hx)b_y]$$
$$y'' - y'(hx' + h'x) = \frac{T'}{1+\delta}[-x'b_s + (1+hx)b_x].$$

We consider only the transverse planes and not at all the longitudinal one. It is clear that kinematical terms as well as terms coming from the discontinuity of the curvature h of the reference orbit appear, which were not in the simple model used to establish formula (1).

In order to have the equation for the betatron motion around the off-momentum closed orbit we then substitute x for $D\delta + x_\delta$ and keep in the expansion the monomials in x_δ and x'_δ. We obtain :

$$x''_\delta + (h^2+k)x_\delta - (k - k'D + 2h^2 + h'D' - 2Dh^3 - 4hkD)x_\delta\delta - (hD)'x'_\delta\delta = 0$$
$$y''_\delta - ky_\delta + (k - k'D + h'D' - 2hkD)y_\delta\delta + (hD)'y'_\delta\delta = 0 \qquad (3)$$

We observe that the perturbation in δ contains two terms linear in x and x' and hence the chromatic effects cannot be estimated by means of the well known first-order tune-shift formula [8] :

$$\Delta Q = \frac{1}{4\pi}\oint \Delta k(s)\beta(s)ds.$$

An estimate of the importance of the error when using this formula can be found in a report on the design of the LEAR machine at CERN [9]. It appears that the ratio Q'/Q can be wrong by a factor as large as 2 if the calculation is not properly done. However such a dramatic effect may only occur in small machines with a radius of the same order as the dispersion function. For instance in LEP, where the dispersion is of the order of a metre and the radius of curvature three kilometres, the error due to neglecting these terms is as the ratio of these numbers, i.e. of a tenth of unit in Q' for a Q' value of the order of -100.

Nowadays accurate transfer maps to large order in transverse coordinates have been computed for all sorts of magnets by means of modern techniques such as LIE algebra [10] or Differential Algebra [11]. Therefore there is little excuse for not computing the chromatic effects properly in any machine.

6 COMPUTATIONS OF GRADIENT PERTURBATIONS

Being able now to compute chromatic effects, we wish to compensate the detrimental ones. One possibility is to establish formulae giving the effect of gradient perturbations.

The principle of the method used is rather simple: as we deal with perturbations which modify linear terms in the equation of motion, we can use matrix calculus to estimate their effect. The equation of motion contains chromatic terms (monomials in δ) which can be considered as a perturbation. These terms give exactly the perturbation of x' due to an infinitely thin element. As there is no change in x, the transfer matrix of such a perturbation is straightforward. Multiplying the unperturbed Twiss matrix at the place of the perturbation by the transfer matrix of the perturbation makes it possible

to compute the Twiss matrix associated with the perturbed machine and then the new tunes. This procedure can be generalised to higher order.

Before examining this in detail, we must stress that such perturbation calculations are the basis of the computation of off-momentum transfer matrices for any sort of magnetic machine elements. Such computations can be found for instance in [12, 13].

6.1 First-order focussing perturbation

In equations (3), the chromatic terms can be written as :

$$C_z z\delta + C_{pz} z'\delta$$

where z stands for x or y. The transfer matrix associated with such a perturbation of infinitely small length ds is obtained considering that :

- the transverse coordinate z does not change
- the change in z' is merely $z''ds$

The transfer matrix is then:

$$\begin{pmatrix} 1 & 0 \\ -C_z\delta ds & 1 - C_{pz}\delta ds \end{pmatrix} \tag{4}$$

We notice readily that this matrix is not symplectic, i.e. its determinant is not 1. This is due to the term C_{pz} coming from the dipole ends. As an actual dipole has two ends, it is not physical to consider one isolated end. If we compute the off-momentum transfer matrix for a complete dipole, we notice that it is indeed symplectic. Therefore it is legitimate to compute the Twiss matrix of the perturbed machine by multiplying merely the unperturbed Twiss matrix :

$$\begin{pmatrix} \cos\mu + \alpha\sin\mu & \beta\sin\mu \\ -\gamma\sin\mu & \cos\mu - \alpha\sin\mu \end{pmatrix}$$

in which the functions β and α are computed at the place of the perturbation, by the matrix of the perturbation. From the trace of the product, we obtain the tune μ^* of the perturbed machine :

$$\cos\mu^* = \cos\mu - \frac{1}{2}\left[C_z\delta\beta\sin\mu + C_{pz}\delta(\cos\mu - \alpha\sin\mu)\right] ds.$$

The tune μ^* of the perturbed machine is :

$$\mu^* = \mu + \mu'\delta.$$

Expanding $\cos\mu^*$ and identifying the first-order terms in δ, we obtain

$$\mu' = \frac{1}{4\pi}\oint [C_z\beta(s) + C_{pz}(\cot\mu - \alpha)]ds$$

by a mere integration of all first-order effects. It is worth noting that, as $C_{pz} = (hD)'$, its integral is zero since hD is a function which has a one-turn period. The correct first-order formula for the computation of the tunes derivatives is then :

$$\mu' = \frac{1}{4\pi} \oint [C_z \beta(s) - C_{pz}\alpha)] ds. \tag{5}$$

Integrating the term with C_{pz} by parts gives:

$$\oint -C_{pz}\alpha ds = \frac{1}{2} \oint C'_{pz}\beta ds$$

where C'_{pz} is the derivative of C_{pz} with respect to the longitudinal coordinate s. This is the formula given in [14].

6.2 First-order chromaticity correction

We apply now the above results to the case of sextupoles. From Eq. (3), we have:

$$C_x = k'D, \;\; C_{px} = 0 \qquad C_y = -k'D, \;\; C_{py} = 0$$

We put these expressions in Eq. (5) to obtain the changes in the first derivative of the tunes with respect to δ due to sextupoles:

$$\Delta Q'_h = \frac{1}{4\pi} \oint k'(s) D(s) \beta_h(s) ds.$$

and

$$\Delta Q'_v = \frac{1}{4\pi} \oint -k'(s) D(s) \beta_v(s) ds.$$

These expression can be obtain exactly by means of the first-order tune-shift formula.

For practical applications, it is enough to have one set of sextupoles at locations where β_h is large to act mainly on Q'_h and another set where β_v is large to act mainly on Q'_v. The two above equations lead to a set of two linear equations with the two strengths of the two sextupole sets as unknowns. The same system can obviously be used to compute increments of sextupole strengths required to make increments of tune derivatives. This is applied routinely to accelerator control.

6.3 Higher-order focussing perturbation

There are two origins of the higher-order terms :

- those appearing in the expansion of the equation of motion
- combined effect of gradient perturbations of lower order.

As an example of higher-order calculations, we give below the main lines of the computation of tune variations in δ^2 in the horizontal plane.

The terms appearing in the equation of motion are obtained from an expansion of this equation up to third order in transverse coordinates and second-order in δ. Then we

substitute x for $x_\delta + D\delta + D_2\delta^2$, expand and keep the monomials linear in x_β and x'_β. The monomials in δ are given in equation (3). The monomials in δ^2 are :

$$\begin{aligned} x\delta^2 \quad & [-h'D'_2 + (2h^3 + 4hk + k')D_2 - 2hh'DD' + 3(kh^2 - h^4/2 + hk' + k''/6)D^2 \\ & +D'(2h^2 + 3k/2) + D(4hk - k'D - 2Dh^3) + 2h^2 + k] \\ x'\delta^2 \quad & [-h'D_2 - hD'_2 - hh'D^2 + (4h^2 + 3k)DD' - 3hD']. \end{aligned} \tag{6}$$

The combined effect of the gradient perturbations of first-order is obtained by generalising the first-order calculation in section 6.1. The unperturbed Twiss matrix is now divided into two sub-matrices A and B, and the matrices of perturbations P1 and P2 similar to that given in Eq. (4) are inserted to make the perturbed Twiss matrix:

$$A \times P1 \times B \times P2. \tag{7}$$

The second-order chromatic term associated with the perturbation of index i in the equation of motion can be written:

$$[Ci_x x\delta + Ci'_x x\delta^2 + Ci_{px} x'\delta + Ci'_{px} x'\delta^2]ds$$

where Ci_x and Ci_{px} are extracted from Eq. (3), and Ci'_x and Ci'_{px} are extracted from Eq. (6). The tune with the chromatic perturbations is obtained from the trace of the matrix given by (7).

In order to have the second derivative of the tune, we expand the general expression of the trace of the matrix with the tune expanded to second-order in δ :

$$\cos(\mu + \mu'\delta + \mu''\frac{\delta^2}{2}) = \cos\mu - \mu'\delta\sin\mu - \frac{1}{2}(\mu'^2 + \mu'')\delta^2\cos\mu \tag{8}$$

and then we identify the terms in δ and δ^2 in Eq. (8) and in the expression of the trace of the matrix (7). We obtain after lengthy calculations:

$$\begin{aligned} 2d\mu'/ds &= C1_x\beta_1 + C1_{px}(\cos\mu - \alpha_1\sin\mu) + C2_x\beta_2 + C2_{px}(\cos\mu - \alpha_2\sin\mu) \\ d^2\mu''/ds1ds2 &= -d\mu'^2/ds1ds2 \\ &\quad +C1'_x\beta_1\tan\mu + C1'_{px}(1 - \alpha_1\tan\mu) + C2'_x\beta_2\tan\mu + C2'_{px}(1 - \alpha_2\tan\mu) \\ &\quad +C1_x C2_x\beta_1\beta_2\sin\mu_1\sin\mu_2/\cos\mu \\ &\quad +C1_x C2_{px}\beta_1\sin\mu_2(\cos\mu_1 - \alpha_2\sin\mu_1) \\ &\quad +C2_x C1_{px}\beta_2\sin\mu_1(\cos\mu_2 - \alpha_1\sin\mu_2) \\ &\quad +C2_{px}C1_{px}(\cos\mu_2 - \alpha_2\sin\mu_2)(\cos\mu_1 - \alpha_2\sin\mu_1) \end{aligned} \tag{9}$$

The second-order derivative of the tune with respect to momentum is then obtained from the double integral:

$$Q'' = \oint ds2 \int_0^{s2} \frac{d^2\mu''}{ds1ds2} ds1.$$

In Eq. (9), we recognise in the expression of $2d\mu'/ds$ a mere sum of first-order effects identical to those given by Eq. (5).

In the second-order terms of Eq. 9), we have three sorts of terms:

- the first one is a mere 'first-order product'
- the second one is the contribution of the second-order terms of the equation of motion calculated with the first-order formula
- the third one is made from a series of products of first-order terms with phase factors, i.e. the effect of coupling between first-order terms. These terms generalise the second-order formula established in Ref. [8].

6.4 Off-momentum closed orbit to higher order

The position of the off-momentum closed orbit is, to second-order in δ :

$$D\delta + D_2\delta^2$$

We substitute x for this expression in the equation of motion and factorize in powers of δ. Setting to zero the term in δ in this equation, we obtain the standard equation of the dispersion [8]:

$$D'' + (k + h^2)D = h$$

Setting to zero the term in δ^2 in this equation, we obtain the equation of D_2 :

$$D_2'' + (k + h^2)D_2 = h'DD' - h^3D^2 + hD'^2/2 - 2hkD^2 - k'D^2/2 + kD + h$$

It is worth noting that the right hand side contains both coefficients of the equation of motion and D and D', which come from the solution of the equation of the dispersion, whilst the left hand side is the unperturbed equation of motion. This makes it possible to compute D_2 easily from the periodic forced solution of the equation of motion [8]. This recurrence property extends to higher orders, simply because the terms containing the highest-order derivative come from the terms linear in coordinates which are by definition the unperturbed equation of motion.

6.5 Higher-order chromaticity correction

The problem of correction is now less straightforward than for the linear case, because we will have to make a compromise between the contributions of all orders, and very often it is not worth having too many variables to do the job. Therefore the best thing is to make a numerical minimisation of all contributions.

Formulae based on a generalisation of the Courant and Snyder second order perturbation formula [8], with contributions of multipoles up to decapole, are old fashioned but relatively easy to implement. For instance the program HARMON [15] is based on such formulae for the chromaticity correction up to third order in δ and to second-order in betatron amplitude. They are enough to make the third-order correction in a large machine, but not for a machine with a radius of curvature comparable with its dispersion function. For the latter case formulae like (9) should be used.

A promising and elegant alternative method is the Differential Algebra technique [16] which makes it possible to obtain the contribution of variables, e.g. multipole excitations, to the chromatic monomials in the transfer map of a whole machine. From the transfer

maps it is possible to compute the expansion of the tunes with respect to momentum, i.e. the tune derivatives. This new method has the advantage that the computation time grows less fast with the expansion order than analytical methods like the above one.

We cannot leave the subject of higher-order chromaticity correction without mentioning how the problem of non-linear terms in transverse coordinates can be solved. This can be done with perturbation theories which make it possible to compute contributions to a given order. Examples can be found for instance in Ref. [1]. The formalism of gradient perturbations can be generalised to this case. For instance the above mentioned HARMON program contains the computation of anharmonicity and resonance effects. Of course the more general Differential Algebra technique can also be applied to this. The correction of non-linearities in the transverse coordinates is important in machines where the fields associated with multipoles are at least of the order of several thousandths of the quadrupole fields at the useful aperture. This occurs for instance in the synchrotron light sources [17] or in superconducting machines where the field defects are much larger than in conventional ones.

7 CHROMATICITY OF A MACHINE WITH LOW-β INSERTIONS

A low-β insertion is a part of a machine in which the envelope functions β are made small in both planes in order to reduce the beam size. This is the case in storage rings at the crossing points. Because of the energy spread in the beam, the dispersion functions have also to be made small in order not to destroy the effect of the low-β . Usually the dispersion functions are made zero in the whole insertion because the RF cells, which must be put in regions without dispersion, are conveniently placed close to them. In such a region without dispersion, a local correction, as explained in section 3, cannot be made.

In order to make a local correction, we could try to let the dispersion function oscillate in the insertion quadrupoles and make it zero at the crossing point only. However, if there is no bending magnet between the quadrupoles close to the crossing point, the dispersion has opposite values in these quadrupoles because the phase advance is close to π. The sextupoles for the local correction have then opposite strengths and their non-linear kicks act in phase, which necessitate a compensation with additional sextupoles.

In fact a solution exists for a non-local correction. The problem is the 'off-momentum mismatch' of the insertion, the latter being matched on central orbit, to a lattice which is chromatically corrected. Thus, rearranging the sextupoles in the lattice to force the off-momentum β's to take values close to those which make an 'off-momentum matching', is the way to a solution.

We examine here how to estimate the effect of this off-momentum mismatch, which is a particular case of an insertion mismatch due to a gradient perturbation [18].

7.1 Global estimation of a gradient perturbation

We consider a perfect machine at the end of which the betatron functions have the values β and α. By definition the betatron functions have also the values β and α at the beginning of the machine.

We introduce in this machine a certain gradient perturbation. The effect of this perturbation can be computed exactly with the transforms of β and α through the perturbed machine, which are β^t and α^t, as well as the associated phase advance μ^t defined by :

$$\mu^t = \int_0^C \frac{ds}{\beta^{t,s}} \tag{10}$$

where $\beta^{t,s}$ is the transform of β at the point of longitudinal coordinate s. Indeed these quantities are enough to obtain the perturbed Twiss matrix the expression of which is [19] :

$$\begin{pmatrix} \sqrt{\frac{\beta^t}{\beta}}(\cos\mu^t + \alpha\sin\mu^t) & \sqrt{\beta\beta^t}\sin\mu^t \\ \frac{1}{\sqrt{\beta\beta^t}}((1+\alpha\alpha^t)\sin\mu^t + (\alpha^t-\alpha)\cos\mu^t)) & \sqrt{\frac{\beta}{\beta^t}}(\cos\mu^t - \alpha_t\sin\mu^t) \end{pmatrix}$$

It is important to note that β^t and α^t are not true Twiss functions : they have the same meaning as Twiss functions in a transfer line. The actual β-function at the end of the perturbed machine β^* can be obtained from the second element of the first line of this matrix :

$$\beta^* = \sqrt{\beta\beta^t}\sin\mu^t / \sin\mu^*$$

where the new tune μ^* can be computed from the trace of the perturbed matrix :

$$2\cos\mu^* = \left(\sqrt{\frac{\beta^t}{\beta}} + \sqrt{\frac{\beta}{\beta^t}}\right)\cos\mu^t + \left(\alpha\sqrt{\frac{\beta^t}{\beta}} - \alpha^t\sqrt{\frac{\beta}{\beta^t}}\right)\sin\mu^t \tag{11}$$

Putting :

$$\theta = -\arctan\frac{\alpha\sqrt{\frac{\beta^t}{\beta}} - \alpha^t\sqrt{\frac{\beta}{\beta^t}}}{\sqrt{\frac{\beta^t}{\beta}} + \sqrt{\frac{\beta}{\beta^t}}} = \frac{\alpha^t\beta - \alpha\beta^t}{\beta + \beta^t}$$

We can transform equation (11) into :

$$\cos\mu^* = \cos(\mu^t + \theta)\sqrt{1 + \frac{1}{4}\left(\sqrt{\frac{\beta^t}{\beta}} - \sqrt{\frac{\beta}{\beta^t}}\right)^2 + \frac{1}{4}\left(\alpha\sqrt{\frac{\beta^t}{\beta}} - \alpha^t\sqrt{\frac{\beta}{\beta^t}}\right)^2} \tag{12}$$

In order to obtain this expression, there is a trick consisting of adding 4 to the sum of the squares of the coefficients of the trigonometric functions in equation (11), so that the sign plus in the first one can be changed to minus. As the term under the square root is always larger than 1, there are values of $\mu^t + \theta$ for which $\cos\mu^*$ is larger than 1, even if the unperturbed $\cos\mu$ is smaller than 1 : the gradient perturbation has opened 'gradient stop-bands'.

It is worth noting that the expression under the square root can be used as a measure of mismatch when trying to match an insertion [18]. In the case of an imperfect matching, minimising this term guarantees that the stop-bands associated with the mismatch have the minimum width.

7.2 Application to a chromatic perturbation

We expand μ^t and $\mu*$ in power series of the relative momentum deviation δ :

$$\mu^t = \mu + \mu^{t'}\delta + \frac{1}{2}\mu^{t''}\delta^2 + ...$$

μ being the on-momentum phase advance.

$$\mu^* = \mu + \mu^{*'}\delta + \frac{1}{2}\mu^{*''}\delta^2 + ...$$

We expand also β^t and α^t :

$$\beta^t = \beta + \beta'\delta + ...$$
$$\alpha^t = \alpha + \alpha'\delta + ...$$

β and α being the on-momentum values. We put all these expansions in (11), expand in powers of δ and identify the terms with the same power of δ on both sides. We obtain for the first-order :

$$\mu^{*'} = \mu^{t'} - \frac{\alpha\beta' - \beta\alpha'}{2\beta}$$

This is already interesting : the first derivative of the tune cannot be obtained simply by the expansion of the integral in (10). To second-order, we obtain after substitution of $\mu^{*'}$ with its above expression :

$$\mu^{*''} = \mu^{t''} - \frac{1}{4}\cot\mu\left[\left(\frac{\beta'}{\beta}\right)^2 + \left(\frac{\alpha\beta' - \beta\alpha'}{\beta}\right)^2\right] + \frac{\alpha\beta'' - \beta\alpha''}{2\beta} + \frac{\beta'}{2\beta^2}(\alpha'\beta - \beta'\alpha) \qquad (13)$$

In this expression of the second-order derivative of the tune with respect to momentum, $\mu^{t''}$ is a mere contribution of second-order aberrations, and the other terms describe the effect of the off-momentum mismatch of the β-function to first and second-order. It must be recalled that β' is not the derivative of the β-function with respect to momentum. It is the derivative with respect to momentum of the transform of β, the on-momentum value, through the machine. The computation of this β' can be easily done from the derivative of the expression of the transform of the β-function through a thin quadrupole of length l, which is :

$$\frac{\beta'}{\beta} = -kl\beta_0 \sin 2[\mu - \mu_0]$$

where the unlabelled optics parameters refer to the point of longitudinal coordinate s where the derivative is computed and the quantities labelled 0 refer to the quadrupole location. Taking the derivative of this expression with respect to s gives :

$$\alpha\frac{\beta'}{\beta} - \alpha' = kl\beta_0 \cos 2[\mu - \mu_0]$$

From these expressions, the first-order contributions are easily obtained by integration over the machine. Such a calculation has been implemented in MAD [7] (command TWISS CHROM) in order to estimate quickly the importance of these basic effects. Substituting the two above expressions in equation (13), it appears that the contribution of a single quadrupole to the off-momentum mismatch effect goes with $(kl\beta_0)^2$. It is worth comparing this with the second-order contribution, which comes from the variation of K with momentum (Eq. (2)), which is $+2kl\beta_0$.

7.3 Non-linear chromaticity due to a low-β insertion

In the case of a low-β insertion, several quadrupoles contribute to the chromatic effects but there is at least a strong one, close to the crossing point, which has a dominant effect. In order to give an idea of the order of magnitude of this effect, we consider the case of LEP under physics conditions, in the vertical plane. The β-value at the interaction point β^* is 5cm. The closest quadrupole is at 3.7m, it has a length of 2m and a strength k of 0.164m^{-2}. The β-value at the quadrupole centre is about 400m, and the expression $kl\beta_0$ has a value of about 130. From this number we deduce that the effect of the off-momentum mismatch induced by this quadrupole, which is proportional to its square, is larger than its second-order effect, due to the expansion of K to second-order in δ, by two orders of magnitude. For the other quadrupoles of the insertion, the product $kl\beta_0$ is always smaller than that of the above one by more than one order of magnitude, and their contributions add-up with phase factors which make their effect even smaller. This is also for the cross terms described in section 6.3.

Under these conditions we can describe the second derivative of the tune with respect to momentum by means of the term which is a sum of squares in equation (13). In order to make this formula easier to use, we express the tunes in units of 2π and we consider that there are $\mathcal{N}_s$ equal super-periods in the machine with one symmetrical low-β insertion each. As said above, the only contributions we have to consider are those of the low-β quadrupoles. If we take the origin of the super-period at the symmetry point of the low-β insertion, the phase of one low-β quadrupole is $\pi/2$ and that of the second one (close to the end of the super-period) is $2\pi Q/\mathcal{N}_s - \pi/2$. Using the above formulae for β'/β and its derivative, we obtain finally :

$$Q'' \approx -\frac{\mathcal{N}_s \cos^3 \frac{2\pi Q}{\mathcal{N}_s}}{\pi \sin \frac{2\pi Q}{\mathcal{N}_s}} (Kl\beta)^2 \qquad (14)$$

Putting the numbers given above for the LEP low-β quadrupoles in the vertical plane, we obtain a value of Q'' of -6.5$\times$10^4 with a fractional part of the tune of 0.19 and an integer part multiple of four (tunes used from 1991 to 1994 for physics), the super-periodicity of LEP being four. The parabolic variation of the vertical tune due to this Q'' is enough to produce a betatron instability, i.e. make a decrease of the vertical tune of 0.19, for a momentum deviation of $\pm 2.4 \times 10^{-3}$. The calculation done with all the quadrupoles in the machine shows that this instability occurs at about $\pm 3 \times 10^{-3}$ as shown on Fig. 2.

Such a strong non-linear variation of the tune with momentum can be compensated with special sextupole arrangements as shown below, if no parameter entering formula (14) can be changed. For LEP it has been estimated that the tunes with an integer part multiple of the super-periodicity were better for the beam-beam interaction, as the linear part of this interaction makes a decrease of β^*. This implies that $\sin(\frac{2\pi Q}{\mathcal{N}_s})$ is small and Q'' is large in the vertical plane. As a consequence, sextupole families have been foreseen for LEP.

If there is some freedom for acting on the parameters entering formula (14), it is quite easy to find situations where this variation does not occur. For instance, if the tune is close to a quarter integer modulo one half, the chromaticity correction with two sextupole families is quite acceptable. This is what has been done to test LEP with a 90°

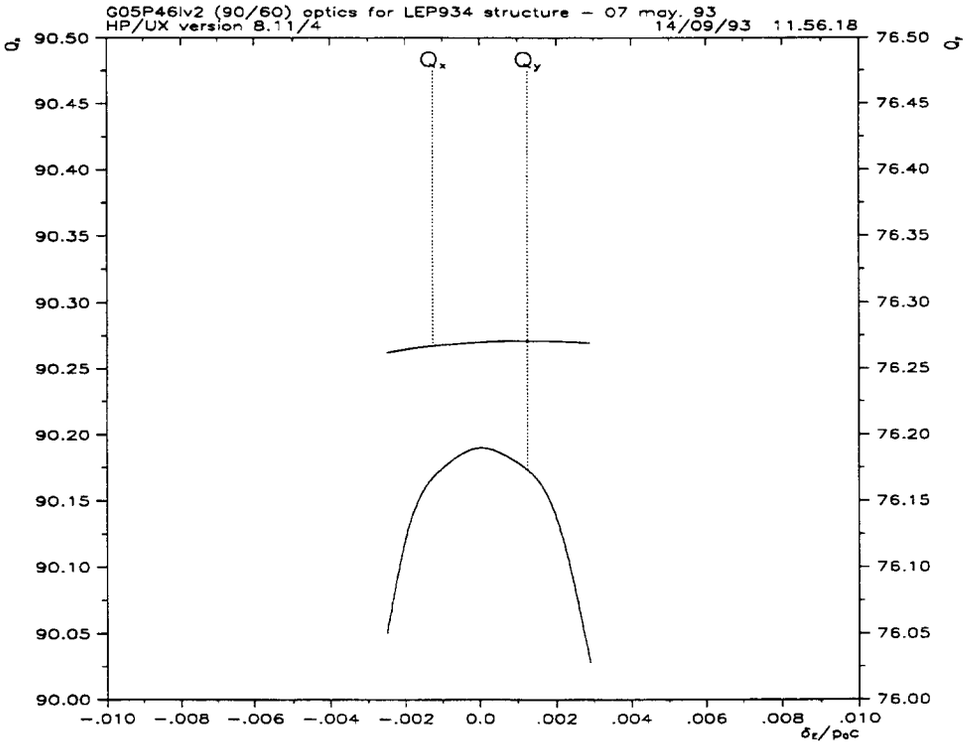


Figure 2: Tunes versus momentum for LEP under physics conditions at 46GeV, G05p46lv2-optics, Q_h=90.27, Q_v=76.19, β^* of 5cm and 1.25m, two sextupole families.

phase advance in the arc cells, whilst the sextupole families were cabled for a 60° phase advance. Choosing the tunes:

$$Q_h = 91.30 \quad Q_v = 97.20$$

made it possible correct the chromaticity with two sextupole families for a β^* of 5cm. Another possibility is to increase β^* for intermediate optics as shown below.

7.4 Scaling law for the second-order chromaticity

We have seen above that the main contribution to the second-order chromaticity is that of the quadrupole close to the crossing point, mainly because the β-value is large in this quadrupole. This large value is associated directly with the small β-value at the crossing point. The effect of the quadrupole is indeed to change the sign of the derivative with respect to the longitudinal coordinate of the β-function so that this function takes smaller values and can be matched to the lattice. This is only approximate as the matching of the insertion consists of a more subtle adjustment of the quadrupole strength, but this estimate makes it possible to obtain an interesting qualitative analysis. This change of sign is expressed, in the thin-lens approximation, by:

$$-2\alpha = Kl\beta$$

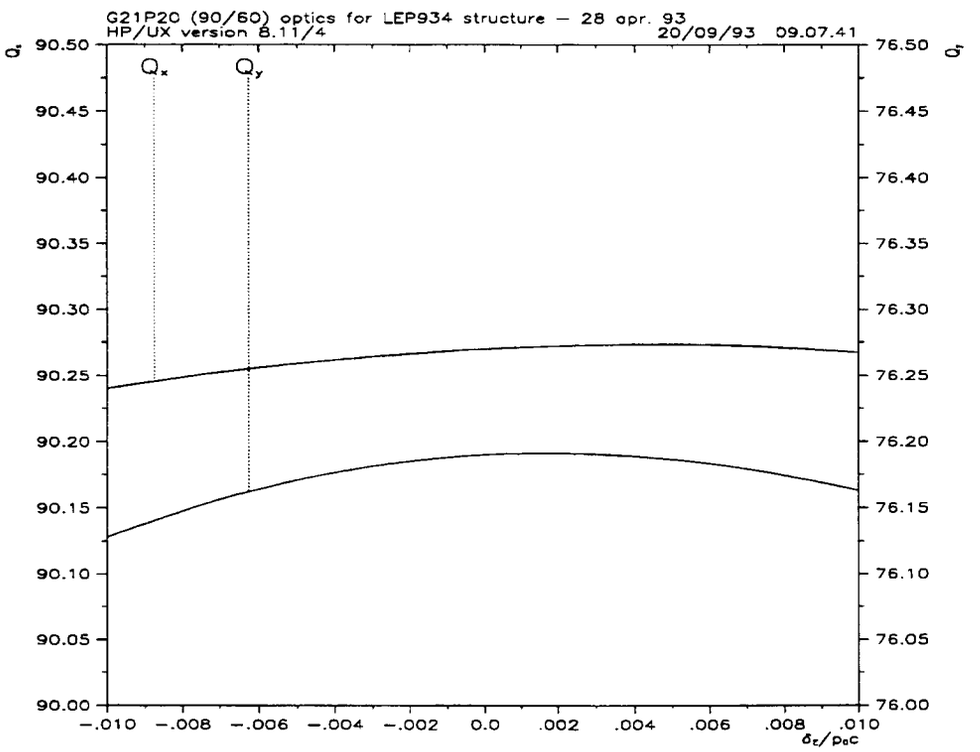


Figure 3: Tunes versus momentum for LEP at injection at 46GeV, G21P20v2-optics, Q_h=90.27, Q_v=76.19, β^* of 0.21m and 2.5m, two sextupole families.

where α is related with the value of the β-function at the crossing point by:

$$\alpha = -\frac{L}{\beta^*}$$

L being the distance between the crossing point and the centre of the quadrupole. From these two equations, we obtain the estimate of the value of the term $kl\beta$ entering formula (14):

$$kl\beta \approx \frac{L^2}{\beta^{*2}}$$

from characteristic quantities of the insertion. For the case of LEP the β^*'s are reduced by a factor 3 at injection, which reduces Qy'' by a factor 9 (and mainly reduces the maximum β-value in the machine). Thus for the injection optics, the chromaticity correction with two sextupole families is quite acceptable, as shown on Fig. 3.

8 CHROMATICITY CORRECTION WITH SEXTUPOLE FAMILIES

The off-momentum mismatch due to a low-β insertion is so evident that even for the first large machine with tunes suitable for the beam-beam interaction, it was proposed to make an off-momentum matching of the insertions by means of sextupole families

Family label	β_v at δ=-0.002	β_v at δ=+0.002
2	343	99
1	57	355
3	282	173
2	347	102
1	58	356
3	277	169
2	351	106
1	60	357
3	271	164

Table 1: Off-momentum β_v's at the nine first SD sextupoles for the present LEP. G05P46lv2 optics, two sextupole families. The sextupoles are placed in the arc FODO cells with a vertical phase advance of 60° . Thus there is a π vertical phase advance between two successive sextupoles of each family (sextupoles with the same label). Each family has its own power supply, there are 32 sextupoles per octant. As this number is not an even multiple of 3, this ensemble of sextupoles does not constitute a second-order achromat (two sextupoles are in excess). The on-momentum β_v's are about 150m.

[20, 21, 22]. One of the proposals was to introduce a 'sextupole insertion' to perform this matching. Unfortunately this leads to the use of strong localised sextupoles which result in important non-linear geometric aberrations, so condemning the project or implicating a complicated scheme for the correction of the non-linear geometric aberrations. An analysis of such a failure can be found in [23].

Rather than trying to make a sextupole insertion, it is much more efficient to change the off-momentum β's in the machine cells by breaking the off-momentum cell periodicity via re-shuffling of the sextupoles into families. However it is of prime importance to make sure that the arrangement so obtained is efficient, as the tolerance of the phase advance per cell is tight. Therefore the best procedure to make sextupole families is first to adjust the first tune derivatives with two sextupole families and then to inspect the beating of the β-functions on the off-momentum closed orbits. This beating gives a simple solution to the compensation of the higher-order tune derivatives, because it makes the off-momentum β-values different at each sextupole. In order to show how this can be exploited, it is best to give an example. The off-momentum β's at some sextupole locations in the beginning of a LEP arc are shown in Table 1.

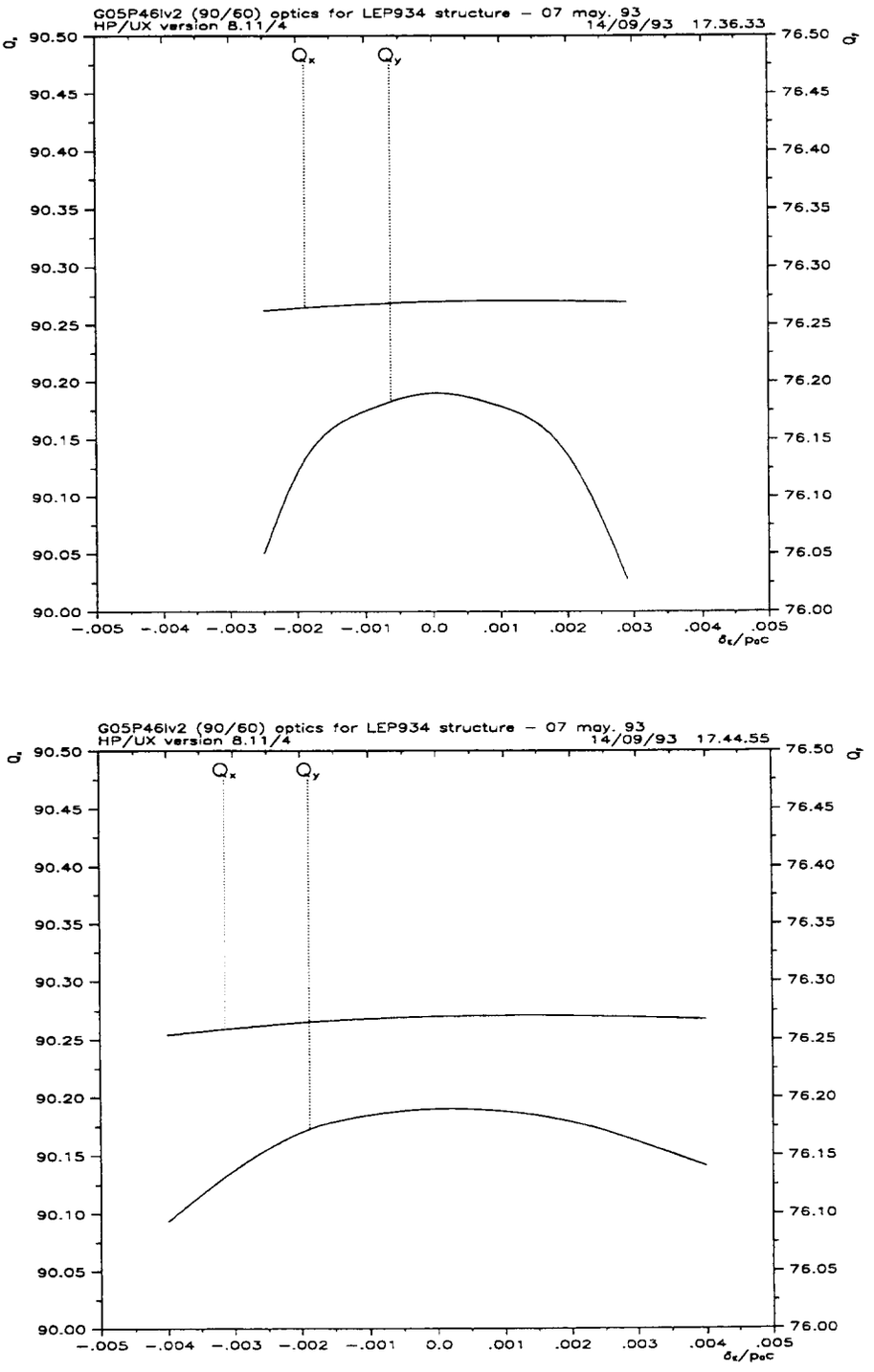


Figure 4: Tunes versus momentum for the LEP physics optics G05P46lv2. The upper graph is the same as Fig. 2 with the horizontal scale changed. For the lower graph the K' of the SD2 family has been incremented by +0.04 and that of the SD1 family has been incremented by -0.04.

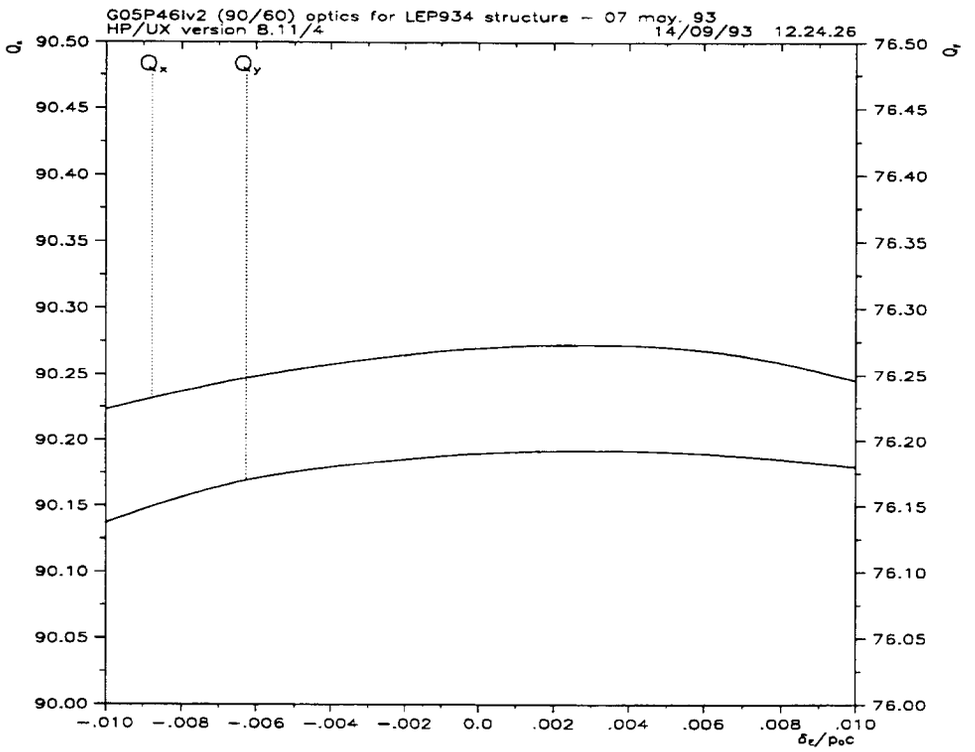


Figure 5: Tunes versus momentum for LEP under physics conditions at 46GeV, G05p46lv2-optics, Q_h=90.27, Q_v=76.19, β^* of 5cm and 1.25m, five sextupole families.

Thanks to the difference between the β-values at the sextupole locations, it is possible to increase certain sextupole strengths and decrease others, whilst keeping Q' constant, and hence make a tune-shift on an off-momentum orbit. As Q' is kept constant, this tune-shift is due to tune derivatives of order higher than, or equal to, two. Taking the numbers in Table 1, if a positive increment is applied to the sextupoles labelled 1 and the opposite increment applied to the sextupoles labelled 2, positive tune shifts are made on both negative δ and positive δ, which corrects a little the variation of the tunes with momentum. The efficiency of this action is shown on Fig. 4. A complete correction has been done by means of the HARMON program, which uses a third-order perturbation theory. The associated variation of the tunes with momentum is shown on Fig. 5. The improvement is impressive, comparing this figure and Fig. 2.

For this LEP example the vertical phase advance per cell is exactly 60° and the extrema of the off-momentum modulation of the vertical β-function are well located at the SD sextupoles. For a vertical phase advance per cell of 55°, the modulation would be dephased by 10° at each cell since it goes with twice the phase advance. This makes a change of sign of the modulation after 18 cells. As there are 30 cells per arc, it is not possible to use regular families. A more complicated distribution of sextupoles into three families has to be done, as for the first LEP project [24].

Regular sextupole families associated with a phase advance per cell which is a simple

fraction of π allow chromaticity correction schemes free from second-order aberrations [25]. It is possible to devise similar schemes completely free from geometrical aberrations on the reference orbit by grouping the sextupoles in pairs separated by a π phase advance. Under this condition, the non-linear kicks due to two sextupoles of a given pair are equal because of the quadratic variation of the sextupole field, and do not produce any effect outside the pair. This sort of arrangement is referred to as 'non-interleaved sextupole schemes'. Their drawback is obvious : for the off-momentum trajectories the compensation of the non-linear kicks does not hold any more and the dynamic aperture decreases very quickly with the momentum deviation. An example extracted from the LEP design study can be found in [26], it is shown in this paper how to optimise such schemes. For the present LEP they have been discarded because of the existence of multipole components in the arc dipoles which destroy the non-interleaved condition. However they are potentially interesting because of the large dynamic aperture they guaranty for small momentum deviations.

9 CHROMATICITY CORRECTION BY MEANS OF MULTIPOLES

We have seen in section 3 that a variation of the tunes with δ^n can be compensated with a $2(n+2)-pole$. In the equation of motion, a right $2(n+2)$-pole introduces a term :

$$\frac{K^{(n)}}{(n+1)!}\left[x^{n+1}-(n+1)x^{n-1}y^2+.....\right]$$

substituting x for $D\delta+x_\beta$ and keeping only the linear terms in x_β we obtain

$$\frac{K^{(n)}}{n!}(D\delta)^n x_\beta$$

The contribution of a term like this to the n^{th} derivative of the tunes with respect to momentum is simply obtained using the first-order formula for gradient perturbations. If there are many such terms, their contributions are merely added. However, for a correct computation of the chromatic effects to a given order, all the contributions of multipoles of lower order have to be added by means of the formulae in section 6.3 which describe the coupling between gradient perturbations. For instance if we want to have chromatic effects in δ^3, we must add contributions of :

- decapoles which are merely summed
- octupoles and sextupoles which are computed with the second-order formulae given in section 6.3
- sextupoles which are computed with a third-order formula, i.e. coupling between three gradient perturbations. Such a formula is obtained by a generalisation of what is done in section 6.3.

The implementation of the decapole and octupole contributions are quite straightforward once the second-order formalism applied to sextupoles is available. For instance they have been included in a negligible time in the HARMON program, used as a MAD routine.

There is another very simple way of making such corrections. It consists of making successive corrections in increasing order in δ by means of multipoles of increasing order.The variation of the tunes with momentum has to be recomputed after each correction. This poor man's method has the advantage of taking into account all the higher-order effects mentioned above and is extremely accurate thanks to the iteration process.

Rather than correcting the higher-order tune derivatives due to parasitic multipoles, it is possible to reduce these derivatives by merely decreasing the dispersion function. This is why 90° lattices are attractive. It is probably possible to go further if sextupole families are not needed, i.e. if there is no need of a phase advance per cell which is a simple fraction of π. However we must keep in mind that a phase advance per cell of 120° is probably forbidden because the systematic sextupole components act in phase for the excitation of the third-order non-linear resonances. In conclusion, using a phase advance per cell substantially larger than the magic value of 60° , could be a solution to the compensation of higher-order chromatic effects due to parasitic multipoles, as long as the increase of the sextupole strength needed to correct the first-order chromaticity does not ruin the effort.

10 CHROMATICITY MEASUREMENTS

At fixed machine settings, changing the RF frequency changes the beam energy, i.e. the beam goes on an off-momentum closed orbit. Indeed the RF frequency F_{RF} is a harmonic of the revolution frequency f_{rev} :

$$F_{RF} = H f_{rev} \tag{15}$$

where H is the harmonic number. The revolution frequency is, by definition :

$$f_{rev} = \beta c / C$$

where βc is the velocity of the particles, c being the speed of light. The lengthening of the off-momentum closed orbit with respect to the central one is, by definition :

$$\Delta C / C = \alpha_c \delta.$$

Taking the logarithmic derivative of equation (15), we obtain :

$$\frac{\Delta F_{RF}}{F_{RF}} = \frac{\Delta f_{rev}}{f_{rev}} = \frac{\Delta \beta}{\beta} - \frac{\Delta C}{C}$$

From the relativistic relationship between β and the momentum, we deduce:

$$\frac{\Delta \beta}{\beta} = \frac{1}{\gamma^2} \delta$$

Putting this in the previous equation, we obtain the relationship between a change of the RF frequency and a change of the beam momentum :

$$\frac{\Delta F_{RF}}{F_{RF}} = \delta \left(\frac{1}{\gamma^2} - \alpha_c \right)$$

This formula is used to compute δ from a given change of the RF frequency. Measuring the tunes for various RF frequencies makes it possible to compute the variation of the tunes with momentum, which is the most important part of the chromaticity.

If the RF frequency is kept constant and the dipole field is varied, the momentum of the beam changes but it stays on the same closed orbit. For a separated function machine, this is equivalent to changing the normalised field strengths of the focussing magnets. Doing this on the central orbit produces a tune change proportional to the momentum change, which is the measure of the derivative of the tunes with respect to momentum without any effect coming from the non-linear magnets. This is a means to obtain the 'natural' chromaticity of the linear machine without any correction.

References

[1] B. Autin, Non-linear betatron oscillations and A. Ropert, Dynamic aperture. Both lectures in the Proceedings of the CAS third advanced accelerator physics course, Ed. S. Turner, CERN 90-04 (1990).

[2] M. Sands, The head-tail effect : an instability mechanism in storage rings, SLAC-TN-69-8.

[3] E. Keil, Single-particle dynamics, Linear machine lattices, Proceedings of the first course of the International School of Particle Accelerators of the Ettore Majorana Centre for Scientific Culture, Erice ,November 1976, CERN 77-13 (1977).

[4] Computer codes for particle accelerator design and analysis : a compendium, LA-UR-90-1766. Second edition, May 1990.

[5] K. Brown, A first- and second-order matrix theory for the design of beam transport systems and particle spectrometers. SLAC report 75, or Advances in Particle Physics 1, 71-134 (1967).

[6] K. Steffen, High energy beam optics, Inter-science Publishers, New-York, 1965.

[7] H. Grote and F.C. Iselin. The MAD program (Methodical Accelerator Design), version 8.10, User's reference manual, CERN/SL/90-13(AP), (Rev. 3, January 19, 1993).

[8] E. D. Courant and H. S. Snyder, Theory of the alternating-gradient synchrotron, Annals of Physics, vol. 3, 1-48 (1958).

[9] J. Jaeger and D. Moehl, Comparisons of methods to evaluate the chromaticity in LEAR, PS/DL/LEAR/Note 81-7, (1981).

[10] A.J. Dragt, R.D. Ryne, L.H. Healy and F. Neri, MARYLIE 3.0 (1985), University of Maryland.

[11] M. Berz, Differential Algebraic description of beam dynamics to very high orders. Particle Accelerators Vol. 24, Number 2 (1989). ISSN 0031-2460.

[12] B. Autin and A. Verdier, Focussing perturbations in alternating gradient structures, CERN ISR-LTD/76-14, (1976).

[13] M. Bassetti, Perturbative method for chromatic and sextupolar effects, 2^{nd} order solution. DESY report PET-77/35, (1977).

[14] W. Hardt, J. Jaeger and D. Moehl, A general analytical expression for the chromaticity of accelerators rings, PS/DL/LEAR/Note 82-5, (1982).

[15] M. Donald, A users guide to the HARMON Programs, PEP note 311.

[16] M. Berz, Private communication (Seminar on Differential Algebra given at CERN in 1990).

[17] A. Ropert, Sextupole correction scheme for the ESRF, second advanced ICFA beam dynamics workshop, Lugano, April 1988. CERN 88-04 (1988).

[18] A. Verdier, Linear optics design of the superconducting low-β insertion in the Intersecting Storage Rings at CERN, Nuclear Instruments and Methods, 207 (1983) 301-307.

[19] K.G. Steffen, H. Hultschig and W. Kern, Use of generalised amplitude and phase functions in designing beam transport systems, DESY/A 2-70 (1960).

[20] E. Keil, The behaviour of betatron oscillations in the vicinity of half-integral structure resonances, PEP-121 (1975).

[21] B. Autin, Sextupole and octupole schemes in very high energy electron rings, CERN ISR-LTD/76-37, (1976).

[22] B. Montague, Linear optics for improved chromaticity correction, LEP Note 165 (1979).

[23] J. Hagel, A local chromaticity correction scheme applied to LEP, LEP Note 552 (1985).

[24] J.R.J. Bennett et al., Design concept for a 100GeV e^+e^- storage ring (LEP), CERN 77-14, (1977).

[25] K. Brown. SLAC-PUB 2257.

[26] A. Verdier, La luminosité à haute énergie pour le LEP11 avec β^* nominaux selon les arrangements hexapolaires. LEP Note 396 (October 3 1982).

PLANNING SEXTUPOLE FAMILIES IN A CIRCULAR COLLIDER

P.J. Bryant
CERN, Geneva, Switzerland

It is now standard practice to equip circular colliders with a strong low-β insertion in each beam-crossing region. In an electron machine such as LEP, these insertions will be flanked on either side by a long series of FODO cells for the RF cavities. Dispersion suppressors will then match the long-straight sections to the normal arcs, so that the low-β insertion and the RF sections are kept dispersion-free. Proton machines have a similar construction, except that the RF requirements are usually less and the dispersion suppressors can be moved closer to the low-β insertion. Since the long-straight sections are dispersion-free, it is not possible to use sextupoles anywhere within the crossing region for chromatic correction, which is a severe disadvantage since the final doublet in a low-β insertion is usually the biggest single source of chromatic error in such a lattice. If left uncorrected these errors will make the betatron amplitudes at the collision points highly non-linear with the momentum and a severe loss of luminosity will result. Correction sextupoles can still be placed in the main arcs to control the natural chromaticity as before, but something more sophisticated is required to account for the higher-order chromatic effects caused by the distant low-β quadrupoles and the side effects of having strong localised sextupoles in the arcs. This problem led the designers to formulate a theory for the chromatic effects in terms of a first-order perturbation of the lattice functions for off-momentum particles, which is developed in a paper by B.W. Montague in Ref. [1], to which this note can be considered as an addendum. This formulation defines two chromatic variables a and b and an average betatron phase advance, ψ, as

$$a = \mathrm{Limit}_{\Delta p \to 0} \frac{(\beta_1 - \beta_0)}{(\beta_1 \beta_0)^{1/2}} \frac{1}{\Delta p / p_0}, \qquad b = \mathrm{Limit}_{\Delta p \to 0} \frac{(\alpha_1 \beta_0 - \alpha_0 \beta_1)}{(\beta_1 \beta_0)^{1/2}} \frac{1}{\Delta p / p_0} \tag{1}$$

$$\psi = \frac{1}{2}(\mu_0 + \mu_1) \tag{2}$$

where α, β and μ have their usual meanings for the betatron lattice functions, and the subscripts $_0$ and $_1$ denote the central orbit and an off-momentum orbit respectively. The variables a and b are defined in the limit as the momentum offset between the orbits tends to zero. It is shown in [1] that these variables can be used to describe a complex variable that is known as the w-vector.

$$w = \frac{1}{2}(b + \mathrm{i}a) \tag{3}$$

and, furthermore, that the w-vector will propagate in an achromatic lattice at twice the phase advance, ψ, i.e.

$$w = \frac{1}{2}(a^2 + b^2)^{1/2} e^{\mathrm{i}2\psi} \tag{4}$$

whereas in chromatic regions, both the amplitude and phase of the w-vector are changed.

The principle problem is to find a way of compensating chromatic errors far from their sources, while controlling the resonance excitation due to the correction sextupoles so that a reasonable dynamic aperture can be preserved. The approach described here makes a matched insertion for the w-vector, rather as would be done to "capture" a closed-orbit distortion from

an experimental magnet by placing two dipoles downstream to force the distortion into a local bump. Taking the specific example of a circular collider with low-β insertions, the aim would be to keep the β^* at the crossing point constant with respect to the momentum deviation, $\delta = \Delta p / p_0$. If this is not done, there will be a loss in luminosity. First, the chromatic variables a and b are set to zero at the crossing point and tracked through the structure of the machine to the arcs where the sextupoles are situated. The strategy is then to let the w-vector propagate through the arc as far as the centre, while progressively reducing its amplitude, with a series of correction sextupoles. Thus, the machine would appear achromatic at each crossing point and at the centre of each arc. By achromatic, it is meant that at these privileged points a and b are zero and there is no first-order amplitude dependence of the lattice functions on δ. At intermediate points, a and b would be non-zero and probably large. The integrated effect of such an insertion over one turn will cause a change in the global tune spreads Q_x' and Q_z', but this can also be controlled by suitably powering the sextupoles.

In the paper by Montague [1], the changes in the chromatic variables in the thin-lens approximation are given by,

$$\begin{aligned} &\text{For a quadrupole,} \quad \Delta a \equiv -\beta_0 k_0 \Delta s, \quad \Delta b = 0 \\ &\text{For a sextupole,} \quad \Delta a \equiv -\beta_0 D k_0' \Delta s, \quad \Delta b = 0 \end{aligned} \tag{5}$$

assuming the effective length of the field component $\Delta s \to 0$, but $(k_0 \Delta s)$ and $(k_0' \Delta s)$ remain finite and are given by the lens strengths. k_0 represents the normalised gradient. More exact expressions for long lenses can be found in Ref. [2].

In this simplified model with thin lenses the w-vector will rotate at twice the betatron frequency through achromatic regions and suffer jumps in its imaginary part a at quadrupoles and at sextupoles in regions of finite dispersion. This is a rather familiar concept, being similar for example to the normalised betatron phase space diagrams for closed-orbit bumps where the ions' position vector rotates with constant amplitude at the betatron frequency Q punctuated by jumps in Y' due to the dipole kicks.

If now we wish a series of sextupoles to add constructively, we should simply arrange for the w-vector to rotate by $2n\pi$ between sextupoles. Furthermore by choosing a different series of sextupoles shifted in phase with respect to the first series, we can by vectorial addition in the complex w-phase space create any w-vector we require.

In a symmetric FODO lattice, it is clearly convenient to start with a sextupole of length ℓ_s placed beside each quadrupole of length ℓ_q. Looking at Eq. (5) we see that by setting the sextupoles, so that

$$k_0' \ell_s = k_0 \ell_q / D, \tag{6}$$

we can mutually cancel their chromatic effects making each pair an achromatic doublet. The entire FODO lattice is then achromatic. Clearly the closer the sextupoles are placed to the quadrupoles the more efficient this compensation becomes, and since the F- and D-sextupoles see the most extreme betatron amplitudes, this makes them more orthogonal for controlling the chromaticity in the two planes. Finally this arrangement also gives the most space for the dipoles, which is important in proton machines in order to squeeze in the maximum bending and is equally important in electron machines, but in that case it is to spread out the bending as much as possible to limit synchrotron radiation losses. We take this chromatically corrected lattice as our starting point.

Usually the lattice arcs contain many cells so that it will be possible to build up the amplitude of the w-vector slowly. This keeps the individual sextupole strengths to a minimum, reducing nonlinearities and resonance excitation. We will now divide the sextupoles up into

2(N+1) interleaved families, as shown in Fig. 1. The first group of (N+1) periods will contain the first members of each family, the second group the second members of each family and the third group the third members and so on.

$$\{S_F,S_D;S_{F1},S_{D1};S_{F2},S_{D2};\ldots;S_{FN},S_{DN}\}\{S_F,S_D;S_{F1},S_{D1};S_{F2},S_{D2};\ldots;S_{FN},S_{DN}\}$$
$$\leftarrow \quad \text{1st group} \quad \rightarrow \leftarrow \quad \text{2nd group} \quad \rightarrow$$
$$\{S_F,S_D;S_{F1},S_{D1};S_{F2},S_{D2};\ldots;S_{FN},S_{DN}\}\ldots\{S_F,S_D;S_{F1},S_{D1};S_{F2},S_{D2};\ldots;S_{FN},S_{DN}\}$$
$$\leftarrow \quad \text{3rd group} \quad \rightarrow \quad \leftarrow \quad \text{Last group} \quad \rightarrow$$

Fig. 1 Fully interleaved sextupole families

As described above we start with all the F-sextupoles equal and all the D-sextupoles equal and with the values set so as to compensate the natural chromaticity. If we now increment the SF family by $\Delta k'_{SF}$ we will see at the exit to the first member a step in the w-vector parallel to the a-axis. The lattice has been made achromatic, so this vector will rotate at twice the frequency of the betatron phase advance and with constant amplitude. We know that for the most efficient build-up in the amplitude of the vector, it is desirable that by the time it reaches the next member of its family (N+1) cells later the phase 2ψ has turned exactly $2n\pi$. This is illustrated in Fig. 2(a). If this is true for one family then it will be true for all families except that the vectors from successive families will be separated by twice the betatron phase advance between the sextupoles. Fig. 2(b) shows how the SF1-vector would look in a 60^0 FODO lattice and similarly Fig. 2(c) shows what the SF2-vector would look like.

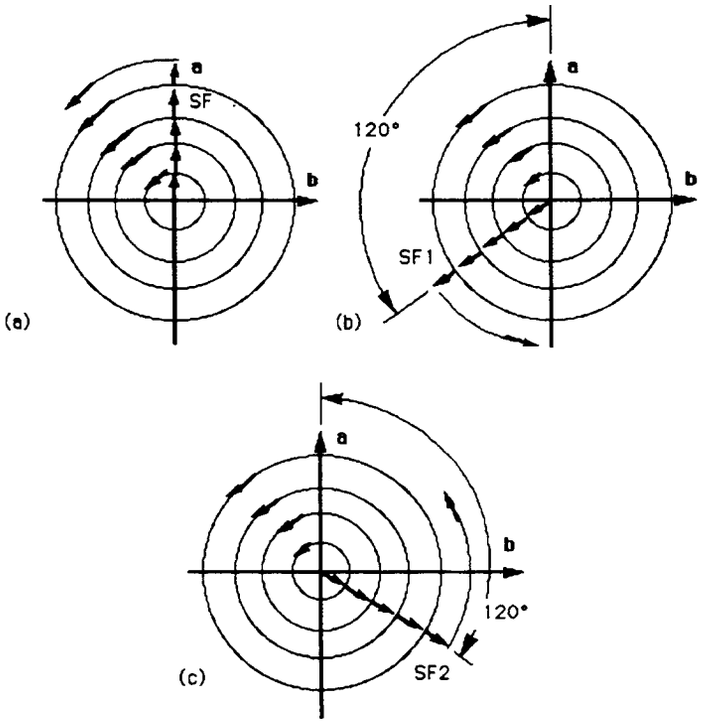


Fig. 2 Periodic build-up of w-vectors in a 60° phase advance FODO lattice
(a) SF vector, (b) SF1 vector, (c) SF2 vector

How many independent vectors do we need? The minimum should be six: two variables for w_x, two for w_z and a further two for the chromaticities Q'_x and Q'_z. Finally, we do not

want to excite the $3Q = p$ integer resonance, so subsequent members in a family should be in antiphase for this resonance condition and ideally there should be an even number of lenses. Summarising these conditions we find,

$$\left.\begin{aligned} 2(N+1)\mu_0 &= 2n\pi \quad \text{(phase condition)} \\ 3(N+1)\mu_0 &= (2m+1)\pi \quad \text{(anti - resonance condition)} \end{aligned}\right\} \tag{7}$$

where N, n and m are all integers.

Dividing the two equations gives $3n = (2m + 1)$ and since n and m are both integers n can only be odd. Equations (7) can therefore be solved to give the possible phase advances,

$\mu_0 = \pi/2$ for $n=1$, $N=1$ and $m=1$ i.e. 4 families

$\mu_0 = \pi/3$ for $n=1$, $N=2$ and $m=1$ i.e. 6 families

$\mu_0 = \pi/4$ for $n=1$, $N=3$ and $m=1$ i.e. 8 families

$\downarrow$ etc.

$\mu_0 = 3\pi/4$ for $n=3$, $N=3$ and $m=4$ i.e. 8 families

$\mu_0 = 3\pi/5$ for $n=3$, $N=4$ and $m=4$ i.e. 10 families

$\downarrow$ etc.

The most popular phase advance per cell is $\pi/3$ or 60° (as in Fig. 2). This satisfies all of our conditions and it is the lowest order case which gives the required six families. Another favoured case is the $\pi/2$ or 90° phase advance, because it provides stronger focusing, but with only four families it has insufficient variables to make a universal compensation scheme and a small trick has to be employed to provide the missing flexibility. The higher-order cases such as $3\pi/5$ or 108° phase advance will work perfectly well, but the 10 families are a little in excess of what is really needed. So we shall confine ourselves to looking more closely at the $\pi/3$ and $\pi/2$ possibilities. Figure 3 shows the relative positions of the SF-families in these two lattices.

Figures 3 (a), (b) and (c) correspond to the 60^0 lattice and Figure 3 (d) to the 90^0 lattice case. From Fig. 3(a), it is easy to see that no net w-vector will be excited if all three SF-sextupoles families are incremented equally. This action will, however, change the compensation of the lattice and cause a net change in the global tune spreads Q_x' and Q_z'. Naturally, the same applies if all D-families are equally incremented, but the values will be different since the lattice functions are different. This is precisely the situation we need for controlling the global values of the tune spreads Q_x' and Q_z', but without affecting w. If we now increment the current in just one family we will excite a net w-vector. In the $\pi/3$ case, we need only to increment the SF1 and SF2 families to reach anywhere in the ab plot while the SF family remains constant. In Figs 3(b) and (c), we are using the orthogonal sum and difference modes to move in the ab plot. Once the w-vector that we want is set, we can go back and retune the global tune spreads by incrementing all sextupoles of the same type equally. Thus the six 'handles' for adjusting the scheme are (SF2-SF1), (SF1+SF2), (SD2-SD1), (SD1+SD2) and equal increments in all sextupoles of the same type. In a symmetric FODO lattice the vectors from the D-sextupoles will sit midway in phase between those of the F-sextupoles. Each family will excite vectors in both planes but as with the global chromaticity the widely different betatron functions will make the F-families stronger in the horizontal plane and the D-families stronger in the vertical plane. Thus, a set of six equations will allow us to decide all settings for given parameters in a $\pi/3$ lattice.

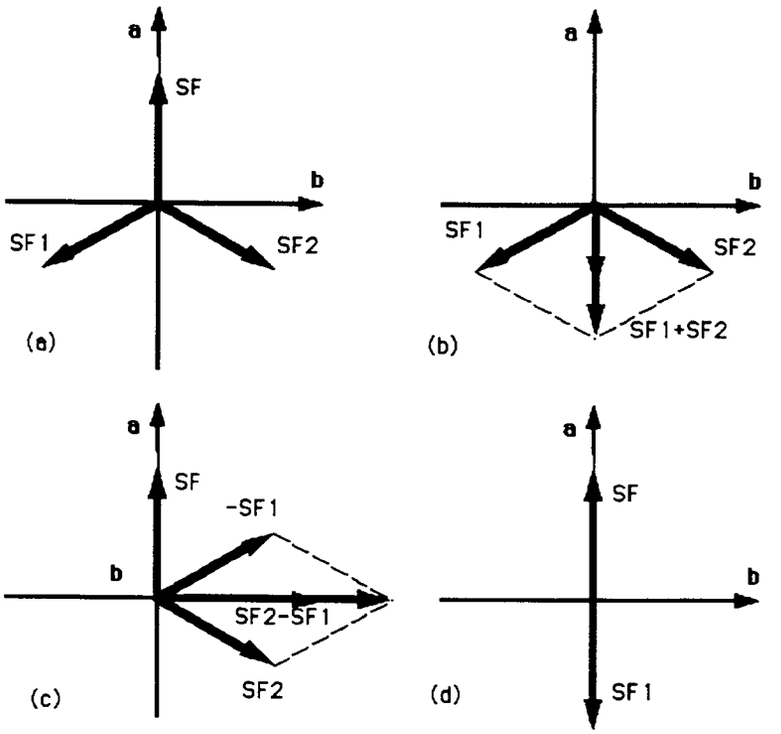


Fig. 3 w-vectors due to F-sextupoles in the $\pi/3$ and $\pi/2$ lattice cells

In the $\pi/2$ case [see Fig. 3(d)] the SF and SF1 are degenerate, so it is not possible to reach in any direction other than the a-axis. The same situation applies to the SD and SD1 families, but it is still possible to control the global tune spreads by equally incrementing all sextupoles of the same type. The missing flexibility is restored by conspiring with the position of the first sextupole and the linear lattice matching to bring the w error vector to the first sextupole with the correct phase i.e. parallel to the a-axis. This overcomes the problem of too few independent sextupole families, but the scheme is rather inflexible.

Figure 1 showed a fully interleaved layout for the sextupoles. An alternative layout is the non-interleaved scheme shown in Figure 4.

S_F ------ S_F ------ S_F ------ S_F ------	S_D ------ S_D ------ S_D ------ S_D ------
← SF family separated by $n\pi$ →	← SD family separated by $n\pi$ →
S_{F1} ------ S_{F1} ------ S_{F1} ------ S_{F1} ------	S_{D1} ------ S_{D1} ------ S_{D1} ------ S_{D1} ------
← SF1 family →	← SD1 family →
S_{F2} ------ S_{F2} ------ S_{F2} ------ S_{F2} ------	S_{D2} ------ S_{D2} ------ S_{D2} ------ S_{D2} ------
← SF2 family →	← DS2 family →

Fig. 4 Fully non-interleaved sextupole scheme with even numbers of sextupoles in each family and adjacent lenses in the same family separated by $n\pi$ in betatron phase

One feature of this layout is that we have abandoned the fully compensated achromatic lattice as the starting point (i.e. not all quadrupoles will have a sextupole). Also, unless the arcs are very long there will be relatively few sextupoles per family so that individual lens strengths will have to be increased.

The interleaved choice is the more natural and the more successful. The non-interleaved scheme was suggested on the reasoning that there would be fewer non-linearities if phase shifts between members of a single family were not perturbed by the non-linear influence of other interleaved sextupole families. For very small momentum errors, this proved to be true and dynamic apertures in the LEP machine for example are far greater with the non-interleaved scheme. However, once the momentum error is increased the aperture drops to zero far more quickly than for the interleaved scheme. This is attributed to the individual sextupole strengths being much higher and causing more resonance excitation. On balance the interleaved scheme is more effective.

One could also envisage partially interleaved schemes. One possibility would be interleaving the SF1 and SD1 families and the SF2 and SD2 families, while keeping these pairs of families non-interleaved. All other sextupoles could be connected to the SF and SD families to ensure a fully achromatic lattice. Since the betatron amplitudes are at their extreme values in the F and D lenses the SFs are relatively invisible to the SDs and the scheme has some similarity to the fully non-interleaved one. If we abandon the fully achromatic lattice the SF and SD families could be interleaved as a separate group further on in the lattice, so that we would have three non-interleaved pairs of interleaved families. Figure 5 illustrates the first of these possibilities.

Once the compensation scheme has been designed using the linear theory, a program such as HARMON [3] should be used to refine the results. HARMON minimises the excitation of a wide range of resonances and finds the best sextupole settings to equalise the various non-linearities. There is still no guarantee that the beam will be stable and it is normal to use one of the many tracking programs such as PATRICIA [4] to simulate the beam behaviour with random closed-orbit distortions and other errors. The aperture in which the beam is found to be stable is known as the *dynamic aperture*. A detailed review of the analytical and numerical tools available for calculating the dynamic aperture can be found in Ref. [5]. With strong low-β insertions and superconducting magnets, which are liable to have a rather poorer field quality than conventional magnets, it is sometimes difficult to have a dynamic aperture as large as the physical one.

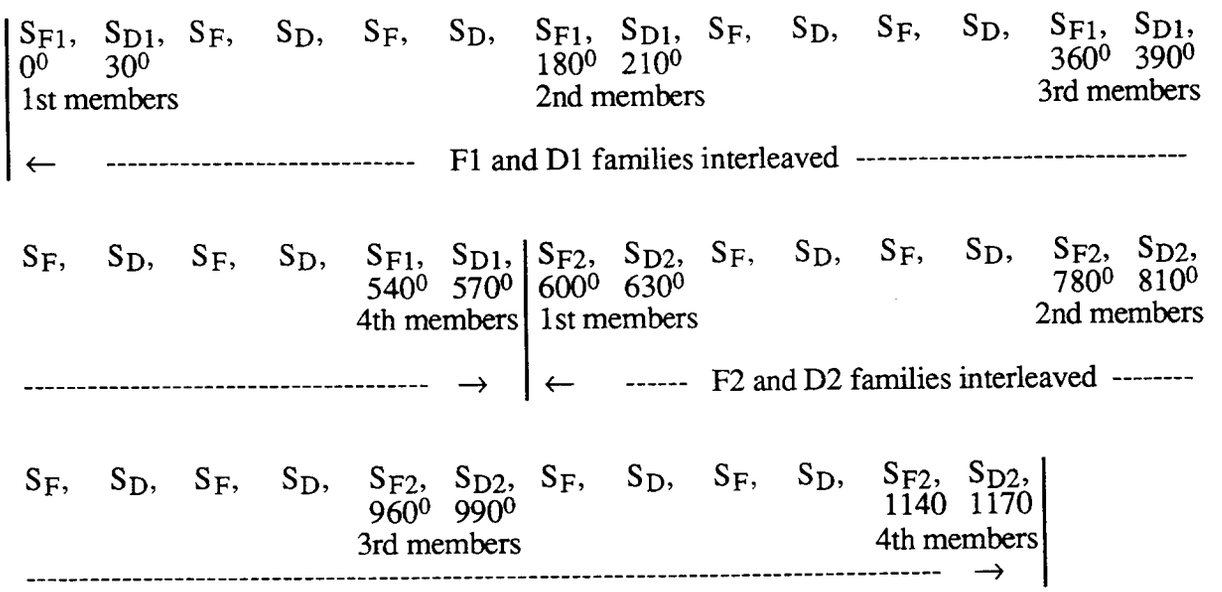


Fig. 5 Partially interleaved families with four members per family embedded in a chromatically compensated 60° lattice

REFERENCES

[1] B.W. Montague, *Chromatic Effects and Their First Order Correction*, Proc. CAS, Advanced Accelerator Physics School, Oxford, 1985, CERN 87-03, (1987).

[2] B. Autin, A. Verdier, *Focusing Perturbations in Alternating Gradient Structures*, CERN ISR-LTD/76-14 (1976).

[3] M. Donald, *Chromaticity Correction in Circular Acceleration and Storage Rings*, PEP Note 311, (July 1979), (program HARMON).

[4] H. Wiedemann, PEP Note 220 (1976) (program PATRICIA).

[5] G. Guignard and J. Hagel, Sextupole correction and dynamic aperture : numerical and analytical tools, *Particle Accelerators*, 18 (1986), 129-65.

DYNAMIC APERTURE

W. Scandale
CERN, Geneva, Switzerland

Abstract
Modern hadron colliders are made with high-field superconducting dipoles and quadrupoles that, in general, exhibit unintentional imperfections of the guiding field-shape due to construction tolerances, persistent currents, and saturation. Although the departure from linearity is typically small, a few units in 10^{-4} at 1 cm from the axis, this phenomenon has a profound influence on the single-particle beam dynamics. The non-linear fields are at the origin of two effects: the betatron tunes change with the amplitude and the momentum of the circulating particles, and, for certain combinations of the horizontal, vertical, and synchrotron tunes, non-linear resonances are excited. As a consequence, the stability of particle motion can be considerably reduced, up to the point that the dynamic aperture can become smaller than the physical aperture of the magnets. Analytical, numerical and experimental approaches have been used to evaluate and possibly compensate mechanisms which lead to particle losses.

1. INTRODUCTION

The original incentive to study the stability of the motion in non-linear dynamical systems is neither recent nor prompted by the necessities of accelerator design. In the second half of the last century, the development of Celestial Mechanics already provided most of the concepts that are applied nowadays to describe the orbital motion in particle accelerators. The mathematical development of this subject, which long preceded the age of computers, did in fact anticipate most of the quantitative features of non-linear dynamics, including the concept of dynamic aperture, that at present numerical simulations and experimental analysis give so much evidence for. The so-called 'region of stochasticity' of the phase space was in fact termed 'region of instability' in the late 1920s by G.D. Birkhoff [1], and the 'homoclinic points', around which the stochastic behaviour can be observed, were introduced in the last century by H. Poincaré [2]. The first part of my lecture will be devoted to schematically recaling the basic methods of non-linear dynamics routinely used in particle accelerators.

In the last decade, more 'pragmatic approaches' have been introduced by the accelerator designers to optimize the specifications and the cost of the superconducting guiding magnets for future hadron colliders [3]. Semi-analytical or purely numerical tools have been introduced to estimate the dynamic aperture as a function of various machine parameters. The effect of higher-order terms in the non-linear perturbation has been investigated with normal form techniques and related Lie algebra formalism that provide a natural generalization of the Courant–Snyder analysis for linear dynamics. On the other hand, the extraordinary progress of modern computers has made available a constantly extended numerical capacity, which allowed new simulation and analysis techniques to be explored. In the second part of this paper I shall describe how the stability of motion is investigated in modern superconducting accelerators, using as an example the case of the Large Hadron Collider (LHC) to be built in the LEP tunnel at CERN.

Despite the recent progress in understanding non-linear dynamics and in performing numerical simulations, our ability to predict the dynamic aperture in particle accelerators is quite limited. The theoretical methods available are not yet rigorously mastered, and the models for computer simulations are in general unable to include all the relevant phenomena characterizing the motion in particle accelerators. The operation of colliders may provide a wealth of

information that can be exploited to reduce the gap between computer models or analytical understanding and reality. However, extracting information in full detail on non-linear dynamics from working accelerators may be extremely difficult because of the complexity inherent in their modes of operation. To overcome this difficulty, dedicated experiments have recently been performed in already well-understood hadron colliders, like the CERN–SPS or the FNAL–Tevatron, in which clear experimental conditions have been established to eliminate spurious phenomena and to bring into evidence the short-term and the long-term effect on particle motion of well-known non-linear perturbations. The third part of this paper will be devoted to a review of the main results of these experiments.

2. THEORETICAL ASPECTS OF THE CONFINEMENT

2.1 Hamiltonian motion

A *n*-D Hamiltonian system has *n* degrees of freedom and its dynamical state is represented by a point *X* in a 2*n*-D phase space with generalized coordinates and momenta:

$$X = (q;p) = (q_1, p_1; q_2, p_2; \ldots; q_n, p_n). \tag{1}$$

The motion is described by an Hamiltonian *H* and obeys Hamilton's equations:

$$H = H(q_1, p_1; q_2, p_2; \ldots; q_n, p_n; t) \qquad \begin{cases} \dot{q}_i = \dfrac{\partial H}{\partial p_i} \\ \dot{p}_i = -\dfrac{\partial H}{\partial q_i} \end{cases} \tag{2}$$

The change of the Hamiltonian *H* with time is given by:

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial t} + \sum_{i=1}^{n} \dot{q}_i \frac{\partial H}{\partial q_i} + \dot{p}_i \frac{\partial H}{\partial p_i} = \frac{\partial H}{\partial t} \quad . \tag{3}$$

The second part of (3) comes from (2) and tells us that if the system is conservative, then *H* is independent of time and its value is conserved.

It has been shown by many authors [4–6] that the motion of charged particles in circular accelerators is governed by a 3-D Hamiltonian $H(x,p_x,y,p_y,t,-E;s)$ and that the dynamical state of the particle can be represented by a point in the 6-D phase space $(x,p_x,y,p_y,t,-E)$, where (x,p_x,y,p_y) are the transverse coordinates and their conjugate momenta and $(t,-E)$ is the third coordinate and its conjugate. In this representation, the independent variable of the motion is the azimuthal distance *s*, rather than the time *t*. The Hamiltonian is explicitly dependent on *s* and not on *t*, therefore the system is conservative.

It can be demonstrated in general that the motion of a system with a *n*-D time-dependent Hamiltonian is equivalent to that of a (*n*+1)-D time-independent Hamiltonian with an additional degree of freedom, expressed by the phase space variables $(t,-E)$.

On the other hand, a *n*–D time-independent Hamiltonian system can be described as a time-dependent system with (*n*–1)-D degrees of freedom, by choosing any generalized coordinate as the new time and the conjugate coordinate as the new time-dependent Hamiltonian.

2.1.1 Symplectic and area-preserving conditions

In a *n*–D Hamiltonian system, the Hamiltonian flow M is the operator that transports points in the phase space according to Hamilton's equations. It is usual to express it in terms of a map:

$$\begin{pmatrix} Q \\ P \end{pmatrix} = \mathtt{M} \begin{pmatrix} q \\ p \end{pmatrix}. \tag{4}$$

The Jacobian M is the derivative matrix of the map at the point of linearization:

$$\text{Jacobian:} \qquad \mathrm{M} = \left(\frac{\partial(Q_1, P_1; Q_2, P_2; \ldots; Q_n, P_n)}{\partial(q_1, p_1; q_2, p_2; \ldots; q_n, p_n)} \right) . \tag{5}$$

Using Hamilton's equations it is easy to prove that the Jacobian of a Hamiltonian map obeys symplectic conditions. This has been demonstrated, in particular, for particle motion in circular accelerators [7]. The usual way to write symplectic conditions is:

$$\mathrm{M}^{\mathrm{t}} \mathrm{J} \mathrm{M} = \mathrm{J} . \tag{6}$$

In the case of two degrees of freedom J is a 4×4 matrix:

$$\mathrm{J} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} . \tag{7}$$

The area enclosed by a loop C of the phase space is defined as the algebraic sum of the partial areas for (q_i,p_i). That area is also called the Poincaré integral. If the loop C moves in a Hamiltonian flow, then the Poincaré integral is an invariant of the motion and vice-versa if the area of all closed loops is conserved for all times, it is possible to find a single-valued Hamiltonian that governs the motion [8].

$$\text{Poincaré integral:} \qquad A_1 = \oint \sum_{i=1}^{n} p_i \, \mathrm{d} q_i \ . \tag{8}$$

$$\text{Area-preserving} \ \frac{\mathrm{d}A_1}{\mathrm{d}t} = 0 \ . \tag{9}$$

2.1.2 Eigenvalues, eigenvectors and linear stability

The eigenvalues and the associated eigenvectors of the Jacobian M obey the equation:

$$\mathrm{M}X = \lambda X . \tag{10}$$

The eigenvalues are the $2n$ characteristic roots of the $2n^{\text{th}}$ order equation for λ:

$$\det(\mathrm{M} - \lambda \mathrm{I}) = 0 \qquad \text{with I = identity matrix.} \tag{11}$$

In order to have stable motion for repeated applications of M, each eigenvalue must satisfy:

$$\lim_{k \to \infty} \lambda^k = \text{bounded value} \qquad \Leftrightarrow \qquad \left| \lambda^k \right| \le \rho \quad \text{with} \ \ \rho \in \Re \ \ \text{and} \ \ k \in \mathsf{N} . \tag{12}$$

The set of $2n$ vectors $X_i = (X_{ij})$ that obey Eq. (10) for the $2n$ eigenvalues λ_i are the eigenvectors. They can be used to diagonalize the Jacobian M:

$$\text{Diagonalization matrix:} \qquad \mathrm{A} = \left(X_{i,j} \right) \qquad \text{with } i, j = 1, \ldots, 2n \tag{13}$$

$$\text{Diagonalized Jacobian:} \qquad \Lambda = \mathrm{A}^{-1} \mathrm{M} \mathrm{A} \tag{14}$$

$$\text{Diagonalized eigenvectors:} \quad U_i = \mathrm{A}^{-1} X_i \qquad \text{with } i = 1, \ldots, 2n . \tag{15}$$

The symplectic conditions for M or Λ imply that eigenvalues must come in quadruples:

$$\lambda,\ 1/\lambda,\ \lambda^*,\ 1/\lambda^* \ . \tag{16}$$

Each pair λ, λ^* of eigenvalues corresponds to a normal mode of the system, i.e. a possible independent mode of behaviour.

The new Jacobian Λ is diagonal, and describes a linearized motion with n independent degrees of freedom.

For λ complex with $|\lambda| \neq 1$ and Im $\lambda \neq 0$, the quadruple is symmetrical with respect to the real axis and the unit circle (case of complex instability). For Im $\lambda = 0$, the quadruple becomes pairs lying on the real axis (hyperbolic case). For either of these cases the linearized motion is unstable. For $|\lambda| = 1$, the quadruple degenerates to a pair lying on the unit circle and the linearized motion is stable (elliptic case). For $\lambda = \pm 1$, there is marginal stability (parabolic case).

2.1.3 Liouville theorem

The Liouville theorem states that the n-D Hamiltonian flow M preserves the hypervolume in the $2n$-D phase space. This follows from the property that, in a Hamiltonian flow, the derivative of the state vector X is incompressible, i.e.:

$$\dot{X} = \left(\frac{\partial H}{\partial p_1}, -\frac{\partial H}{\partial q_1}, \ldots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial q_n} \right), \tag{17}$$

$$\mathrm{div}\left(\dot{X}\right) = 0 \ . \tag{18}$$

The Liouville theorem is important in statistical mechanics since it implies that the phase space density is invariant. It does not prevent an initially regular volume from being deformed by non-linear forces in an intricate form, imbedding a large fraction of empty space. In accelerators this phenomenon is called filamentation and for all practical purposes it implies an increase with time of the macroscopically observable region occupied by a particle beam. For instance, a kicked beam will exhibit an enlarged emittance in a profile monitor, since filled and empty parts of the phase plane are more and more intricately mixed with each other under the effect of non-linearities.

The hypervolume in the phase space can be obtained by constructing Poincaré's integral invariants of higher and higher hierarchy, the last one of which is the Liouville integral:

$$\begin{cases} A_1 = \oint \sum_{i=1}^{n} p_i \, \mathrm{d}\, q_i = \iint \sum_{i=1}^{n} \mathrm{d}\, q_i \, \mathrm{d}\, p_i \\ A_2 = \iint\iint \sum_{i,k=1}^{n} \mathrm{d}\, q_i \, \mathrm{d}\, p_i \, \mathrm{d}\, q_k \, \mathrm{d}\, p_k \\ A_n = \int \mathrm{d}\, q_1 \mathrm{d}\, p_1 \ldots \mathrm{d}\, q_n \, \mathrm{d}\, p_n \end{cases} . \tag{19}$$

The area-preserving property is equivalent to the Liouville theorem if $n = 1$; it is more demanding, instead, if $n > 1$.

2.1.4 Canonical transformations

A transformation of the phase space coordinates that preserves the form of Hamilton's equations is called canonical, or symplectic, or contact transformation.

The Hamiltonian flow described by Eq. (4) is a canonical transformation.

Canonical transformations preserve area (Poincaré invariants) and hypervolume (Liouville invariant).

An important example of canonical transformation consists in changing variables into action-angle variables of the linearized motion, see for instance Ref. [5].

2.1.5 Poisson bracket

The Poisson bracket of two dynamical variables F and G is defined as:

$$[F,G]=\sum_{i=1}^{n}\frac{\partial F}{\partial q_i}\frac{\partial G}{\partial p_i}-\frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q_i}\ . \tag{20}$$

This can be used to express in a more compact form the Hamilton's equations and the evolution of a dynamical variable with time:

$$\begin{cases}\dot{q}_i=\left[q_i,H\right]\\ \dot{p}_i=\left[p_i,H\right]\end{cases} \tag{21}$$

$$\dot{F}=[F,H]+\frac{\partial F}{\partial t}\ . \tag{22}$$

The Poisson brackets obey the following fundamental relations:

$$\left[q_i,q_j\right]=0 \qquad \left[p_i,p_j\right]=0 \qquad \left[q_i,p_j\right]=\delta_{i,j} \tag{23}$$

$$\text{Jacoby identity:} \qquad [A,[B,C]]+[B,[C,A]]+[C,[A,B]]=0\ . \tag{24}$$

Using Eqs. (23) and (24) it is possible to define a Lie algebra structure on the space of Hamiltonian vector fields [8] and to represent a Hamiltonian flow as a Lie series:

$$\mathrm{M}\ X=\mathrm{e}^{:H:}X=\sum_{i=1}^{\infty}(:H:)^i\frac{X}{i!}, \tag{25}$$

where the operator $:H:$ is defined in terms of Poisson brackets of the Hamiltonian:

$$\begin{cases}:H:X=\left[H,X\right]\\ (:H:)^2X=\left[H,\left[H,X\right]\right]\\ \ldots\\ (:H:)^iX=\underbrace{\left[H,\left[H,\ldots,\left[H,X\right]\right]\right]}_{i\ \text{times}}\end{cases} \tag{26}$$

2.1.6 Poincaré mapping

A natural way to correlate Hamiltonian flows and discrete maps is based on the concept of surface of section introduced by Poincaré [2].

In a n-D time-independent Hamiltonian system, the total energy H is conserved and this confines the motion in a so-called energy shell of the phase space, that is a $(2n–1)$-D set of the phase space. On the energy shell one of the phase space coordinates, say the momentum p_n, can be expressed as a function of the other coordinates, since:

$$H(q,p)=E_0\ . \tag{27}$$

The condition q_n = constant identify a $(2n–2)$-D surface in the energy shell, that represents a surface of section for the trajectory. The successive intersections of the trajectory with the surface of section define the Poincaré map that iteratively correlates points in the reduced phase space $(q_1,p_1;\ \ldots;q_{n-1},p_{n-1})$. The Liouville theorem implies that hypervolume in the reduced phase space is conserved. If additional independent invariants of motion exist, the

Poincaré section will lie on a surface of dimensionality less than (2*n*–2)-D, otherwise it will fill densely a bounded region of the (2*n*–2)-D phase space. Projections of the surface of section into the *n*–1 phase planes (q_i,p_i), with i = 1,..., *n*–1, are a useful way to visualize the motion. Separable Hamiltonians for the phase coordinates q_i, p_i imply area-preserving motion in the (q_i,p_i) phase plane and trajectory intersections lying on a smooth curve. Non-separable Hamiltonians, instead, generally produce annular shapes of finite area in any (q_i,p_i) phase plane projection.

In particle accelerators, intersections of 6-D trajectories with H = E_0 and $t = k \cdot T$, where T is the revolution period, provide the Poincaré map of the transverse motion at a given azimuthal position. The condition $t - t_0$ = constant, where t_0 is the time of arrival of the synchronous particle, say at the location of the RF cavity, provides a stroboscopic view of the transverse motion with the synchrotron frequency.

2.2 Integrable systems

A *n*-D dynamical system is integrable if it has *n* independent integrals of motion, i.e., in 2*n*-D action-angle phase space, the Hamiltonian has the general form:

$$H(J,\theta) = H_0(J) \qquad \text{with} \quad J = (J_1,...,J_n) \quad \text{and} \quad \theta = (\theta_1,...,\theta_n) \ . \tag{28}$$

The *n* actions J_i, with i = 1,..., *n*, are the *n* invariants from which the energy conservation can be deduced. Each action invariant reduces by one the dimensionality of the phase space set that contains the trajectory. The overall motion is thereby confined to a *n*-D invariant torus parametrized by the *n* action variables, on which the *n* angle variables run.

Rigorously speaking, invariant tori exist under the condition that the *n* integrals of motion be independent and in involution [8], that is, the following relations hold for all i and j:

$$\begin{cases} [H, J_i] = 0 & \text{invariance of } J_i \\ [J_i, J_j] = 0 & J_i, J_j \text{ in involution} \\ \sum_{i=1}^{n} a_i \, \mathrm{d}J_i = 0 \Rightarrow (a_1,...,a_n) = 0 & J_1,...,J_n \text{ independent} \end{cases} \tag{29}$$

For each pair of conjugate coordinates (θ_i,J_i) the motion is 2π-periodic for θ_i and follows closed curves, topologically equivalent to circles. The frequency of each mode is:

$$\omega_i(J_1,...,J_n) = \frac{\partial H}{\partial J_i}(J_1,...,J_n) \ . \tag{30}$$

A resonance occurs when the frequencies ω_i are rationally related by a set of *n* integers s_i not all equal to zero:

$$\sum_{i=1}^{n} s_i \, \omega_i = 0 \ . \tag{31}$$

In the generic non-resonant case, the trajectory is dense on the torus, i.e. it passes arbitrarily close to every point of the torus.

In the resonant case the trajectory closes onto itself after s_i revolutions of the phase θ_i, i.e. it is a periodic orbit, and the invariant torus degenerates into invariant tori of lower dimension. Since the integers s_i can be arbitrarily large, there are periodic orbits arbitrarily close to one another in action space.

In each (θ_i,J_i) phase plane, the Poincaré section of a resonant trajectory is a set of s_i parabolic fixed-points that depends on the initial condition of the angle. All of them lie on a circle parametrized by the resonant action invariant.

For non-linear systems there are trajectories with resonant frequency arbitrarily close to any trajectory with non-resonant frequency. This means that for an arbitrarily small change of the initial conditions, an orbit can move from a non-degenerate to a degenerate torus. This 'structural instability' plays the role of the devil in non-linear dynamics.

A simple example of a 2-D integrable system that performs linear motion is the time-independent oscillator with uncoupled degrees of freedom described by the Hamiltonian

$$H_0(J) = \omega_1 J_1 + \omega_2 J_2 \ . \tag{32}$$

An example of a non-linear system is the oscillator with amplitude-dependent frequencies described by the Hamiltonian:

$$H_0(J) = H_1(J_1) + H_2(J_2) \ . \tag{33}$$

The solution of the Hamilton equations is:

$$\begin{cases} \theta_1(t) = \theta_1(0) + \dfrac{\partial H_1(J_1)}{\partial J_1} t = \theta_1(0) + \omega_1(J_1) t \\ \theta_2(t) = \theta_2(0) + \dfrac{\partial H_2(J_2)}{\partial J_2} t = \theta_2(0) + \omega_2(J_2) t \end{cases} \tag{34}$$

The projection for t integer provides the so-called twist map:

$$\begin{cases} J_{1,k+1} = J_{1,k} \\ \theta_{1,k+1} = \theta_{1,k} + 2\pi\alpha\left(J_{1,k+1}\right) \\ J_{2,k+1} = J_{2,k} \\ \theta_{2,k+1} = \theta_{2,k} + 2\pi\alpha\left(J_{2,k+1}\right) \end{cases} \tag{35}$$

Invariant tori and their Poincaré projections in the phase plane (θ_1, J_1) for case (35) are illustrated in Fig. 1. Assuming given values of J_1 and J_2, the rotational frequencies and their ratio r are also fixed:

$$\begin{cases} \omega_1 = \omega_1(J_1) = 2\pi\alpha(J_1) \\ \omega_2 = \omega_2(J_2) = 2\pi\alpha(J_2) \\ r = \dfrac{\omega_1}{\omega_2} \end{cases} \tag{36}$$

For r irrational, the trajectory maps onto the entire torus surface and its projection maps onto the circle of radius J_1. For r rational, the torus degenerates into a closed curve and its section is a finite set of fixed points.

2.3 Non-integrable systems

A n-D integrable system slightly perturbed in such a way that the Hamiltonian is also a function of the angles is, in general, non-integrable. Its Hamiltonian has the general form:

$$H(J,\theta) = H_0(J) + \varepsilon H_1(J,\theta) \ . \tag{37}$$

The perturbation term $\varepsilon H_1(J,\theta)$ produces non-linear coupling and resonances among degrees of freedom, so that, near resonances, the regular solutions are strongly perturbed while the trajectories in the phase space undergo a change in topology. The singularity can be removed at an isolated resonance by an appropriate canonical transformation, as described for instance in Ref. [9] for particle dynamics.

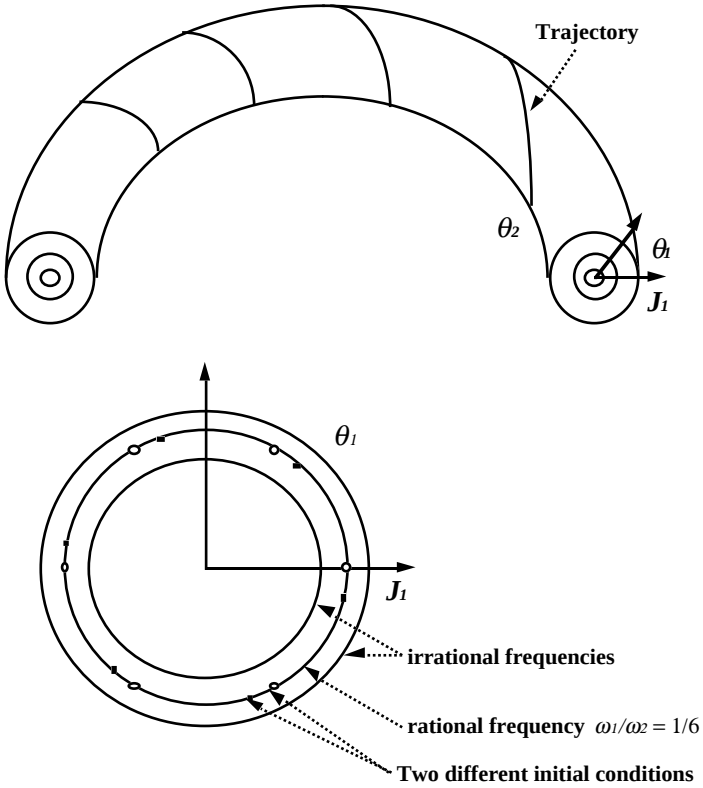


Fig. 1: Motion of an integrable system with two degrees of freedom.
Upper part: the motion lies on a torus J_1 = const. J_2 = const.
Lower part: Poincaré surface of section after a large number of trajectory intersections.

However, in a non-integrable system, the general problem of the confinement requires deep investigations on the interaction between resonances and on the existence and accessibility of invariant sets in the phase space.

2.3.1 The Kalmogorov–Arnold–Moser theorem

The Kalmogorov–Arnold–Moser (KAM) theorem states that in a *n*-D system with the Hamiltonian (37) some of the invariant tori of the unperturbed motion are preserved under sufficiently small and smooth perturbations, provided there is a sufficiently large non-linearity. Those invariant tori are preserved that are sufficiently far from all resonances in units of resonance width.

The perturbed torus is parametrized by n phase variables $\phi = (\phi_1,...,\phi_n)$ and by ε:

$$\begin{cases} J = I + V(\phi,\varepsilon) \\ \theta = \phi + U(\phi,\varepsilon) \end{cases} \tag{38}$$

U and V are periodic functions in ϕ that vanish for $\varepsilon = 0$, in which case, ϕ coincides with θ, the unperturbed phase of the motion.

The requirement of a sufficient non-linearity implies that, in the vicinity of non-resonant points of the phase space, there be some domain of J where the eigenfrequencies are linearly independent, or non-resonant:

$$\sum_{i=1}^{n} s_i \omega_i \neq 0 \qquad \text{with} \qquad \omega_i = \frac{\partial H_0}{\partial J_i} \ . \tag{39}$$

In the neighbourhood of resonances the perturbation induces a distortion of the orbit, and a variation of the action, which in turns modifies the frequency and brings the system out of resonance. This mechanism limits the overall change of the perturbed action and eventually allows perturbed KAM tori to stay in the vicinity of the unperturbed invariants. Since there are an infinite number of resonances, which approach sufficiently close to any point in the phase space, invariant tori are preserved far from low-order resonances provided the width of the neighbouring higher-order resonances converges sufficiently rapidly with the order s of the resonance. This requires the perturbing part of the Hamiltonian to be sufficiently smooth and differentiable for higher-order terms of the Fourier expansion to be small.

KAM invariants have diophantine frequencies:

$$\left|\vec{s}\cdot\vec{\omega}\right| \geq \gamma\left|\vec{s}\right|^{-\eta} \quad \text{with} \quad \vec{s} = \left(s_1,\ldots,s_n\right), \quad \vec{\omega} = \left(\omega_1,\ldots,\omega_n\right) , \tag{40}$$

with γ and η positive constants, the first depending on ε, the second on n.

Therefore, for increasing values of the perturbation, KAM invariants persist for those irrational frequencies that are hardest to be approximated by rational ones. The Chirikov criterion discussed below will clarify this point in terms of resonances and resonance width.

The KAM theorem, called in this form in recognition of the work of Kolmogorov, Arnold and Moser in the 1950s and 1960s, has an exceedingly small range of validity, which prevents any quantitative application to physical systems and in particular to beam dynamics in particle accelerators. However, its proof brings to our attention the crucial assessment that regular motion can be observed with certain restrictions also in non-integrable systems.

2.3.2 The Poincaré–Birkhoff theorem

The Poincaré–Birkhoff theorem applies to 2-D non-integrable systems and gives a deep insight on the breaking mechanism of invariant curves.

Let us consider as an example the 2-D oscillator with amplitude-dependent frequencies described by the Hamiltonian (33), to which a small non-integrable perturbation is added. The surface of section with $H = E_0$ and θ_2 = constant, identify a Poincaré map. The unperturbed motion is characterized by the ratio r of the two linearized frequencies, in particular, rational values of r characterize fixed points of the map:

$$r\left(J_1, J_2\right) = \frac{\omega_1}{\omega_2} = \frac{q}{s} \ . \tag{41}$$

Equations (41) identify a circle of fixed points of order s, in the phase plane (θ_1, J_1).

The Poincaré–Birkhoff theorem states that generically $2s$ of them survive the perturbation and appear in chains of alternating elliptic and hyperbolic fixed points. Regular phase space trajectories encircle the elliptic points and a separatrix trajectory connects the hyperbolic points. The typical pattern is that of a chain of islands. Orbits in the neighbourhood of elliptic fixed points in general reveal higher-order resonances and fixed points which have similar characteristics to those just analysed above, but on a finer scale. The driving term is proportional to $1/s!$, the order of the considered resonance. Orbits in the vicinity of hyperbolic fixed points are the connection of four curves, two of them are incoming separatrices, the two others are outgoing separatrices. Incoming and outgoing separatrices interconnect the set of hyperbolic fixed points of a resonance. The movement along the separatrices becomes increasingly slow as an hyperbolic fixed point is approached. Incoming and outgoing separatrices intersect each other in an intricate way, and in the vicinity of hyperbolic fixed points the number of the intersections, also called homoclinic points, diverges to infinity. In addition, along the separatrix there is an infinity of secondary resonances corresponding to multiples of s. Each secondary resonance has its own set of alternating elliptic and hyperbolic points, and its own separatrices. Separatrices of primary and secondary resonances in general follow wild trajectories that intersect each other in heteroclinic points. Homoclinic and heteroclinic points fill densely the space surrounding the circle of fixed points of the unperturbed system,

characterized by Eq. (41). In that region KAM tori cannot exist: the phase-space trajectory is generically stochastic, since it changes topology at each singularity crossing. The size of the stochastic region around the separatrix grows as the perturbation amplitude increases, however, there is not an abrupt transition to stochasticity for some critical value of the perturbation.

2.3.3 The Chirikov criterion of overlap of the resonances

Chains of resonant islands have a characteristic width that increases with the perturbation strength and that can be approximately estimated by an appropriate expansion technique [9]. For small perturbations, neighbouring chains of different order resonances are eventually well separated and can be described as isolated resonances. Chirikov suggested [10] that the largest perturbation parameter for which adjacent primary chains of islands overlap is a good estimate for the value of the perturbation required to destroy the last KAM invariant between these islands. This criterion has an intuitive appeal, since we already know that regions around separatrices are in fact stochastic, therefore we cannot expect regular motion in between two chains of resonances whose separatrices intersect each other.

However, the above criterion is too severe a condition for stochasticity. In fact, the last KAM torus is destroyed well before resonance overlapping occurs, because of the interaction between the two primary resonances that modify the topology of the phase space, and because of the existence of secondary resonances lying between the primary ones.

The criterion has to be used with care since higher-order perturbations may have to be estimated to find the appropriate resonance width. Whenever it holds, there is a transition from local to global stochasticity. In other terms, stochastic areas surrounding isolated islands become interconnected over a large fraction of the phase space, where the motion is no longer regular.

2.4 Definition of dynamic aperture

Rigorously speaking, the dynamic aperture is the innermost radius of the region in the phase space, where the motion is stable. Trajectories with initial conditions in this domain remain confined for ever. Posed in this form, the question of the stability of the motion can only be tackled by a purely mathematical approach that apparently has little if any practical significance. The kind of stability we are talking about can be defined as follows:

$X(t)$ is stable if for all t and for any $\varepsilon > 0$ there is a $\delta > 0$ such that:

$$|X(0)| < \delta \qquad \Rightarrow \qquad |X(t)| < \varepsilon \ . \tag{42}$$

For practical applications the concept of confinement has to be defined within an appropriate time scale. In particle accelerators, two regimes of interest can be identified in terms of the number of stable revolutions along the ring.

The short-term dynamic aperture requires confinement for 10^3 to 10^4 revolutions. This is relevant for transient phenomena such as the injection process or beam observations. At injection, the initially large amplitude of oscillation due to steering errors is in general reduced with feed-back devices: confinement at large amplitude for about a damping time is thus sufficient to avoid particle losses. On the other hand, dedicated experiments are generally planned to explore the phase-space parameters as a function of the amplitude. Measurements, which in general require up to a few thousand turns, can be performed up to the short-term dynamic aperture.

The long-term dynamic aperture instead concerns steady-state regimes lasting 10^7 to 10^9 revolutions, like injection or collision plateaux.

Phenomena that cannot be treated by Hamiltonian mechanics, like external noise, dissipation, quantum effect, and gas scattering in general play an essential role in the investigation of short-term and long-term dynamic aperture. Nevertheless, the abstract problem of the Hamiltonian confinement has a strong significance for the practical confinement.

2.4.1 Phase space in an accelerator with non-linear elements

In a linear machine with irrational tunes the motion is stable and regular all around the closed reference orbit near the magnetic axis.

The non-linear fields add a tune dependence with the amplitude, which shifts tunes to rational values, provoking resonant phenomena accompanied in the phase space by islands of finite area surrounded by thin chaotic layers. The islands and the chaotic layers exist throughout the entire phase space. However, at small amplitude, trajectories follow invariant surfaces, the KAM tori, and remain stable for indefinite time. As the amplitude increases, the islands become larger until they overlap. Analytically the chaotic limit is well defined for phenomena with 2-D Hamiltonian only since, in these cases, the domain with regular motion is foliated in 2-D tori that encircle the origin and any existing layer of isolated higher-order resonances. When the resonances overlap, the chaotic layers become interconnected and the particle motion is no longer bounded. With more degrees of freedom, the tori have higher dimensions and therefore no longer form an impassable barrier. Particles in stochastic layers, even close to the origin, may escape through the entire phase space, owing to the so-called Arnold diffusion. However, for all practical purposes, the border between mostly regular and mostly chaotic trajectories still exists and defines the long-term dynamic aperture. The confinement inside this border requires weak perturbations and is limited by local-scale stochasticity. Outside the long-term dynamic aperture, the motion is globally stochastic until, for some value of the amplitude, a separatrix going to infinity is encountered that makes the motion totally unstable. Particles with initial conditions in the large-scale stochastic area have totally erratic trajectories; however, they generally survive for a decreasing number of turns as the initial amplitude increases. The short-term dynamic aperture is the border of the region where the survival time is still comparable to the duration of transient phenomena in particle accelerators.

The topological properties of the phase space described here are pictorially illustrated in Fig. 2.

2.4.2 Basic procedures for confinement

The resonances play a crucial role for both theoretical and practical confinement problems. Therefore the basic step to finding dynamic apertures requires the careful identification of all the resonances near the working region of interest in the phase space.

The basic way to enlarge the dynamic aperture is to reduce as much as possible the resonance strength, i.e. to make as small and as smooth as possible the perturbing non-linear terms.

In particular, reducing the strength of low-order resonances and choosing the working eigenfrequencies far from them increases the short-term dynamic aperture. Smearing out the perturbation reduces the width of higher-order resonances, and increases the stability domain.

Residual external noise produces unwanted modulations of the eigenfrequencies and satellite resonances which in turn enlarge the stochastic area close to separatrices and reduce the threshold for overlapping criterion. Powerful noise reduction methods are needed to improve confinement.

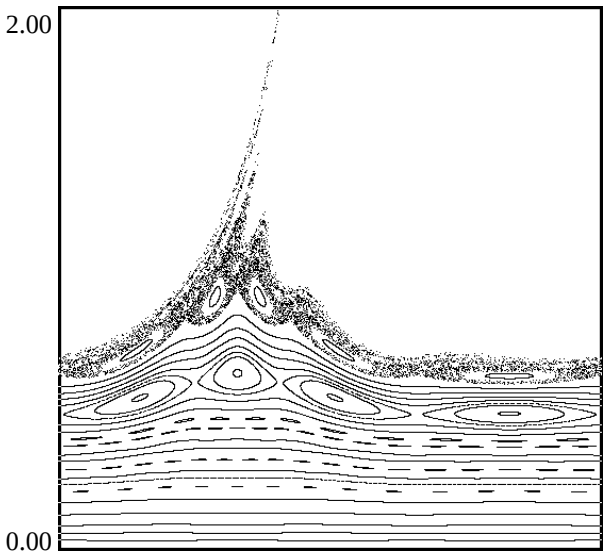
2.4.3 Stability domain of the Hénon map

The Hénon map [11] is a simple 1-D non-linear system that can be used to illustrate some of the concepts relative to confinement. In terms of accelerator physics, the Hénon map corresponds to the horizontal motion in a FODO cell with one sextupole, in thin-lens approximation.

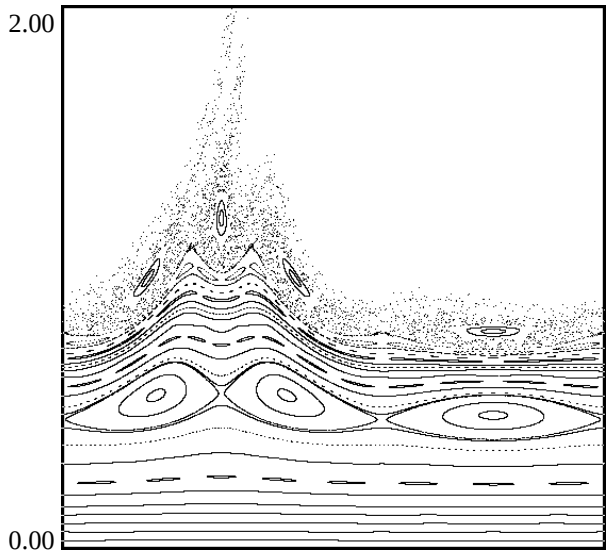
The Hamiltonian is:

$$H(x,p,s) = \omega^2 \frac{x^2}{2} + \frac{p^2}{2} - \omega \frac{x^3}{3} \sum_{k=-\infty}^{\infty} \delta(s - k\ell) \qquad \text{with} \quad \frac{\omega}{2\pi} \in \left]0, \frac{1}{2}\right[\,, \tag{43}$$

where ω is the frequency of the linearized motion and ℓ is the length of the cell.



a) Betatron frequency near the 1/6 resonance



b) Betatron frequency near the 1/5 resonance
Abscissa: angular coordinate in units of 2π
Ordinate: action coordinate in arbitrary units

Fig. 2: Phase space topology of a non-linear system.

The equations of motion are:

$$\begin{cases} \dot{x} = p \\ \dot{p} = -\omega^2 x + \omega x^2 \sum_{n=-\infty}^{\infty} \delta(s - k\ell) \,. \end{cases} \tag{44}$$

The Hénon map solves them iteratively turn after turn:

$$\begin{pmatrix} x_{k+1} \\ p_{k+1}/\omega \end{pmatrix} = \mathrm{R}(\omega T)\left(\begin{pmatrix} x_k \\ p_k/\omega \end{pmatrix} + \begin{pmatrix} 0 \\ x_k^2 \end{pmatrix}\right) \quad \text{with} \quad \mathrm{R}(\omega t) = \begin{pmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{pmatrix}; \tag{45}$$

where x_k, p_k are the position and the conjugate momentum, respectively, at the location ℓ and at the times $k \cdot T$, where T is the transit time through the cell.

Introducing normalized coordinates $\bar{x}_k = x_k$, $\bar{p}_k = p_k/\omega$, and assuming T as the unit time, the linearized part of the Hénon map can be expressed as:

$$\mathrm{L}_k\begin{pmatrix} \bar{x}_k \\ \bar{p}_k \end{pmatrix} = \mathrm{R}(\omega)\begin{pmatrix} 1 & 0 \\ 2\bar{x}_k & 1 \end{pmatrix} \circ \begin{pmatrix} \bar{x}_k \\ \bar{p}_k \end{pmatrix}. \tag{46}$$

The Hénon map (45) has two fixed points, one of which is the origin:

$$\begin{pmatrix} \bar{x}_\mathrm{S} \\ \bar{p}_\mathrm{S} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} \bar{x}_\mathrm{U} \\ \bar{p}_\mathrm{U} \end{pmatrix} = \begin{pmatrix} 2\tan\frac{\omega}{2} \\ -2\tan^2\frac{\omega}{2} \end{pmatrix}. \tag{47}$$

From the inspection of the trace of L, the origin is found to be stable and the second fixed point unstable, for any value of the frequency.

$$\mathrm{Tr}(\mathrm{L}_k) = 2\cos\omega + 2\bar{x}_k\sin\omega \Rightarrow \begin{cases} \mathrm{Tr}\left(\mathrm{L}_{\bar{x}_k=\bar{x}_\mathrm{S}}\right) = 2\cos\omega \in]-2,+2[\Rightarrow \text{ stable} \\ \mathrm{Tr}\left(\mathrm{L}_{\bar{x}_k=\bar{x}_\mathrm{U}}\right) = 2 + 4\sin^2\left(\frac{\omega}{2}\right) \in]2,6[\Rightarrow \text{unstable} \end{cases} \tag{48}$$

The phase space portrait resulting from the iteration of the Hénon map exhibits complicated structures, as in Fig. 3, where the eigenfrequency is chosen in the vicinity of the fourth-order resonance. The dashed area around the origin and inside the four resonant islands is stable. The winding curve that encircles the stable area is the separatrix, obtained by iterating repeatedly a point initially close to the unstable fixed point. A large multitude of homoclinic points appears, that announces chaotic motion. Above the separatrix the motion is unstable. Chains of higher-order islands are visible as the phase space is explored at a finer scale, as in Fig. 4, where the eigenfrequency is chosen in the vicinity of the fifth-order resonance.

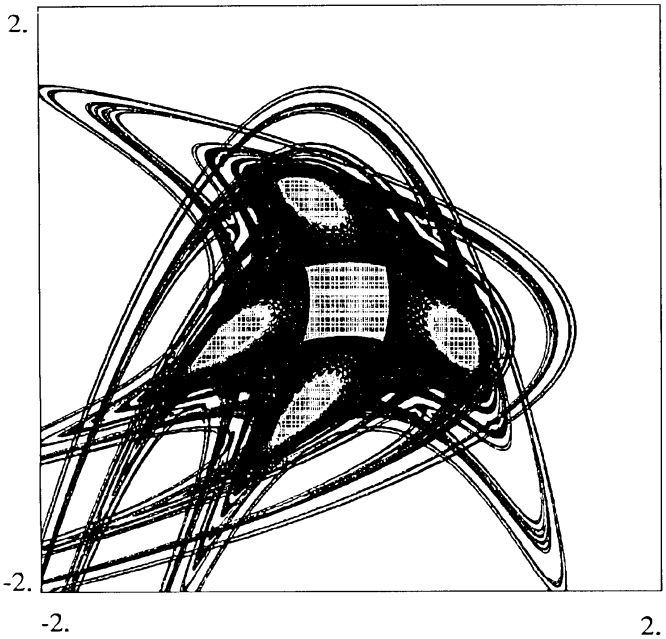


Fig. 3: Phase space portrait of the Hénon map near the 1/4 resonance.

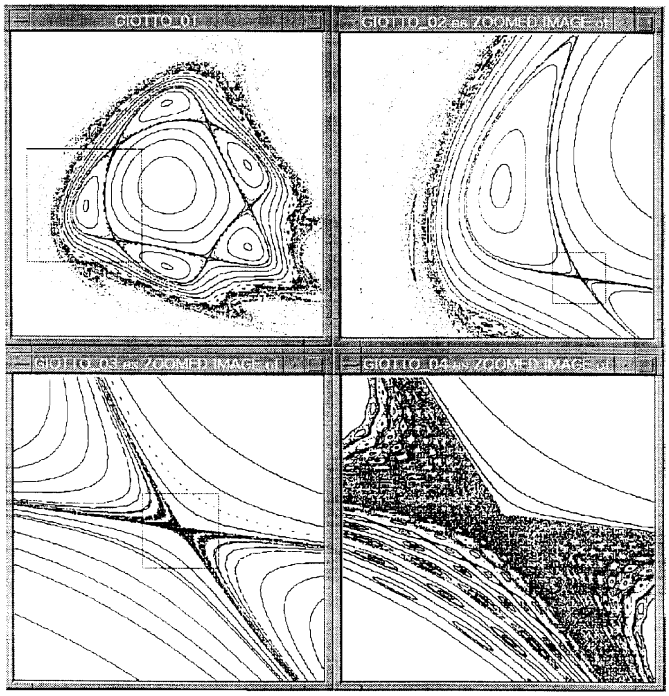


Fig. 4: Phase space portrait of the Hénon map near 1/5 resonance.

The standard way to find the dynamic aperture requires iteration of an appropriate number of the Hénon map for increasing values of the initial amplitude [12]. The domain of stability generally shrinks as the eigenfrequency approaches rational values, as in Fig. 5.

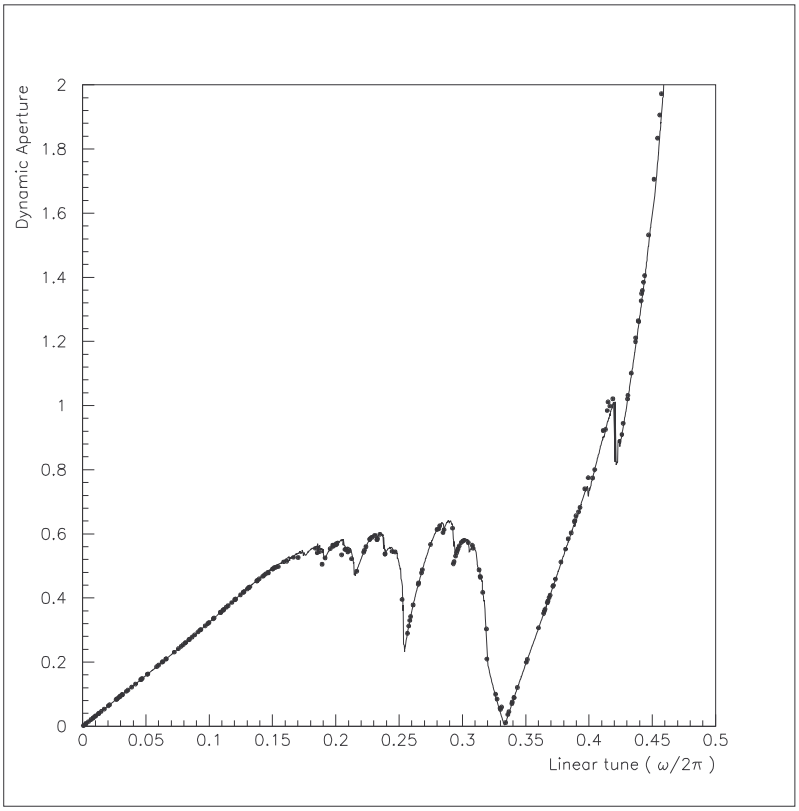


Fig. 5: Dynamic aperture of the Hénon map.

3. CONFINEMENT IN PARTICLE ACCELERATORS

The motion of charged particles in circular accelerators is basically governed by the magnetic field of the guiding dipoles and the focusing quadrupoles. Intentional and non-intentional non-linear fields are in general also present, the side-effect of which is to induce losses at large amplitude. Sextupoles are used to reduce the chromaticity, and octupoles make the tune dependent on the amplitude, which is sometime exploited to improve the current-dependent behaviour. In hadron accelerators, the destabilizing action of chromaticity sextupoles is self-compensated to a large extent due to the regularity of the lattice. However, a strongly focused lattice is usually necessary to reduce the sensitivity to field errors, and this in turn increases the strength of chromaticity sextupoles. Unintentional multipoles due to unavoidable imperfections of the guiding and focusing fields introduce additional non-linearities, which represent the greatest hazard. Longitudinal fringeing fields, instead, play a non-negligible role in small machines only, therefore, hard-edge representation of the magnetic fields along the azimuth is considered a valid approximation for large accelerators. The situation can be very different for different machine types. For instance, in modern electron accelerators, the chromatic aberration is even larger, due to the stronger focusing required to reduce the transverse beam size. Consequently, sophisticated sextupole corrector schemes are used, that include elements in zero-dispersion regions to partly compensate resonances, whilst strongly non-linear insertion devices, like wigglers and undulators, are often present.

Practical methods to study the confinement have to be adapted to the specific case under investigation. It is outside of the scope of this lecture to give an exhaustive overview of the known problems and of the proposed solutions. Instead, we will limit ourselves from now on to discussing the confinement in large hadron accelerators with superconducting magnets, even if some of the concepts illustrated in this section are of a more general and wider applicability.

In hadron colliders of the new generation, compromises must be found between making magnetic fields as uniform as possible and keeping magnet costs low. This is a difficult challenge for superconducting magnets, whose quality depends on the mechanical tolerances of the coil geometry, rather than on the shape of the poles. Both in the Tevatron and in the Hera magnets, typical deviations from uniformity have been limited to about one part in ten thousand at 2.5 centimetres from the magnet axis. Similar values, properly extrapolated with the inner coil diameter and the superconducting filament size, are expected to be reached in the magnets of the SSC, the LHC, and RHIC. Much smaller imperfections are in general present in warm magnets.

The single-particle approach provides a sufficiently simple, reliable, and coherent model of the real accelerator to investigate performances related to non-linear dynamics. The key issue is to estimate the stability of the motion over the operational cycle of the accelerator. Both analytical and numerical tools are used to estimate the dynamic aperture as a function of various machine parameters. Improvements of the linear lattice and correction schemes are studied to reduce the influence of the non-linear forces, and to specify upper limits for the magnet imperfections. The final validation is, in general, performed with numerical simulations in which the particle position is tracked element by element around the machine for large numbers of turns. In this section two subjects are reviewed: the tools by which predictions on beam stability are formulated, and the applications on accelerator design.

3.1 Tools for dynamic aperture estimates

In hadron accelerators non-linear problems are usually addressed for 2-D transverse degrees of freedom, neglecting the coupling with longitudinal motion. For numerical studies and computer tracking simulations, more realistic 3-D models are preferred.

3.1.1 Transfer map for a single magnet

Charged particles travelling through linear magnets obey the 2-D equations of motion:

$$\frac{\mathrm{d}\vec{z}}{\mathrm{d}s} + \mathrm{K}(s)\vec{z} = 0, \tag{49}$$

with:

$$\vec{z}(s)=\begin{pmatrix} z_1(s)\\ z_2(s)\\ z_3(s)\\ z_4(s)\end{pmatrix}=\begin{pmatrix} x(s)\\ p_x(s)\\ y(s)\\ p_y(s)\end{pmatrix},\quad \mathrm{K}(s)=\begin{pmatrix} 0 & 1 & 0 & 0\\ k_1(s)+\rho^{-2}(s) & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & -k_1(s) & 0\end{pmatrix}. \tag{50}$$

In single dipoles or quadrupoles, K is constant with respect to *s*, therefore Eq. (41) can be analytically solved piece-wise, see for instance Ref. [13]. The solution is in general expressed as a linear transport matrix L, that obeys the usual composition rules:

$$\vec{z}(s)=\mathrm{L}(s,s_0)\vec{z}(s_0)\ , \tag{51}$$

$$\mathrm{L}(s_0,s_0)=\mathrm{I},\ \text{the identity map}, \tag{52}$$

$$\mathrm{L}(s_1,s_0)=\mathrm{L}(s_1,s)\circ\mathrm{L}(s,s_0)\ . \tag{53}$$

The matrix L also satisfies the linear equation:

$$\frac{\mathrm{dL}(s,s_0)}{\mathrm{d}s}=\mathrm{K}(s)\circ\mathrm{L}(s,s_0)\ , \tag{54}$$

the general solution of which can be written as:

$$\mathrm{L}(s_0+\ell,s_0)=e^{\ell\mathrm{K}}=\sum_{i=0}^{\infty}\ell^i\frac{\mathrm{K}}{i!}^i. \tag{55}$$

In an extended multipolar magnet with non-linear field-shape components a forcing term is present in the equation of motion.

$$\frac{\mathrm{d}\vec{z}}{\mathrm{d}s}+\mathrm{K}(s)\vec{z}=\vec{f}(x,y;s)\quad\text{with}\quad \vec{f}(x,y;s)=\begin{pmatrix} 0\\ f_x(x,y;s)\\ 0\\ f_y(x,y;s)\end{pmatrix}. \tag{56}$$

The perturbation $\vec{f}(x,y;s)$ is expressed in terms of normal and skew multipolar strengths up to order *m*, called k_l and j_l respectively, that correspond to the relative field-shape deviations b_l and a_l, respectively, measured at a distance *R* from the magnetic axis.

$$\begin{cases} f_x(x,y;s)=\mathrm{Re}\left[\sum_{l=2}^{m}\frac{1}{l!}(k_l+ij_l)(x+iy)^l\right]=\mathrm{Re}\left[B_0\sum_{l=2}^{m}(b_{l+1}+ia_{l+1})\left(\frac{x+iy}{R}\right)^l\right]\\ f_y(x,y;s)=-\mathrm{Im}\left[\sum_{l=2}^{m}\frac{1}{l!}(k_n+ij_l)(x+iy)^l\right]=-\mathrm{Im}\left[B_0\sum_{l=2}^{m}(b_{l+1}+ia_{l+1})\left(\frac{x+iy}{R}\right)^l\right]\end{cases} \tag{57}$$

In general, Eq. (56) cannot be analytically solved. Approximate solutions can be found, instead, either by neglecting higher-order contributions to the particle trajectory, or by localizing the entire non-linear content of the multipolar components in a few points along the azimuth.

In the first approach, called thick-lens approximation, approximate trajectories are found across extended magnets by a recursive method, which initially starts from the linear approximation:

$$\begin{cases} \vec{z}_1 = \mathrm{L}(s,s_0)\vec{z}(s_0) = \mathrm{L}(s,s_0)\vec{z}_0 \\ \vec{z}_k(s) = \mathrm{L}(s,s_0)\vec{z}_{k-1}(s_0) + \int_{s_0}^{s} \mathrm{L}(s',s_0)\vec{f}\left[x_{k-1}, y_{k-1}; s'\right]\mathrm{d}s' \,. \end{cases} \tag{58}$$

At each step *k*, the final coordinates are expressed as a polynomial expansion of order *k* in the initial conditions. The resulting map is called the Taylor map of order *k*:

$$z_{j,k}(s_0+\ell) = \sum_{i=1}^{k} \sum_{(i_1+i_2+i_3+i_4)=i} c_{j;i_1,i_2,i_3,i_4} z_1^{i_1}(s_0) z_2^{i_2}(s_0) z_3^{i_3}(s_0) z_4^{i_4}(s_0) \qquad \text{for } j = 1,\ldots,4\,. \tag{59}$$

At the order *k*, the symplectic conditions (6) are violated by terms of order $\left|\vec{z}(s_0)\right|^{k+1}$.

The second approximation method for solving Eq. (56) is based on the assumption that the non-linear perturbation is concentrated in one or more locations inside the magnet. Usually, the non-linearity is localized in the middle of the magnetic length:

$$\vec{f}(x,y;s) = \vec{f}(x,y;s)\ell\delta(s-s*) \qquad \text{with} \quad s^* = s_0 + \frac{\ell}{2}\,. \tag{60}$$

The transfer map, also called the thin-lens approximation, becomes:

$$\vec{z}(s_0+\ell) = \mathbb{K}\left(s^*\right)\vec{z}(s_0) \qquad \Leftrightarrow \qquad \begin{cases} \vec{z}\left(s^*\right) = \mathrm{L}\left(s^*, s_0\right)\vec{z}(s_0) \\ \vec{z}(s_0+\ell) = \mathrm{L}\left(s_0+\ell, s^*\right)\left[\ell\vec{f}\left(x^*, y^*; s^*\right) + \vec{z}\left(s^*\right)\right]. \end{cases} \tag{61}$$

Terms of the order of ℓ^2/ρ^2 are neglected. The symplecticity (6) is exactly obeyed.

3.1.2 Tracking simulations

As seen in the previous section, a realistic description of the accelerator structure, fully adequate to compute single-particle trajectories, is difficult, if not impossible. Approximations must be tailored to the specific problem under investigation. Simplifications are also imposed by limitations in computing power.

Simulations with thin-lens approximation and symplectic integrators of the equation of motion, like Eq. (61), are considered as the master tool for quantitative estimates of particle behaviour in large accelerators like the LHC, since the length of individual magnets is small compared with the accelerator radius. They provide symplectic solutions for a dynamical system that approximates sufficiently well the entire accelerator. A sequence of linear transfer matrices interleaved with localized polynomial non-linearities should be computed. Reliable results are easily obtained since computer rounding errors can be kept under control [14]. However, vast computing power is required to get reliable estimates of the dynamic aperture as a function of various lattice and beam parameters. The thin-lens description is usually extended to guiding and focusing fields and, in general, does not imply relevant changes to orbit functions.

There are several computer codes by which large accelerators can be described. In the case of the LHC, two of them are routinely used for numerical simulations: MAD [15], developed at CERN, and SIXTRACK [16], developed at DESY. Both of them have scalar versions to be processed in the modern farms of workstations as well as vectorized versions to make use of modern parallel processors.

3.1.3 Maps

The use of non-linear maps in single-particle beam dynamics has been intensively investigated in the last fifteen years. It is outside the scope of this paper to present this activity in an exhaustive manner. Excellent review articles are available, like those in Refs. [17,18].

Here we shall limit ourselves to describing those concepts that have been applied in the design of the LHC.

In linear lattices, particle coordinates can be propagated along the accelerator azimuth by Twiss transfer matrices. The use of maps can be extended to non-linear dynamics with some precautions. This extension, originally motivated by the need to speed-up long-term tracking simulations in hadron colliders, in fact provides a powerful tool for handling dynamical quantities, like the tune dependence with the amplitude and the momentum, the distortion functions and the smear, the higher-order non-linear invariants, and finally the Fourier harmonic coefficients of the resonance driving terms. Non-linear matrices can be constructed very efficiently with differential algebra techniques using Taylor expansion to some higher order of algebraic operators [19]. One-turn Taylor maps resulting from the composition of all the linear and non-linear elements in the accelerator are inherently not symplectic because of the higher-order truncation, therefore inappropriate to preserve the volume in phase space. One of the ways to restore symplecticity is to replace the truncated map with a normal form [20–23], that is an integrable map, represented by a rotation of an angle depending on the amplitude of the orbit. Mathematically this operation implies the neglect of the non-integrable part of the map and the finding of a polynomial transformation of coordinates that conjugate the original map to a simpler and more symmetric map, which has explicit invariants, and explicit iteration formulae. The conjugation equation of a given map with its normal form is formally the same as the similarity transformation that brings a matrix to its diagonal form:

$$\Phi(\varsigma)^{-1}\mathrm{M}\,(Z_\mathrm{c})\Phi(\varsigma)=\mathrm{U}(\varsigma)\ . \tag{62}$$

Courant–Snyder coordinates for M are complex, i.e. $Z_\mathrm{c} = (x\text{-}ip_x, x+ip_x, y\text{-}ip_y, y+ip_y)$, in the case of 2-D degrees of freedom. The new coordinates are $\varsigma=\left(\varsigma_1,\varsigma_1^*,\varsigma_2,\varsigma_2^*\right)$.

The transformation from the new to the old coordinates is tangent to the identity, since:

$$Z_\mathrm{c}=\Phi(\varsigma)=\varsigma+\sum_{i\geq 2}\Phi_i(\zeta)\ , \tag{63}$$

where $\Phi_i(\zeta)$ are homogeneous polynomials of order i in the new coordinates ς.

The normalized map $\mathrm{U}(\varsigma)$ contains the same linear part Λ of M, and can be expressed as a polynomial expansion:

$$\mathrm{U}(\varsigma)=\Lambda\varsigma+\sum_{i\geq 2}U_i(\zeta)\ , \tag{64}$$

where $U_i(\zeta)$ are polynomials of order i that obey special symmetry rules [23].

Formally, one can always build a normal form U of a non-linear map M. However, Φ and U are in general divergent in any open neighbourhood of the origin, because of the non-existence of an analytic invariant of motion for the original map. In spite of this, normal forms truncated at an order m can be used to interpolate the orbits of M in judiciously chosen domains. The truncation, in general, implies that the symplectic conditions are obeyed up to an order m in the new coordinates. On the other hand, the truncated normalized map has an optimized order of accuracy. Above it, the approximation is improved at lower amplitude and worsened at higher amplitude. The domain of convergence is limited by resonances of low order that are allowed by the truncated Taylor map. There are ways to handle the first limiting resonance, with resonant normal forms [24], which have not yet been made of practical use.

In the LHC, the mapping approach based on Taylor expansion and normal form is used to evaluate the dependence of tune-shift on the amplitude and the momentum due to systematic field-shape imperfections [25]. In this way it is possible to identify the multipoles that are more dangerous for the stability of the motion, taking into account the quite strong higher-order cross terms, and to define and optimize the most suitable scheme of multipolar correctors.

Higher-order Taylor maps are also used to estimate the dynamic aperture in a faster way than with the usual element-by-element tracking [26]: however, in the LHC, this approach is non-controversial only for simulations up to few 10^4 turns. By increasing the order of the map, the violation of area-preserving transport can be made arbitrarily small, but the map size grows exponentially and the computing speed decreases accordingly. An interesting result is that one can correlate the accuracy of the truncated Taylor map tracking to the size of the high-order terms in the map. Alternatively, one can restore the symplecticity of the Taylor map by a linear scaling transformation to the particle coordinates at each turn [26].

3.1.4 Early indicators of chaos

Early indicators of chaotic motion have been used to speed up the estimate of the dynamic aperture in several accelerators and colliders, and, in particular, in the LHC. The exponential divergence of two initially very close trajectories is a criterion for chaos, a linear growth indicating regular motion. The exponential coefficient, called Lyapunov exponent [27], can be used to localize stochastic layers in the phase space and eventually to identify the stability border below which its value is zero.

Rigorously speaking, the maximal Lyapunov exponent is defined as:

$$\lambda = \lim_{\mathrm{d}(0)\to 0, t\to\infty} \frac{1}{t} \log \left| \frac{\mathrm{d}(t)}{\mathrm{d}(0)} \right| \quad \text{with} \quad d(t) = \left\| \vec{d}_1(t) - \vec{d}_2(t) \right\|, \tag{65}$$

where $\vec{d}_1(t)$, $\vec{d}_2(t)$ are the phase space coordinates of two particles initially very close to each other. The routine way to evaluate it is to track simultaneously two particles with a slightly different initial amplitude, and to compute periodically and plot their mutual distance d in the phase space. A schematic illustration of the regular and chaotic evolution of d with time is shown in Fig. 6.

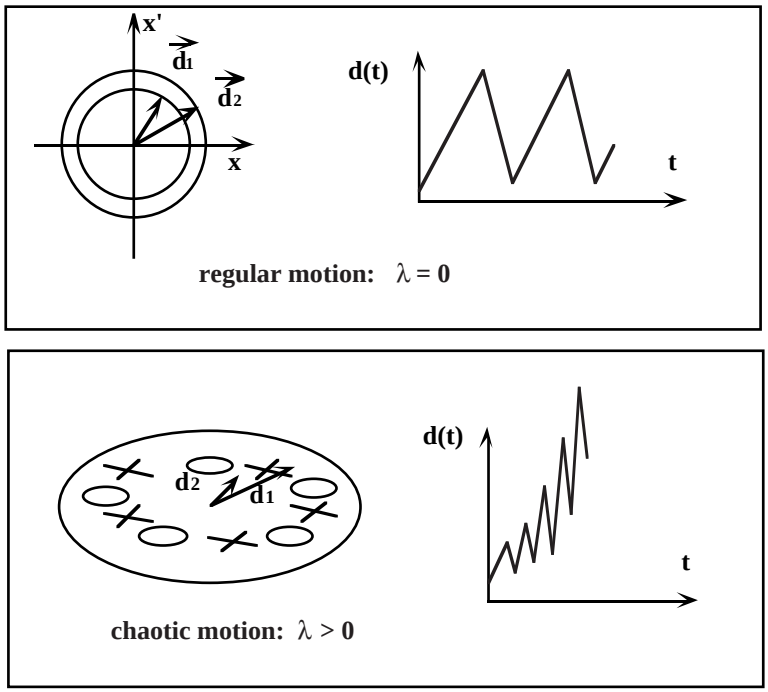


Fig. 6: Graphic method to evaluate the Lyapunov coefficient. The distance d of two initially very close particles is plotted as a function of time.

An equivalent method, better adapted to non-linear maps M, is based on the analytical evaluation of the Jacobian M in the phase-space domain of interest. The eigenvalues of M^k evolve exponentially if the motion is chaotic:

$$\lambda = \lim_{k\to\infty} \frac{1}{k} \log\left(\text{max. eigenvalue of } \mathrm{M}^k\right) . \qquad (66)$$

The predictability of the two methods is enhanced when the non-linear deformation of the phase space is removed by a normal forms type change of coordinates [28].

It is currently admitted that through early indicators of chaos a conservative estimate of the dynamic aperture can be obtained with less computing power than for standard element-by-element tracking.

3.1.5 Figure of merit and data processing

The 'linear aperture', based on smear and tune-shift with the amplitude, was widely used in the past [29] to estimate non-linear effects, since threshold values for detuning and amplitude distortion were considered sufficient to ensure long-term stability. However, the validity of this extrapolation has not been confirmed by more profound studies. Therefore, intensive tracking and sophisticated data-processing are preferred nowadays to estimate the dynamic aperture, after a preliminary selection of rather few significant cases, on the basis of short-term simulations [30]. Results are presented in the graphical form of 'survival plots' that depict the maximum number of stable turns as a function of starting amplitude [31]. Survival plots and early indicators of chaos provide a practical estimate of the stable region.

Dense survival plots are ragged and show a large spread in the survival time close to the chaotic border, rapidly decreasing at larger amplitudes, as shown in Fig. 7. Such an irregular shape reflects the local origin of the particle instability: at moderate amplitude in presence of weaker perturbations, the escape time is largely influenced by microscopic changes of initial coordinates; at large amplitude, instead there are only fast losses.

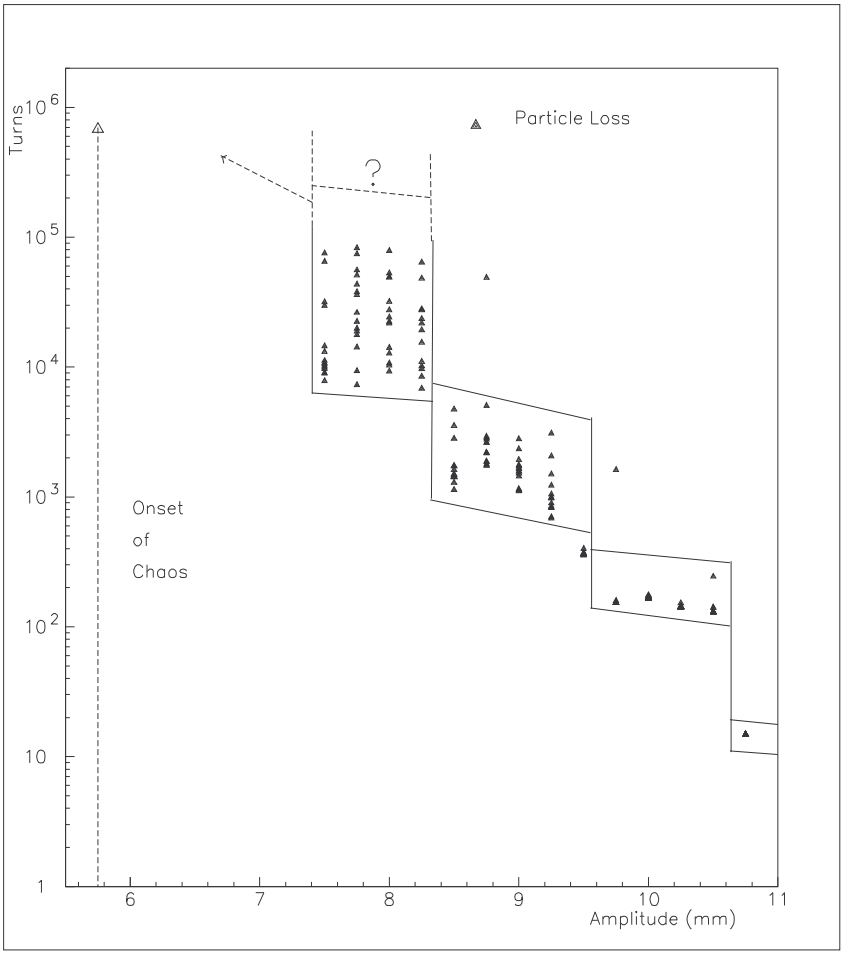


Fig. 7: Dense survival plot.

Under the influence of non-linearities, particles migrate across different nets of resonances. This phenomenon is presented in Fig. 8, where the tunes are plotted in the working diagram as a function of time, for two initially nearby particles undergoing chaotic motion.

In the vicinity of the long-term dynamic aperture, the loss mechanism is in general sudden: the particle may stay confined even for millions of turns and then diverge in a few thousand turns.

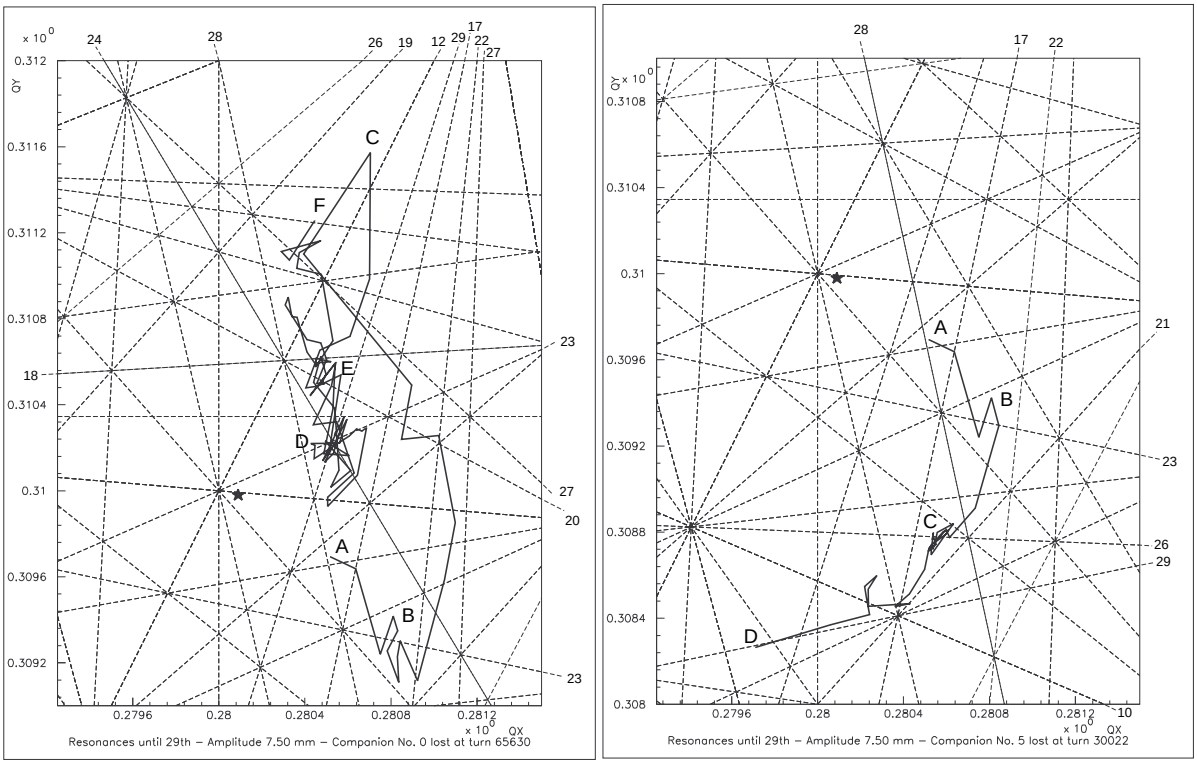


Fig. 8: Evolution of tunes with time of two initially close particles undergoing chaotic motion.

3.2 Applications to accelerator design

Hadron colliders, like the LHC, must operate with negligible loss for long periods, up to 10^8 turns, in spite of the unavoidable field shape imperfections. An upper limit to unintentional multipoles and practical compensation strategies have to be devised for a safe operation. This implies a thorough understanding of the influence of the non-linearities on the long-term behaviour of particle trajectories. Analytical methods available are not yet fully exploited. Numerical simulations are too cumbersome and time-consuming for an exhaustive overview of all the possible situations. Nevertheless, remarkable progress has been made through heuristic approaches proposed in the recent past, based on the investigation of simplified non-linear models of the lattice and on the use of empirical criteria for beam stability.

Too crude simplifications of the lattice structure itself have dramatic effects on non-linear performances. Cell lattice models with only regular cells and no interaction regions show a regular azimuthal pattern of the orbit functions and in particular of the betatronic phase advance leading to unrealistic enhancement of the particle stability. They are in general used for numerical studies of simple dynamical systems such as the Hénon or higher-order maps. A more realistic way to drop the insertions is to replace them with equivalent rotation matrices. Part of the chromatic aberration and some unintentional field errors are disregarded in this way. In addition, this technique makes it particularly difficult, if not impossible, to track in a correct manner off-momentum particles performing synchrotron oscillations. However, relevant information can be gained with less computing power and complexity, especially during the injection plateau, where the stability of motion is mainly determined by non-linear

perturbations in the arcs of regular cells. This approach was used, for instance, to determine the optimum value for the length and the phase advance of the LHC cell [32], fixed to about 100 m and 90°, respectively. More advanced studies are generally based on models with realistic descriptions of the insertions.

The field-shape imperfections are equivalent to multipoles up to large order, which can be expressed as the sum of two parts, one systematic and the other random. The general agreement, based on the magnetic measurements of the Tevatron and Hera [33], is to stop at order 11 in the multipolar expansion and to neglect correlation between random multipoles of different orders. Systematic errors are larger at injection due to persistent currents. Large low-order systematic multipoles, like b_3,b_4,b_5 in superconducting dipoles, or b_6 in superconducting quadrupoles, provoke a sizeable detuning with the amplitude and the momentum, which can be corrected either locally [34], as in the LHC lattice version 2, or using a clever cancellation of the detuning terms by means of Simpson rules [35], as in the LHC lattice version 1. In the latter case, multipolar correctors are to be located near the main quadrupoles as well as at about the middle of each half-cell [25]. Octupolar imperfections can be treated in a global manner. For instance, they have, in the LHC, a particular behaviour related to the symmetry of the magnetic field in two-in-one magnets: the integrated value along the azimuth is expected to be zero, therefore the detuning with the amplitude is expected to be self-compensated without specific correctors. Large higher-order systematic multipoles, like b_7 and b_9, destabilize off-momentum particles and have to be minimized by design: tolerable values for the LHC have been found to be of the order of 2×10^{-6} and 5×10^{-7} units at 1 cm radius, respectively.

Random imperfections, which vary from magnet to magnet owing to manufacturing tolerances, are the main source of non-linear resonances and distortion functions. Statistical distributions can easily be predicted, but are insufficient for a complete knowledge of the non-linear optics, since resonance strengths depend on the specific sequence of the random errors around the ring rather than on statistical properties. Therefore, criteria for magnet design are to be studied on several non-linear lattices, with different sequences of random multipoles. In fact, there are many parameters that limit the stability of the particle motion in hadron colliders, therefore the first task is to identify the most important ones, in order to reduce to a reasonable amount the enormous computing time required for an exhaustive set of simulations.

Heuristic approaches are in general applied, like in the design of the LHC, briefly described hereafter. Parameters routinely considered in the accelerator models are residual closed orbit, linear coupling due to imperfections, synchrotron motion, and residual ripple in main power supplies. Chromaticity and non-linear detuning are corrected with a proper set of correctors. Special cases with some residual uncompensated chromaticity are considered as well. Short-term dynamic aperture is first evaluated by tracking particles of different starting amplitudes for 10^3 turns. This is fast and well-suited for a first exploration of the space parameters, and is also sufficient to reveal the most important features of the non-linear phase space. Ten different seeds are used to fix the test samples of the random errors. Appropriate subsets of them are considered to disentangle the effects of the dipole imperfections from those of the quadrupole imperfections in a machine with a perfect closed orbit and no linear coupling. By choosing three representative seeds in each distribution, i.e. one with the smallest, one with the largest, and one with an average value of the aperture, one can easily check the combined effect of the dipole and the quadrupole errors and identify a limited number of representative sets of non-linear lattice models to be investigated with long-term tracking simulations. With this strategy [30], beam stability has been found to be strongly influenced by linear lattice parameters like tune, residual linear coupling, and peak ß-values in the insertion quadrupoles as well as by a residual chromaticity of a few units. Instead, residual closed orbit associated with magnet misalignment and tune ripple of a few 10^{-4} units showed a weak interference with beam stability. Particles above the stable region are expected to diffuse towards the vacuum pipe at a speed strongly increasing with the transverse amplitude. A set of collimators [36] is used to absorb them before they hit the magnets and provoke an unwanted deposition of energy in the superconducting coils. The transverse position of the primary collimator defines the mechanical aperture of the accelerator. It is basically fixed taking into account the mechanical tolerances of the cold bore and of the thermal beam screen in the main magnets, the expected peak-values of

the closed orbit, of the dispersion, and of the ß-function modulation, and, of course, the optimum value of the separation between the primary and the secondary collimators. For a safe operation, careful matching of physical aperture and stability border is to be performed [30]. With a small mechanical aperture, only trajectories with small amplitude oscillations are allowed, which are weakly perturbed even in presence of large field-shape imperfections, whilst, with larger mechanical apertures, and larger amplitude oscillations, the non-linear perturbation becomes larger and the size of the magnetic errors start to play a leading role for the dynamic aperture. On the other hand, we believe that particles with amplitudes up to the chaotic boundary are stable, although the non-linear perturbations induce a finite smear of their trajectories, whilst particles with larger amplitudes may become unstable after a sufficiently large number of turns. Ideally, the chaotic limit should be equal to the dynamic aperture evaluated in the presence of collimators, in which case only the unstable particles will be intercepted by the collimators. In this respect the mechanical size and the field quality of the LHC at injection have been found to be well matched to a value of 6 σ, i.e. of 7.2 mm, which is considered a wise choice for the needed dynamic and physical aperture.

Strategies of magnet sorting have been invented, by which the magnets are installed in such a sequence in the machines as to minimize the combined non-linear effects. For practical and theoretical reasons, the sorting scheme should be as local as possible and must refer to a limited kind of multipoles. Different solutions have been proposed [37–39]. By introducing a quasi-periodicity of multipoles every two betatronic wavelengths, the harmonic content of non-linearities can be shifted away from harmful frequencies. Alternatively, small groups of magnets, typically ten, are ordered in such a way as to minimize a broad band of non-linear driving terms computed to 2nd perturbative order, contributing to resonances up to order 6. The first method is used in the LHC and the SSC, the last method is used for HERA and is still under investigation for the LHC.

4. REVIEW OF DEDICATED EXPERIMENTS

Since 1986 dedicated experiments have been performed in the CERN SPS and in the FNAL Tevatron to study the beam dynamics in the presence of strong non-linear fields intentionally introduced along the accelerator circumference. The motivation is to study the refinement of aperture and field quality criteria for the design of future large hadron accelerators like the LHC, SSC, or RHIC.

The experimental procedure consists in exciting already existing sextupoles in order to introduce non-linearities in a controlled fashion in an otherwise linear lattice. To probe large amplitudes, either a pencil beam with small emittance and momentum spread is used, to which a large enough coherent deflection is applied, or a large emittance beam is created with repetitive small-amplitude coherent kicks. In the first case, a few hundred turns are sufficient to create a 'hollow' distribution of charges around the central orbit due to non-linear filamentation, in the second case, a few kicks are required to spread out the particles all around the available physical aperture of the accelerator. The proton behaviour is observed with several instruments: current transformers record lifetime, Schottky noise detectors give tune and tune-spread, flying wires provide transverse profile, and orthogonal pairs of position monitors are able to produce a phase-space portrait. Sinusoidal tune modulations are added to simulate the effect on the beam stability of the unavoidable ripples in power supplies.

Interesting experiments have been started more recently at the IUCF cyclotron facility in Indiana University, with the specific purpose of exploring non-linear particle motion, in particular along the resonant islands [40–43]. Only the necessity to be concise forces us to drop their description and to concentrate our attention on the experiments at CERN and FNAL.

4.1 Experimental conditions

Well-understood and clear experimental conditions, with most of the spurious effect eliminated and the phenomena under study carefully isolated, are required to make meaningful comparisons between the experimental results and the numerical simulations of them with computer tracking programs.

In the SPS, low-intensity beams of 10^{11} to 2×10^{12} protons undergo linear motion between 100 GeV/*c* and 250 GeV/*c*, since the remanent fields effects can be neglected, as are the space charge effects, whilst the saturation effects in the magnets are not yet perceptible. An energy of 120 GeV/*c* is finally selected for dynamic aperture studies, so as to benefit from the maximum strength of the added non-linear fields. In the Tevatron, intensities of 10^{10} particles are used, and an operational energy of 150 GeV/*c* is chosen for similar reasons.

Intentional non-linearities are provided by powering independently several sextupoles foreseen for the resonant extraction of the beam at top-energy. Eight sextupoles evenly placed along the ring are routinely used in the SPS, sixteen sextupoles are used in the Tevatron. Positive and negative values of the current are alternated around the rings in such a way as to suppress the third-integer resonances, except in early experiments where that resonance was strongly excited. The chromaticities are corrected using four families of sextupoles in order to obtain an almost flat pattern of the tunes as functions of the momentum deviation.

Closed-orbit deviations and linear coupling are in general well compensated: a residual r.m.s. orbit not in excess of 0.5 mm and a minimum horizontal–vertical tune approach of the order of 0.002 units are currently imposed.

Except for occasional situations, a beam with one single bunch is used in the Tevatron, whilst a debunched beam is used in the SPS.

4.2 Results in the CERN-SPS

4.2.1 Short-term dynamic aperture experiment

Studies of the short-term dynamic aperture were started in 1986 while the SPS was operating in fixed target mode [44,45]. The beam was injected at 14 GeV/*c*, accelerated up to 120 GeV/*c* and allowed to circulate at this energy for eight seconds on a magnetic flat top. The full ring was filled with bunches and the RF was kept on. At the beginning of the 120 GeV/*c* flat top the intentional sextupoles were energized and the beam emittance increased until losses occurred. This was done by repeatedly deflecting the beam with the fast kicker normally used to measure the tunes. In the absence of intentional non-linearities, it was sufficient to pulse the kicker twice at its maximum deflection to reach the machine aperture in the vertical plane, which was of the order of ±15 mm. In these conditions the intensity losses are abrupt in time. When the sextupoles were energized, the acceptance was reduced. The amplitude of the two kicks was therefore lowered so as to produce losses not in excess of 10–20% of the circulating intensity, these losses started just after the second kick and lasted for a few seconds. Following that, in all cases there was another period of a few seconds before the end of the flat top with no visible losses. The size of the surviving beam measured with a high gain of the wire scanner was assumed to be the dynamic aperture of the machine. This mode of operation is schematically illustrated in Fig. 9.

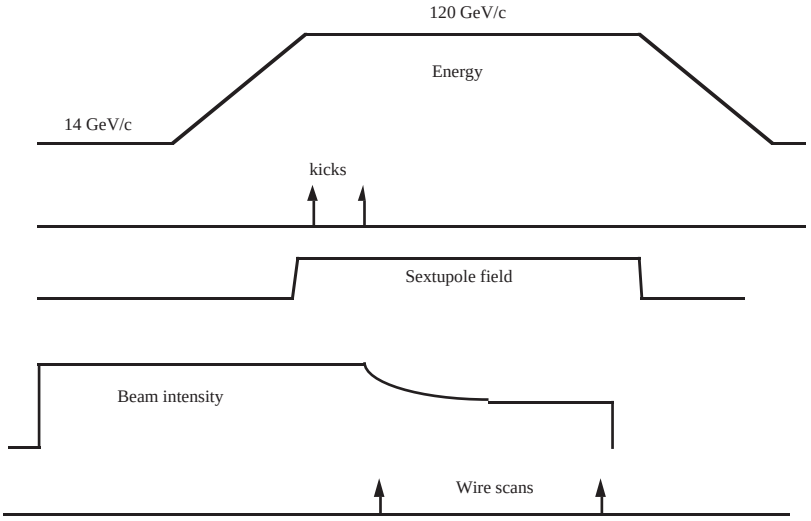


Fig. 9: Mode of operation of the CERN-SPS during the short-term dynamic aperture experiment.

The experiment was performed first with intentional sextupoles exciting the third-integer resonance and later without exciting it. In the latter configuration, several working points around the fifth-order resonances were explored for which a horizontal dynamic aperture of 13–21 mm was measured. Once the synchrotron motion and closed-orbit distortions were taken into account, computer tracking simulations performed over up to 10^5 turns, i.e. a few seconds of real machine time, were able to provide estimates of the dynamic aperture in reasonable agreement with the measured value. The loss mechanism was clearly related to the crossing of the fifth-order resonances, and the upper limit of the detuning and the smear of stable trajectories were found to be of the order of 0.01 units and 0.035 respectively.

In 1988, another experimental session was devoted to the study of the short-term stability, with the same machine conditions as in 1986. At that time, an instrument was made available to record the beam oscillation over several thousands of turns in consecutive position monitors separated by about $\pi/2$ in betatronic phase. Using this instrument, phase portraits could be drawn and tunes could be precisely measured both in pulsed and in storage mode [46].

To explore the effect of the intentional non-linearities, a pencil beam of 1.3 mm r.m.s. size was used, which was obtained by bringing the larger injected beam in proximity with a collimator. Horizontal hollow beams of amplitudes as large as the vacuum chamber could be produced by firing the fast kicker once. Two working points were explored, one near to the fifth- and the other near to the seventh-order resonances. The detuning was measured as a function of the oscillation amplitude, proportional to the kick voltage, and found to be in a very good agreement with the numerical simulations by tracking. The dynamic aperture was measured by adding the amplitude of the coherent oscillation to twice the r.m.s. beam size for which losses were just visible. For comparison, the boundary of the chaotic motion region was computed by numerical simulations with the program SIXTRACK [16]. The agreement between experimental data and simulations was excellent, as for instance for the working point near the fifth-order resonances: the measured aperture was between 15.8 mm and 20.2 mm, the numerical estimate of the chaotic border was 17.6 mm.

With the previous experiments the concept of short-term dynamic aperture was clarified to a large extent. In the presence of strong non-linearities, particles can be lost in a few hundred turns by the effect of higher-order resonance crossing due to the detuning. In these conditions, the Lyapunov coefficient, easily obtainable by numerical simulations, is a good predictive indicator of the chaoticness of the trajectories and of the instability of the motion.

In 1989, the non-linear detuning and its compensation were studied and their effect on dynamic aperture was carefully explored. The measurement of the amplitude-dependent tune shift was again found to be in good agreement with tracking. Figure 10 shows the experimental data of 1989, very close to that of 1988. Moreover, Fig. 10 shows clearly how strongly the particle losses can be influenced by the use of octupoles. By adequately powering them in the SPS we succeeded in reducing the non-linear detuning by roughly an order of magnitude, which at the same time improved the dynamic aperture by 30%. This result suggests that the correction of the non-linear tune-shift in superconducting hadron colliders can be an appropriate method to improve the performance of these very non-linear machines and in particular to reduce short-term losses.

4.2.2 Sextupole-induced diffusion

An early experiment at the Sp$\bar{p}$S, showed that circulating particles have a transverse diffusive behaviour in the presence of localized beam–beam collisions, due to non-linear resonances, tune-spread, and noise. In the context of the studies for the design of large superconducting hadron colliders, an important issue is whether the unavoidable non-linearities in the guide fields are likely to create a diffusion mechanism similar to that induced by beam–beam interactions, since both kinds of non-linearities create tune-spread and resonance excitation, even if in the two cases the dependence with particle amplitude is quite different. This issue was addressed in 1988 at the SPS, with a dedicated experiment on diffusion induced by sextupoles [45].

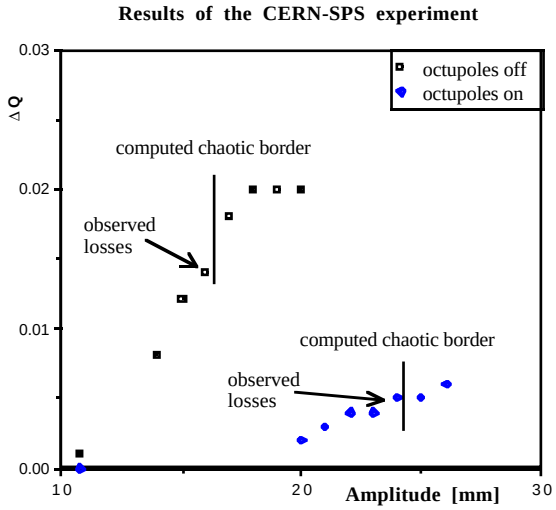


Fig. 10: Compensation of the detuning with the amplitude by means of octupoles.

A debunched beam was stored at 120 GeV/c, and the horizontal beam emittance was slowly increased by repeatedly firing a kicker, until a few per cent losses were noticed. The eight sextupoles were powered so as to avoid excitation of the third-order resonance. The beam intensity was recorded, while certain actions were performed, in order to bring evidence for proton diffusion. A horizontal scraper was introduced close to the edge of the beam in order to clear the tails of the transverse particle distribution. In doing this, a small fraction, typically about 1%, of the circulating particles was lost abruptly. Subsequently, an exponential decay of the intensity was measured. The scraper was then retracted by a few millimetres, typically 2 mm. Consequently, the intensity remained constant for the time required for the particles to fill the gap between the innermost and the retracted position of the scrapers. The time duration with flat intensity could be used to evaluate the average value of the diffusion coefficient. The plateau regime was followed by another period with exponential decay of the intensity, the time constant of which was slightly larger than the initial one. When the scraper was moved back to the innermost position, another abrupt loss of beam was visible, followed by an exponential decay with the same time constant initially observed. We see here a clear signature of a 'diffusion process'.

With the working point near the fifth-order resonances, a diffusion rate of 3 mm/min was measured at an amplitude of 12.6 mm, and 6 mm/min at 15.4 mm. At 18.1 mm the diffusion was so fast that it could not be measured with the procedure described above: this is just across the short-term dynamic aperture sitting between 15.8 mm and 20.2 mm.

Similar measurements were done at working points near to the seventh-order resonances. In this case the largest amplitude at which no significant diffusion could be detected was of 9.2 mm.

The measurement was repeated in the absence of the added sextupoles. There was no sign of diffusion up to an amplitude of 22 mm, well outside the dynamic aperture measured in the previous experiment.

Numerical simulations performed for up to 10^6 turns did not reveal any chaotic motion inside the short-term dynamic aperture. This changed dramatically when a tune modulation of 3×10^{-3} units was introduced in the tracking to take into account the realistic situation of the main power supplies, optimized at that time for the operation in fixed-target mode. This is to be expected from the application of the Chirikov criterion for overlap of satellite resonances. The value of the Lyapunov coefficient computed with tracking now decreased as the particle

amplitude was reduced, and from this, lower and lower diffusion rates could be expected, in qualitative agreement with the experimental results.

With a view to evaluating the predictive power of the detuning and the smear as indicators of stable motion, Fig. 11 was drawn which included all the experimental situations considered in 1986 and 1988. The rectangular box represents the criteria used in 1988 for acceptance of the LHC lattice: it was supposed that in this rectangle the machine would be sufficiently linear to assure a good, comfortable operation. It appeared clearly from the experimental data that the LHC criteria were sufficiently conservative as far as the short-term dynamic aperture was concerned, however, they were certainly insufficient to ensure a good lifetime in storage mode, owing to the magnetic-imperfection-induced diffusion.

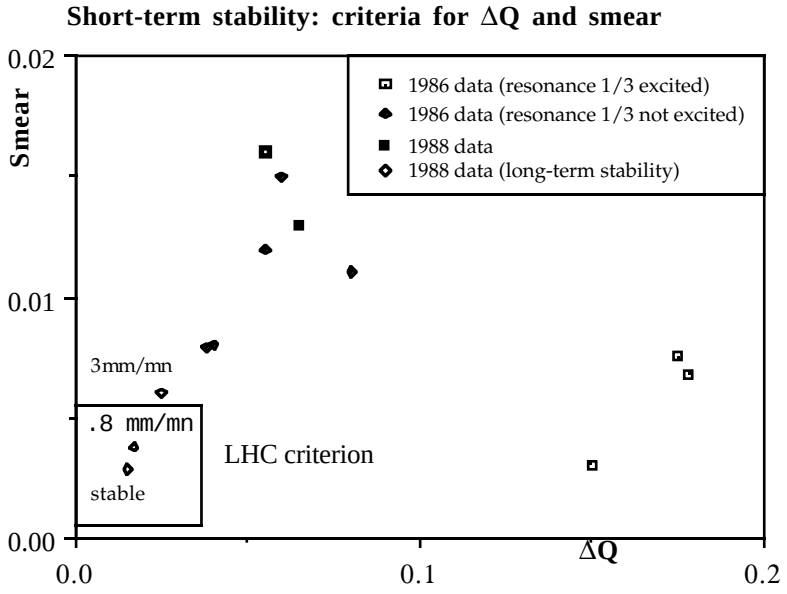


Fig. 11: Predictive power of detuning and smear as indicators of stable motion.

4.2.3 Effects of an external modulation of the tune

Experiments were performed at the SPS in 1989, 1990 and 1991 to get quantitative estimates about the influence of sinusoidal modulations of the tunes on the dynamic aperture [47,48].

As a preliminary step, the time structure and the spectral characteristics of the ripple inherent in the main power supplies were measured in different operational conditions, using the Schottky noise detector for continuous-tune measurements [49]. At a 120 GeV/*c* energy plateau, with the power supply settings appropriate for pulsed operation, the tune variation with time was of the order of $\pm 10^{-3}$ units. The variation was reduced to about $\pm 2 \times 10^{-4}$ units, when the main power supply control was improved as for collider operation. In both cases, the power spectra of the ripple showed a dominant peak at a frequency of 50 Hz, with additional peaks of decreasing power at the higher harmonics of 50 Hz, extending up to 1 kHz.

The effect of the induced tune modulation was studied on a low-emittance debunched beam stored at 120 GeV/*c*. All the circulating particles were deflected horizontally at large amplitude in a 'hollow' beam, by firing a kicker once. The eight sextupoles were powered so as to avoid excitation of the third-order resonance, while provoking non-linear motion. A sinusoidal excitation of the tunes was induced by powering a special quadrupole, located near a focusing main quadrupole. The amplitude of this artificial, controlled ripple can be increased well beyond that of the natural ripple of the main power supplies. Modulation depths up to

about $\pm 2 \times 10^{-3}$ tune units could explored, in a frequency range limited to about 600 Hz by the inductance of the quadrupole. The beam stability was studied with the scraper retraction technique, described in the previous section. The experimental layout is schematically illustrated in Fig. 12.

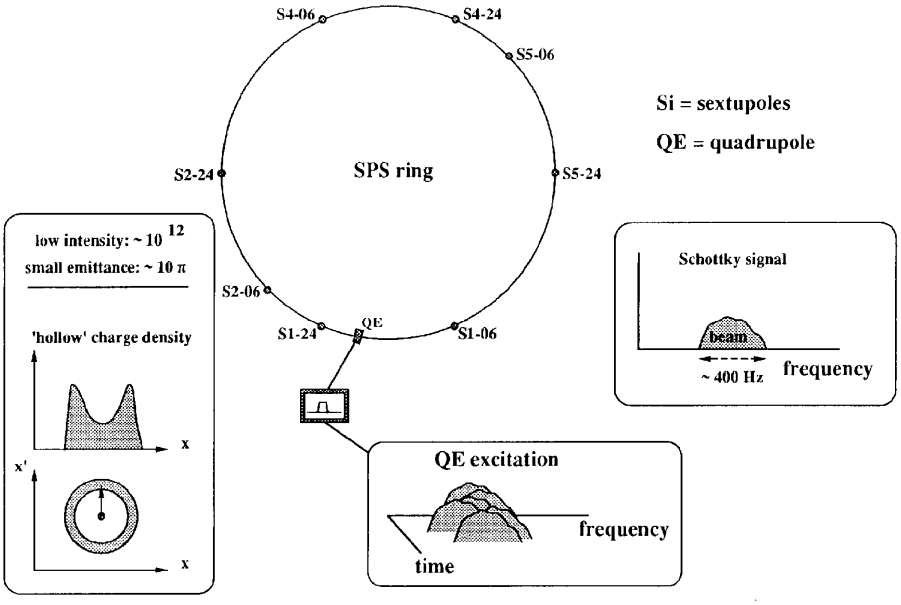


Fig. 12: Mode of operation of the CERN–SPS during the long-term dynamic aperture experiment.

During the experiments, care was taken to compensate for the closed-orbit distortions and the linear coupling. The tune was chosen very carefully so as to minimize the effect of resonances. Octupoles were used to decrease the amplitude-dependent tune shift so as to better locate the working point in between the most dangerous resonances. Even if the particles were initially displaced in the horizontal direction, the net of resonances explored by the beam, in general, provoked horizontal–vertical non-linear coupling, with consequent beam losses in both planes. Therefore, to measure the diffusion speed with the retraction method, both horizontal and vertical scrapings were required.

In storage-mode experiments in 1989 and 1990, the beam size was increased by a single large-amplitude coherent kick. The tune, after applying the kick, was adjusted so as to stay away from the fifth- and seventh-order resonances: Q_H between 0.623 and 0.640, Q_V between 0.529 and 0.540. Measurements with no tune modulations and with modulations of 9 Hz, 40 Hz, and 180 Hz and the same depth of $\pm 8.0 \times 10^{-4}$ tune units were performed. As the scrapers were retracted from the peripheral of the beam, there was the usual period of no noticeable losses. When the first scraper was reached by particles, a drop in lifetime appeared, followed by a second drop when the second scraper was also reached. This is indeed what we can see schematically in Fig. 13. Diffusion speeds of about 0.5 mm/min were recorded without tune modulation and of about 2.5 to 5 mm/min with tune modulation. The diffusion process, however, was almost independent of the modulation frequency.

The importance of the modulation depth was pointed out in a similar experiment performed in 1991. By varying from 1.1×10^{-3} to 2.2×10^{-3} the amplitude of the sinusoidal modulation, an increase of the diffusion speed by a factor of ten was observed for each modulation frequency.

In comparison with the effect of a monochromatic beam excitation, combining two different frequencies in the tune modulation was expected to lead to a sizeable reduction of the dynamic aperture. This effect was tested in the 1991 experiment. Initially there is no ripple and no measurable diffusion. With a tune modulation of 1.65×10^{-3} amplitude and 9 Hz

frequency, the beam lifetime drops to seven hours. The losses stop immediately when the ripple is switched off again. Two combined sinusoidal modulations of 0.8×10^{-3} amplitude each and of 9 Hz and 180 Hz frequency, respectively, produce a severe drop of the lifetime to two hours. In the two situations the same total modulation amplitude was used, however, the effect on beam stability is dramatically larger with two frequencies than with only one frequency. The qualitative explanation of that invokes the Chirikov criterion: with two frequencies, primary resonances have richer spectra of sidebands and more effective overlap.

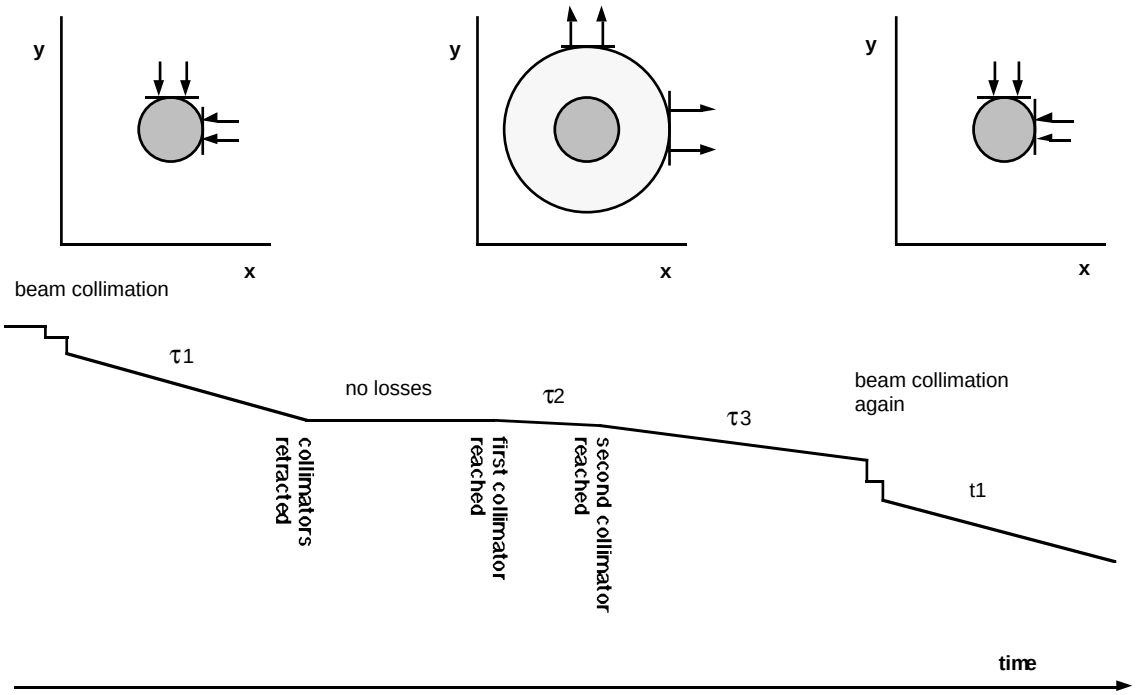


Fig. 13: Illustration of the scraping retraction technique to measure the diffusion coefficient.

From the results discussed above it is clear that in a collider with large non-linear field-shape imperfections, the total ripple amplitude which can be tolerated in the main power supplies is a crucial issue. The present understanding is that tune modulations larger than those observed in the SPS in collider mode, i.e. a few 10^{-4} units, can be quite harmful for beam stability.

4.2.4 Conclusive overview of the SPS experiment

The techniques for the experiment on diffusion and dynamic aperture at the CERN SPS are quite well mastered, although it requires a long and meticulous preparation to obtain clear machine conditions. However, its interpretation is far from being complete.

As far as short-term phenomena are concerned, the predictive power of the Lyapunov coefficient has been well assessed. It has also been proved that the detuning is a parameter strongly influencing the beam stability, therefore its compensation is considered beneficial for the design of future colliders.

Long-term effects are still not fully clarified. The diffusion mechanism itself is likely not to be governed by a simple Fokker–Planck equation. On the other hand, the diffusion process is inherently related to ripple, and empirical criteria are now available for the design of the power supplies of future colliders, which have to be stabilized in current to better than one part per 10^5.

These kinds of studies are still continuing today, with both new experiments and numerical simulations in order to improve the quantitative understanding of the diffusion and the loss mechanisms. The hope is finally to find criteria for the field tolerances of the superconducting magnets for future colliders.

4.3 Results in the FNAL-Tevatron

4.3.1 Tune-shift, smear and short-term dynamic aperture

Studies of the short-term dynamic aperture, non-linear detuning and smear, i.e. the r.m.s. fractional change in amplitude along phase space trajectories, were started in 1985 in the Tevatron [50, 51].

The beam, made of a single bunch of 10^{10} protons, was stored at the injection energy, 150 GeV/*c*. The sixteen sextupoles were ramped up to the desired value of excitation in 10 s. After a further 10 s delay, a coherent betatron oscillation was induced by firing the Tevatron injection kicker. At each of several values of the horizontal tune, a number of kick amplitudes were employed, bounded from above by the onset of beam loss in the Tevatron without sextupoles. The principal data recorded at each condition were the turn-by-turn signals from two monitors, and the beam intensity. This variety of data was analysed to give smear and tune values to be compared with simulations. If the beam were a single particle, then the calculation of the smear from experimental data would reduce to a simple matter of minimizing a least square fit, using the lattice functions and the orbit positions at the monitors as minimization variables. The finite emittance and momentum spread of the beam add difficulties, since the non-linearity introduces a 'decoherence' of the centre-of-gravity motion. The 'decoherence' that arises from chromaticity can easily be corrected. But the dominant source of the 'decoherence' is the non-linear tune variation with the amplitude caused by the sextupoles. The calculation of the phase space distortion in the presence of this effect relies on a reconstruction of a single-particle motion by fitting a Gaussian distribution of the particles to the apparent amplitude reduction of the motion of the beam centroid. Tune dependence with the amplitude, smear and 'decoherence' damping time were measured with particular care at a horizontal tune around 19.4 units. Quantitative comparison between observed and calculated quantities showed an excellent agreement.

In a similar set of experiments, the horizontal emittance of the beam was slowly increased through the introduction of noise into the transverse damper, and the limiting emittance observed as a function of the sextupole excitation. Beam profiles were recorded throughout this process. The limiting beam size was taken to be a measure of the dynamic aperture, when less than the physical aperture, and compared with simulations. The measured aperture was approximately 20% less than the calculated aperture. This discrepancy was ascribed to effects affecting the loss observed over 10^6 turns, that were not included in the few 10^2 turns computer tracking simulations.

4.3.2 Particle beam trapped in resonance islands

In creating 'hollow' beams at the Tevatron, it was observed that, for particular values of the kick amplitude and polarity, some fraction of the beam was trapped in resonance islands. The detection of such trapping turned out to be surprisingly easy, since, in the presence of 'decoherence' of the rest of the beam, the trapped particles continue to oscillate in a coherent fashion. In addition, both the phase-space plot and the spectral analysis of the transverse position signal as a function of time show that the sustained signal is associated with a fractional value of the tune. For instance at a tune near 19.4, the phase space plot typically showed five islands, whilst the Fourier analysis had power components peaked at two-fifths of the revolution frequency. External tune modulation was found to influence the lifetime of the persistent tune lines associated with particles trapped in islands. The interpretation of this phenomenon in terms of Hamiltonian theory is given in Ref. [52]. Comparisons with experimental data are reported in Refs. [51,53]. A concise description of the various arguments is presented here.

In a Poincaré plot (θ,J) representing the horizontal motion in the Tevatron, with a horizontal tune slightly below 19.4, protons launched at small amplitudes follow circular orbits. As the initial amplitude is increased, the tune increases with a parabolic law. Resonance islands appear at an appropriate amplitude, corresponding to rational tunes: five islands are visible as the tune approaches a value of 2/5 units. A proton launched inside one of the islands jumps two islands per accelerator turn, returning to its original island after five turns. Since the proton

never escapes the island in the absence of external tune modulation, a resonant trajectory only visits a localized range of phase, that is the phase is locked and the tune is exactly 2/5 units.

Single-particle motion inside the island is represented by the five-turn Hamiltonian:

$$H_5 = 2\pi\left(Q_0 - \tfrac{2}{5}\right)J + \tfrac{1}{2}U(J)J^2 - V_5 J^{5/2}\cos\ (5\theta), \tag{67}$$

where $U(J)$ parametrize the tune-shift with the action J, and V_5 represents the strength of the resonance. The equations of motion are:

$$\begin{cases} \dot{\theta} = 2\pi\left(Q_0 - \tfrac{2}{5}\right) + UJ - \tfrac{2}{5}V_5 J^{3/2}\cos\,(5\theta) \\ \dot{J} = 5V_5 J^{5/2}\sin(5\theta)\,. \end{cases} \tag{68}$$

The motion of a proton inside one island is well approximated when Eqs. (68) are integrated over five turns. Protons infinitely close to the fixed point at the centre of the island are characterized by the resonant action J_R, and rotate at the island tune:

$$Q_I = \frac{5}{2\pi}\left(UV_5 J_R^{5/2}\right)^{1/2}. \tag{69}$$

Protons away from the centre rotate at a different speed owing to the non-linear detuning with the distance from the fixed point.

Phenomena that modify the distribution of particles inside the island have relaxation periods, until the new equilibrium distribution is reached, lasting of the order of $1/Q_I$ turns.

Tune modulations of the type:

$$Q(t) = Q_0 + q\cos\left(2\pi Q_{\mathrm{m}} t\right) \tag{70}$$

produce four distinct types of dynamical behaviour, in the control parameter space of tune modulation strength and tune (q,Q_{m}).

For low values of q and Q_{m}, the tune modulation is slow and weak, so that $Q(t)$ changes adiabatically. On account of this, the island amplitudes change slowly, but their phases are almost constant. This is the *amplitude modulation* regime.

As the tune modulation strength q increases at constant modulation time Q_{m}, the stable area of the islands shrinks, up to the point where there is no stable area at all and the adiabatic trapping is eventually lost. Close to the fifth-order resonance, the condition for adiabatic trapping is:

$$qQ_{\mathrm{m}} < \frac{1}{5}Q_I^2\,. \tag{71}$$

When the relation (71) is violated the motion becomes *chaotic*.

A tune modulation, rapid but weak, causes the island fixed points to oscillate in phase but not in amplitude. Additional chains of modulational sideband islands appear around the fundamental tune of 2/5 units, spaced by multiples of the modulational frequency:

$$Q_{\mathrm{sidebands}} = \frac{2}{5} \pm \frac{\mathrm{i}}{5}Q_{\mathrm{m}} \quad \text{with } i = 1,2,\ldots\ . \tag{72}$$

Phase averaging techniques, that are rigorous in the fast limit, show that the i^{th} sideband is negligibly small if the modulation strength q is smaller than the tune distance from the fundamental:

$$q < \frac{i}{5}Q_{\mathrm{m}}. \tag{73}$$

This condition, with $i = 1$, is the boundary between *phase modulations*, when it holds, and *strong sidebands*, when it does not hold.

When sidebands are present, the Chirikov overlap criterion holds and *chaos* ensues if:

$$Q_{\mathrm{m}}^{3/4} q^{1/4} < \frac{4}{(5\pi)^{1/4}} Q_I . \tag{74}$$

At the Tevatron, three of the five regions of the control parameter space (q,Q_{m})were made experimentally accessible.

In one experiment q and Q_{m} were linearly ramped for 1 s up to quite large values and then turned off. In another experiment, only Q_{m} was linearly ramped for 1 s, and q was taken constant and small. Tune modulation strength was varied in the range from few 10^{-4} to 10^{-3} units, with modulation frequencies up to 350 Hz. In this way, the boundary between amplitude modulation, phase modulation, and chaotic regions in (q,Q_{m}) space was explored, and found to be in a good agreement with analytical estimates.

Tune modulation strengths not larger than few 10^{-4} units in a frequency range below few hundred Hz were suggested to be considered tolerable in the design of future colliders.

4.3.3 Model for diffusion induced by sextupoles

In this experiment the Tevatron was used to study stochastic effects due to non-linear beam dynamics that cause 'diffusive' evolution of the beam distribution, even in the absence of external sources of noise [54]. Initially, 'hollow' beam distributions were created with 10^{10} particles circulating in one bunch at 150 GeV/c. Then sixteen sextupoles were switched on to observe the effect of non-linearities on the particle density. The transverse beam profile was determined by measuring the instantaneous flux of secondary particles produced by a fine wire flying through the beam repeatedly every minute or so. During a typical run, lasting about 30 minutes, protons performed 10^8 turns in the Tevatron. Under the influence of sextupoles, the beam increased in size, and the macroscopic density was reduced, since, at microscopic level, void regions were mixed with populated regions. Measurements were performed in the vicinity of the 2/5 resonance, with a tune close to 19.4 units. The evolution of the particle distribution was supposed to obey the diffusion equation:

$$\frac{\partial \rho(J,t)}{\partial t} = \frac{\partial}{\partial J}\left(D(J) \frac{\partial \rho(J,t)}{\partial J} \right). \tag{75}$$

Here $\rho(J,t)$ is the distribution function, normalized to the number of particles $N(t)$ that survive at time t, measured by a current transformer, and $D(J)$ is an amplitude-dependent diffusion coefficient to be determined by fitting with the experimental data given by the wire-scanner profile monitor.

The model of diffusive evolution is based on a four-parameter expression for the diffusion coefficient:

$$D(J) = \begin{cases} D_0 & \text{for } J < J_0 \\ D_0 + d_0 \left(J - J_0 \right)^i & \text{for } J \geq J_0 \end{cases} \tag{76}$$

The most important parameter is J_0 which can be interpreted as the phase space radius inside which the non-linear dynamical effects are negligibly small. The parameter D_0 represents a residual uniform diffusion acting on all particles, that will eventually be caused by external noise. In the experimental fitting, D_0 is arbitrarily allowed to vary from configuration to configuration. The increase of diffusion rate above J_0 is parametrized by d_0 and i.

Models of profile evolution are computed by solving Eq. (75) with the diffusion coefficient of Eq. (76), with boundary conditions that forbid particle depletion at the origin and force the absence of particles above the mechanical aperture.

The hollow distribution extends over a small annular surface of the phase space. The diffusion coefficient in Eq. (76) is therefore well determined by experimental fits only in the filled area of the phase space. The extrapolation into regions that are not populated experimentally is clearly invalid.

During the experiment the attempt was made to bracket J_0 by preferentially populating a low-diffusion region in one run and a high-diffusion region in the other. The data collection grid was quite limited for operational reasons.

The experimental fit of J_0 was compared with the stability limit for particle motion estimated with element-by-element computer tracking simulations over a quite small number of turns. The ratio between computed short-term dynamic aperture and the experimental boundary for diffusion is consistently equal to about 2. As a rule of thumb, based on this ratio, the realistic value of the dynamic aperture is about a factor of 2 smaller than the optimistic value based on fast numerical estimates.

4.4 Comparison of the SPS and the Tevatron experimental results

There are three common issues, relevant for non-linear dynamics in particle accelerators, that have been addressed by the experiments at CERN and at FNAL.

The first issue deals with relatively simple dynamical quantities like tune-shift with the amplitude and smear, and refers to a time scale that influences the short-term stability. In the two machines, the experimental results concerning these issues are very clean, and the predictability of analytical and numerical estimates is in general quite good.

The second issue concerns the diffusion induced by sextupoles, the time scale of which is quite large. The common result of the two experiments is that the long-term stability depends heavily upon the choice of the working point and the vicinity of low-order resonances. The reproducibility of experimental data is difficult to achieve and requires a large effort to restore identical working conditions. The numerical models give a satisfactory level of predictability, in general better than 30%, provided that most of the imperfections, like finite closed-orbit, linear coupling, power supply ripple, are correctly accounted for. Predictions in general require large computational power. The effort of understanding predictability was mostly done at the SPS, whilst at the Tevatron the construction of an analytical model of diffusion was of the highest concern.

The third issue points out the effect of tune modulations on long-term stability. It was addressed by substantially different methods in the two machines. At CERN, the local diffusion coefficient was measured with a quite direct method, using the technique of collimation retraction. Working points in the vicinity of 2-D resonances that couple the two transverse degrees of freedom were explored. The investigation of the influence of modulation strength and tune was empirically inspired by the overlap criterion. Predictions were based on brute-force simulation techniques. At FNAL, the phenomenon was studied with particles trapped in 1-D horizontal resonance islands. The predictions were based on a simple analytical model, based on Hamilton's equations. Surprisingly enough, the two approaches concluded with empirical recipes that fix to the same value, i.e. a few parts in 10^{-5}, the maximum tolerable tune ripple in hadron colliders.

5. CONCLUSION

The concept of dynamic aperture is fundamental for single-particle, non-linear beam-dynamics in hadron accelerators. Mature analytical approaches are available to examine it, based on highly sophisticated mathematical methods that describe Hamiltonian systems and the resonant phenomena relative to them. The most enlightening result is the KAM theorem that, roughly speaking, demonstrates the possible coexistence of stable trajectories and chaotic layers in the phase space of non-integrable Hamiltonian systems. The analysis of the Lyapunov coefficient and the Chirikov criterion of overlapping resonances are crucial in the identification of chaos. And finally, the perturbative theory of maps makes available algorithmic rules to compute higher-order contributions to dynamical quantities.

However, our theoretical understanding is, at present, insufficient for practical applications. Therefore, heuristic approaches have been invented, based on extensive simulations with computer programs, to evaluate the stability of the motion in accelerators. They are very useful for accelerator design and performance evaluations, but they are often specific to the particular accelerator under investigation.

Experiments intended to reproduce in a controlled fashion the non-linear behaviour of particle motion have been intensively pursued for several years in various laboratories on both sides of the Atlantic. We have been able to observe effects of non-linear resonances, onset of chaos and diffusive motion, that are, to a large extent, in agreement with analytical or numerical expectations, once the unavoidable imperfections which are present in the real machine are correctly considered.

ACKNOWLEDGEMENTS

I wish to thank E. J. Wilson, the head of the CERN Accelerator School, for providing me with the motivation to write this lecture, M. Giovannozzi for his most enlightening discussions on this topic, E. Todesco, F. Schmidt, J. Gareyte, S. Turner and B. Warnock for carefully reading the original manuscript, and the staff of the Desktop Publishing Service.

REFERENCES

[1] G. D. Birkhoff, *Dynamical Systems* (American Mathematical Society, New York, 1927).

[2] H. Poincaré, *Les Méthodes nouvelles de la Mécanique célèste* (Gauthier-Villars, Paris, 1892).

[3] A. Chao, *Recent Efforts on Non-Linear Dynamics*, SSCL-Prep. 132 Rev. 1 (1992).

[4] A. Schoch, *Theory of Linear and Non-Linear Perturbations of Betatron Oscillations in Alternating Gradient Synchrotrons*, CERN 57–21 (1958).

[5] R. D. Ruth, *Single Particle Dynamics in Circular Accelerators*, SLAC-PUB-4103 (1985).

[6] H. Mais, *Non-Linear Problems in Accelerator Physics* (Inst. Phys. Conf. Ser. No. 131, 1992) p. 1.

[7] E.D. Courant and H. S. Snyder, *Theory of the Alternating-Gradient Synchrotron*, Ann. Phys. (N.Y.) **3** (1958) 1–48.

[8] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).

[9] G. Guignard, *A General Treatment of Resonances in Accelerators*, CERN 78–12 (1978).

[10] B.V. Chirikov, *Universal Instability of Many-Dimensional Oscillator Systems*, Phys. Rep. **52** (1979) 263–379.

[11] M. Hénon, *Numerical Study of Quadratic Area-Preserving Mappings*, Q. Appl. Math. **27** (1969) 291.

[12] M. Giovannozzi, *Analysis of the Stability Domain for the Hénon Map*, Phys. Lett. **A182** (1993) 255.

[13] K.G. Steffen, *High Energy Beam Optics* (Interscience, New York, 1965).

[14] E. Lohrmann and P. Wilhelm, *Rounding Errors in Beam Tracking Calculations*, Part. Accel. **19** (1986) 99.

[15] H. Grote and C. Iselin, *The MAD Program*, CERN/SL/90–13 (AP) Rev. 3 (1993).

[16] F. Schmidt, *SIXTRACK version 1.1*, CERN/SL/AP 91–52 (1991).

[17] Y. Yan, *Applications of Differential Algebra to Single-Particle Dynamics in Storage Rings*, SSCL-500 (1991).

[18] E. Forest et al., *The Modern Approach to Single-Particle Dynamics for Circular Rings* (AIP Conf. Proc. No. 292, Upton, 1992) p. 417.

[19] M. Berz, *Differential Algebraic Description of the Beam Dynamics to Very High Order*, Part. Accel. **24** (1989) 109.

[20] A. Bazzani et al., *Normal Forms for Hamiltonian Maps and Non-Linear Effects in a Particle Accelerator*, Nuovo Cim. **102 B**, No. 1 (1988) 51–80.

[21] A.J. Dragt et al., *Lie Algebraic Treatment of Linear and Non-Linear Beam Dynamics*, Ann. Rev. Nucl. Part. Sci. **38** (1988) 455.

[22] E. Forest, J. Irwin and M. Berz, *Normal Forms Methods for Complicated Periodic Systems*, Part. Accel. **24** (1989) 91.

[23] A. Bazzani, E. Todesco, G. Turchetti and G. Servizi, *A Normal Form Approach to the Theory of Non-Linear Betatronic Motion*, CERN 94–02 (1994).

[24] A. Bazzani, M. Giovannozzi, G. Servizi, E. Todesco and G. Turchetti, *Resonant Normal Forms, Interpolating Hamiltonians and Stability Analysis of Area Preserving Maps*, Physica **D64** (1993) 66–97.

[25] W. Scandale, F. Schmidt and E. Todesco, *Compensation of the Tune-Shift in the LHC, Using the Normal Form Techniques*, Part. Accel. **35** (1991) 53.

[26] R. Kleiss, F. Schmidt and F. Zimmermann, *The Use of Truncated Taylor Maps in Dynamic Aperture Studies*, Part. Accel. **41** (1993) 117.

[27] A.M. Lyapunov, *Stability of Motion* (Academic Press, New York, 1966).

[28] F. Schmidt, F. Willeke and F. Zimmermann, *Comparison of Methods to Determine Long-Term Stability in Proton Storage Rings*, Part. Accel. **35** (1991) 249.

[29] A. Chao et al., *The Criteria Working Group Summary*, CERN 88–04 (1988) p. 152.

[30] Z. Guo, T. Risselada and W. Scandale, *Dynamic Aperture of the CERN–LHC with Injection Optics*, Proc. 5th Advanced ICFA Beam Dynamics Workshop on Effects of Errors in Accelerators, their Diagnosis and Corrections, Corpus Christi, 1991, Ed. A.W. Chao (AIP Conf Proc., No. 255, New York, 1992) p. 50.

[31] F. Galluccio and F. Schmidt, *Towards a Better Understanding of Slow Particle Loss in the Large Hadron Collider*, Proc. 5th Advanced ICFA Beam Dynamics Workshop on Effects of Errors in Accelerators, their Diagnosis and Corrections, Corpus Christi, 1991, Ed. A.W. Chao (AIP Conf Proc., No. 255, New York, 1992) p. 86.

[32] M. Bassetti, J.P. Koutchouk and W. Scandale, A. Verdier, *Optimisation of the LHC Lattice and Chromaticity*, IEEE PAC 87, Washington, 1987 (IEEE, New York, 1987) p. 1284.

[33] P. Schmüser, *Field Quality Issues in Superconducting Magnets*, IEEE PAC 91, San Francisco, 1991 (IEEE New York, 1991) p. 37.

[34] M. Giovannozzi and F. Schmidt, *General Normal Forms Procedure to Correct Tune-Shift and Non-Linear Chromaticity for Large Accelerators like the LHC*, these proceedings.

[35] D. Neuffer, *Multipole Correction in Large Synchrotrons*, CERN 88–04 (1988) p. 179.

[36] *Design Study of the Large Hadron Collider (LHC)*, CERN 91–03 (1991) p. 115.

[37] R. Gluckstern and S. Ohnuma, *Reduction of Sextupole Distortion by Shuffling Magnets in Small Groups*, IEEE PAC 85, Vancouver, 1985 (IEEE, New York, 1985) p. 2315.

[38] F. Willeke, *A Magnet Ordering Procedure for the HERA Electron-Proton Collider*, DESY-HERA 87–12 (1987).

[39] W. Scandale and L. Wang, *Installing Two-in-One LHC Magnets in Ordered Sequence*, CERN SPS 89–41 (AMS) (1989).

[40] S.Y. Lee et al., *Experimental Determination of a Non-Linear Hamiltonian in a Synchrotron*, Phys. Rev. Lett. **67** (1991) 3768.

[41] D.D. Caussyn, et al., *Experimental Studies of Non-Linear Beam Dynamics*, Phys. Rev. **A46** (1992) 7942.

[42] M. Ellison, *Driven Response of the Synchrotron Motion of a Beam*, Phys. Rev. Lett. **70** (1993) 591.

[43] J.Y. Liu et al., *Determination of the Linear Coupling Resonance Strength Using 2-D Invariant Tori*, Phys. Rev. **E49** (1994) 2347.

[44] L. Evans, J. Gareyte, A. Hilaire, W. Scandale and L. Vos, *The Non-Linear Dynamic Aperture Experiment in the SPS*, EPAC 88, Rome, 1988 (World Scientific, Singapore, 1989) p. 619.

[45] J. Gareyte, H. Hilaire and F. Schmidt, *Dynamic Aperture and Long Term Stability in the Presence of Strong Sextupoles in the CERN SPS*, PAC 89, Chicago, 1989 (IEEE, New York, 1989) p. 1376.

[46] A. Burns et al., *The BOSC Project*, EPAC 90, Nice, 1990 (Editions Frontières, Gif-sur-Yvette, 1990) p. 803.

[47] D. Brandt, E. Brouzet, A. Faugier, F. Galluccio, J, Gareyte, W. Herr, A. Hilaire, W. Scandale, F. Schmidt and A. Verdier, *Influence of the Power Supply Ripple on the Dynamic Aperture of the SPS in the Presence of Strong Non-Linear Fields*, EPAC 90, Nice, 1990 (Editions Frontières, Gif-sur-Yvette, 1990) p. 1438.

[48] X. Altuna et al., *The 1991 Dynamic Aperture Experiment at the CERN SPS*, Proc. 5th Advanced ICFA Beam Dynamics Workshop on Effects of Errors in Accelerators, their Diagnosis and Corrections, Corpus Christi, 1991, Ed. A.W. Chao (AIP Conf. Proc., No. 255, New York, 1992) p. 355.

[49] T. Linnecar and W. Scandale, *Continuous Tune Measurement Using the Schottky Detector*, PAC 83, Santa Fe, 1983 (IEEE, New York, 1983) p. 2185.

[50] D.A. Edwards and M.J. Syphers, *An Overview of the Experiment E778*, CERN 88–04, (1988) p. 95.

[51] A. Chao et al., *Experimental Investigation of Non-Linear Dynamics in the Fermilab Tevatron*, Phys. Rev. Lett. **61** (1988) 2752.

[52] S.G. Peggs, *Hamiltonian Theory of the E778 Non-Linear Dynamics Experiment*, CERN 88–04 (1988) p. 131.

[53] T. Satogata et al., *Driven Response of a Trapped Particle Beam*, Phys. Rev. Lett. **68** (1992) 1838.

[54] T. Chen et al., *Measurements of a Hamiltonian System and their Description by a Diffusive Model*, Phys. Rev. Lett. **68** (1992) 33.

LOW EMITTANCE LATTICES

A. Ropert
ESRF, Grenoble, France

Abstract
The new generation of synchrotron light sources is characterized by an extremely high brilliance synchrotron radiation achieved in low emittance storage rings specially designed to accommodate insertion devices. The basic principles guiding the minimization of the emittance are recalled. The magnet lattice configurations that are currently studied as candidates for these storage rings (DFA, TBA, FODO...) are described and the problems associated with these low emittance lattices and the required strong focusing discussed.

1. INTRODUCTION

High brilliance synchrotron radiation can be obtained through the utilization of insertion devices in low emittance storage rings. A small beam emittance is also very important for FEL experimentation. In that case additional requirements must also be fulfilled. In this lecture, emphasis will be placed on storage ring lattices specially designed to maximize the brilliance and achieve the full potential from insertion devices.

The photon flux is an important source characteristic. Nevertheless, for a number of experiments in which the beam is focused at the sample, the true figure of merit is the spectral brilliance B (called in short brilliance) which is defined as the photon flux per unit solid angle and per unit source area, emitted in a relative bandwidth:

$$B = \frac{d^4N}{dt\ d\Omega\ dS\ (d\lambda/\lambda)} \qquad (1)$$

with
N = number of photons
t = time
λ = wavelength
Ω = solid angle
S = source size

Brilliance is generally expressed in units of:

$$\frac{\text{photons per second}}{(\text{mrad})^2\ (\text{mm})^2\ 0.1\%\ \text{bandwidth}}$$

Brilliance of X-rays has increased by many orders of magnitude since the advent of synchrotron radiation sources as shown in Fig. 1.

Apart from diffraction effects, we have $dS\ d\Omega \sim \varepsilon_x\ \varepsilon_y$, where ε_x and ε_y are the transverse particle beam emittances. Therefore the smaller the emittance of the electron beam, the higher the brilliance of the produced radiation. Most existing rings have emittances of 100 nanometer-radians. The new generation of rings under construction has design values between 5 and 10 nanometer-radians. Yet the maximum level of brilliance is reached only at the so-called

diffraction limit when the beam emittance is about equal to a fraction of the wavelength of the radiation. As an example, a diffracted limited beam capable of producing 10 keV X-rays would have an emittance of the order of 0.01 nanometer-radians. This sets the scale for the ideally desirable beam emittance.

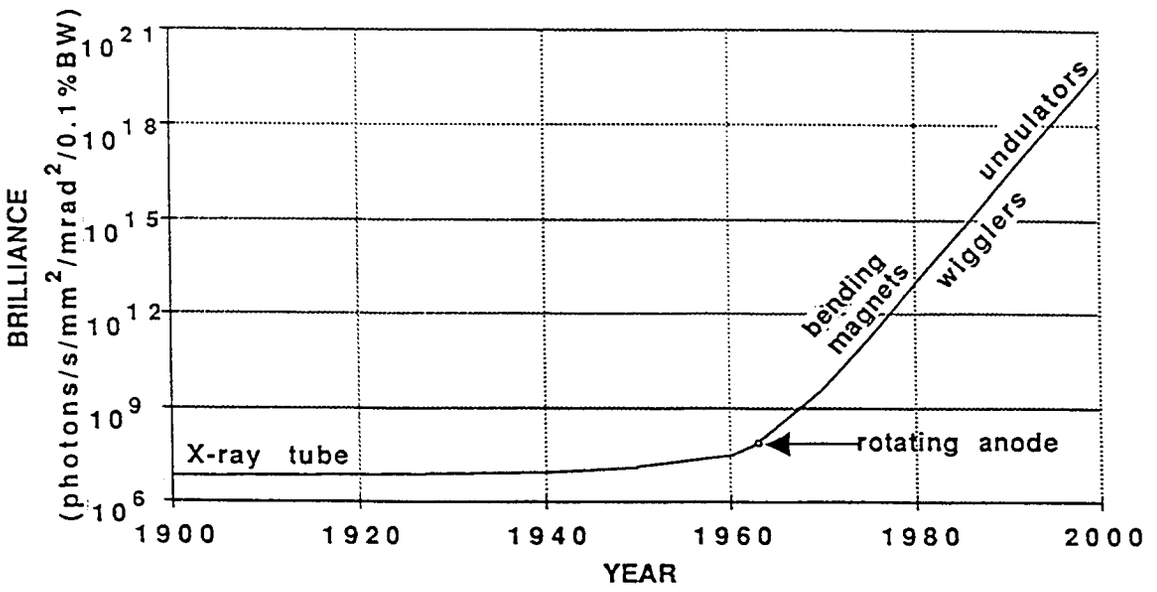


Fig. 1 Increase of brilliance over the last decades

The brilliance can be increased even more if insertion devices are used. Many undulations can be produced along the trajectory, thereby increasing the flux compared to an ordinary bending magnet. For wigglers, the brilliance is enhanced by approximately the number of poles of the device. In an undulator, waves radiated by each electron can mutually reinforce or suppress one another to enhance the radiation of certain wavelengths, resulting in extremely high brilliance.

2. LOW EMITTANCE LATTICES

2.1 Equilibrium emittance

The particle beam emittance in an electron storage ring is determined by an equilibrium between quantum excitation that causes individual particles to oscillate transversally and damping of the betatron oscillations.

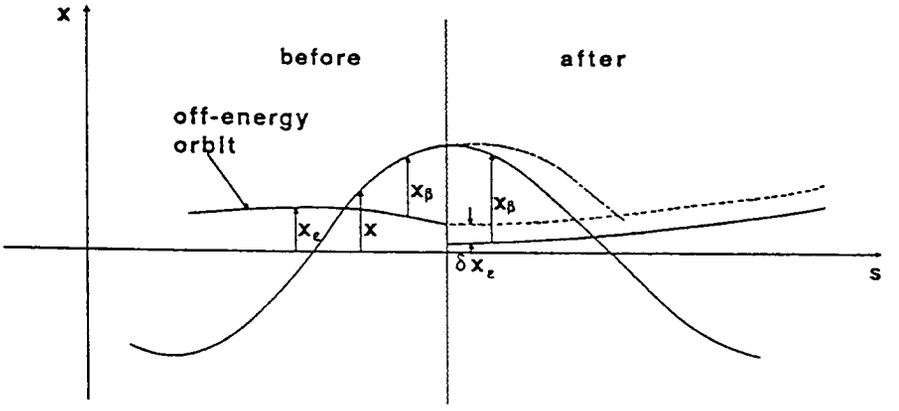


Fig. 2 Effect of energy loss on betatron oscillations

In an undistorted machine, radiation is produced mainly in the normal bending magnets. As a particle emits a photon, it loses energy and this results in a sudden change of its equilibrium orbit which causes a corresponding increase of the betatron oscillation amplitude about the new equilibrium orbit (Fig. 2). This process of continuous increase of the beam emittance is compensated by the damping provided by the RF accelerating system which restores the energy lost by synchrotron radiation only in the longitudinal direction.

The resulting horizontal equilibrium emittance is given by the expression [1]:

$$\varepsilon_{x_0} = \frac{C_q \gamma^2 < H / \rho^3 >}{J_x < 1 / \rho^2 >} \tag{2}$$

where < > means the average around the storage ring.

ρ = bending radius
γ = total energy in mc^2 units
J_x is the horizontal damping partition number
$C_q = \frac{55}{32\sqrt{3}} \frac{h}{2\pi m c} = 3.84\ 10^{-13}$ m

H is the Courant-Snyder dispersion invariant which is specified by the properties of the guide field. It is given by:

$$H = \gamma_x \eta^2 + 2\ \alpha_x\ \eta\ \eta' + \beta_x\ \eta'^2 \tag{3}$$

where α_x, β_x, γ_x are the Twiss coefficients and η, η' are the dispersion function and first derivative respectively.

For an isomagnetic guide field ($\rho = \rho_0$ = Constant in magnets, $\rho = \infty$ elsewhere), Eq. (2) becomes:

$$\varepsilon_{x_0} = \frac{C_q\ \gamma^2 < H >_{mag}}{J_x\ \rho_0} \tag{4}$$

where $<H>_{mag}$ is the average of H taken only in the magnets. That is:

$$< H >_{\text{mag}} = \frac{1}{2\ \pi \rho_0} \int [\ \gamma_x \eta^2 + 2\ \alpha_x \eta\ \eta' + \beta_x \eta'^2\ ]\ ds \tag{5}$$

In the vertical plane, there is radiation damping but ideally no dispersion and hence no quantum excitation, so that in a perfect machine the vertical emittance is zero. The vertical emittance is determined only by coupling between the horizontal and vertical planes due to magnet imperfections and misalignments. The coupling is generally expressed by a constant κ which describes the sharing of the natural emittance ε_{x_0} between the two planes:

$$\varepsilon_x = \frac{1}{1+\kappa}\varepsilon_{x0} \qquad (6)$$

$$\varepsilon_y = \frac{\kappa}{1+\kappa}\varepsilon_{x0}$$

where ε_x and ε_y are the effective emittances. Typically the coupling is of the order of a few %. It can be controlled by means of skew quadrupoles.

2.2 Minimization of emittance

From Eq. (4), it is clear that minimization of emittance, assuming in a first approach J_x to be constant, is achieved by the minimization of the dispersion invariant function. This means that in the bending magnets, where the photons are emitted, the dispersion function η must be low and the β function optimized.

For a single magnet with length L and bending radius ρ, the integral for deriving the average value of H is given by:

$$I = \int_0^L (\gamma\eta^2 + 2\,\alpha\,\eta\,\eta' + \beta\,\eta'^2)\,ds \qquad (7)$$

$$= (\gamma_0\,\eta_0^2 + 2\,\alpha_0\,\eta_0\,\eta'_0 + \beta_0\,\eta'^2_0) + 2\,(\alpha_0\,\eta_0 + \beta_0\,\eta'_0)\,\rho\,(L - \cos\frac{L}{\rho})$$

$$= (\gamma_0\,\eta_0^2 + 2\,\alpha_0\,\eta_0\,\eta'_0 + \beta_0\,\eta'^2_0) + 2\,(\alpha_0\,\eta_0 + \beta_0\,\eta'_0)\,\rho\,(L - \cos\frac{L}{\rho})$$

$$-2\,(\gamma_0\eta_0 + \alpha_0\,\eta'_0)\,\rho\,(L - \rho\sin\frac{L}{\rho}) + \frac{1}{2}\beta_0\left[L - \frac{\rho}{2}\sin(2\frac{L}{\rho})\right]$$

$$-2\,\alpha_0\,\rho^2\left[\frac{3}{4} - \cos\frac{L}{\rho} + \frac{1}{4}\cos(2\frac{L}{\rho})\right]$$

$$+\gamma_0\rho^2\left[\frac{3}{2}L - 2\rho\sin\frac{L}{\rho} + \frac{\rho}{4}\sin(2\frac{L}{\rho})\right]$$

where the index 0 refers to the entrance of the bending magnet. A simplified expression can be obtained by approximating the resulting expression to second order in L/ρ. One gets:

$$I \approx (\gamma_0\eta_0^2 + 2\,\alpha_0\,\eta_0\eta'_0 + \eta'^2_0)\,L + (\alpha_0\eta_0 + \beta_0\eta'_0)\,\frac{L^2}{\rho} \qquad (8)$$

$$-(\gamma_0 \eta_0 + \alpha_0 \eta'_0)\frac{L^3}{3\rho} + \left[\frac{\beta_0 L}{3} - \frac{\alpha_0 L^2}{4} + \frac{\gamma_0 L^3}{20}\right]\frac{L^2}{\rho^2}$$

Optimum values of the optical functions at the beginning of the magnet can then be derived from Eq. (8), in order to achieve the minimum emittance. Figure 3 shows how the emittance changes with variations of α_0 and β_0 away from their optimum in the particular case of zero dispersion at one end of the magnet.

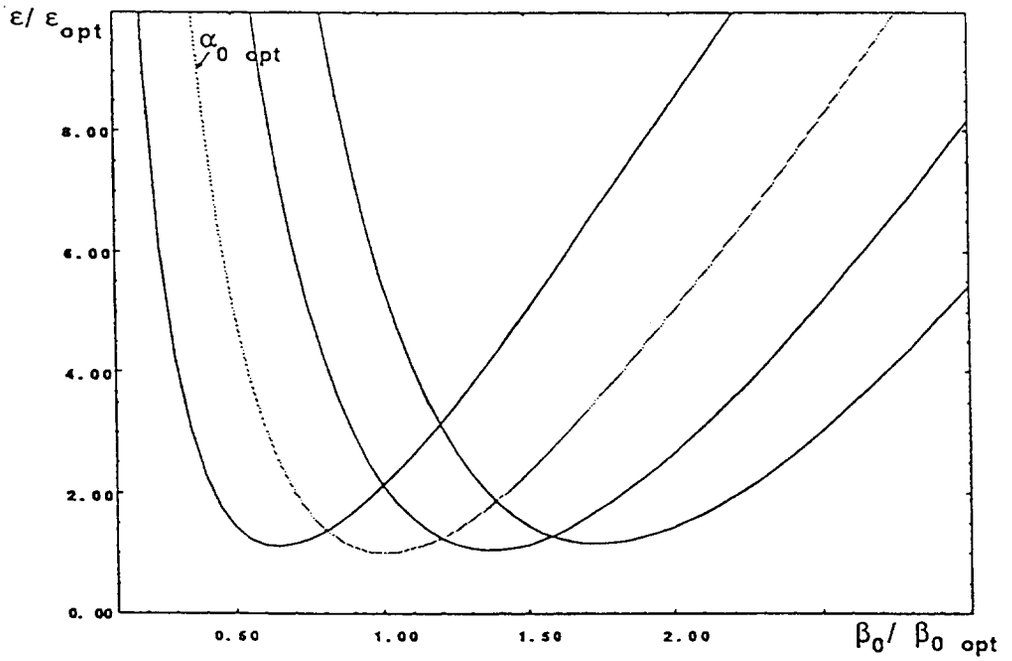


Fig. 3 Beam emittance as a function of initial betatron values

To minimize ε_{x_0}, one can also take into account the variation of J_x. The horizontal partition number is given by [2]:

$$J_x = 1 - \frac{1}{2\pi\rho^2}\int_{mag} \eta\,(1 - 2n)\,ds \qquad (9)$$

where the magnetic field index is defined by: $n = -\frac{\rho}{B}\frac{dB}{dx}$

In a machine with separated functions (i.e. zero field gradient in the dipoles), the quantity J_x is approximately equal to 1. J_x can be made larger by introducing vertical focusing in the dipoles but its maximum value is limited by the rule:

$$\Sigma J_i = J_x + J_y + J_\varepsilon = 4 \qquad (10)$$

which means, in the particular case of no vertical bend, $J_x + J_\varepsilon = 3$, hence $J_x \leq 3$. In practice the maximum tolerable value is $J_x \leq 2$, which results potentially in an emittance reduction of a factor two.

Another possible way to achieve very low emittances is to install damping wigglers in dispersion-free straight sections. For instance, this method would reduce the emittance in PEP by almost one order of magnitude [3] with a total length of installed wiggler magnets of 200 m. This method is applicable to any storage ring but since the damping effect scales proportionally with the bending radius of the ring magnets, it is more efficient in large machines.

3. LATTICE TYPES

3.1 Generalities

The basic structure of low emittance lattices consists of straight sections designed for the installation of insertion devices separated by arc sections. To avoid emittance blow-up by insertion devices, it is usual to adjust the dispersion function in the insertions to zero. In the arcs, the focusing is chosen so as to minimize the chromatic invariant. Minima for H can be found for various types of lattices, involving proper choices for η_x, β_x, and their derivatives in the bending magnet.

A variety of types of lattices are available to design low emittance lattices. They can be divided into two classes:

i) structures without bending magnets in the dispersive section: double focusing achromat (DFA) commonly known as the Chasman-Green lattice [4], expanded Chasman-Green, empty FODO, triplet achromat lattice (TAL).

The double focusing achromat lattice has been used for the NSLS rings in Brookhaven [5]. An expanded Chasman-Green structure is the basis of the conceptual designs of several synchrotron radiation sources under construction: ESRF [6], APS [7], ELETTRA [8]. SUPERACO [9] is already operated with an expanded Chasman-Green structure. The triplet achromat lattice was used in ACO at Orsay [10]; it has the advantage of requiring the shortest circumference.

ii) structures with additional bending magnets in the dispersive region: triplet bend achromat (TBA), FODO.

Triple bend structures are utilized in both the ALADDIN [11] and BESSY [12] synchrotron radiation sources. ALADDIN is a machine optimized primarily to use radiation from the dipole magnet and does not provide zero dispersion in the long straight sections. In the BESSY structure, significant focusing in the vertical plane is provided by the edge focusing of the dipoles. This focusing is necessary in order to limit the amplitude of the vertical beta function.

The triple bend achromat lattice is the logical extension of the DFA. Adding a third bending magnet within the arc section allows extra flexibility. This lattice is proposed for the ALS in Berkeley [13], BESSY 2 [14], SRRC in Taiwan [15].

The FODO lattice is the most commonly used lattice for high energy physics storage rings because of its very compact structure. In contrast to the other types of lattices, the FODO does not naturally provide a space for insertion devices; special lattice sections must be introduced to achieve dispersion-free insertions. The damping rings at Stanford [16] and the SXRL storage ring [17] are good examples of the use of this structure.

3.2 The double focusing achromat

The double focusing achromat lattice or basic Chasman-Green represents the most compact of the structures used in low emittance storage rings. The basic scheme uses two dipole magnets with a focusing quadrupole between them (Fig. 4). The strength of the

quadrupole is adjusted so that the dispersion generated by the first dipole is cancelled by passing through the second dipole.

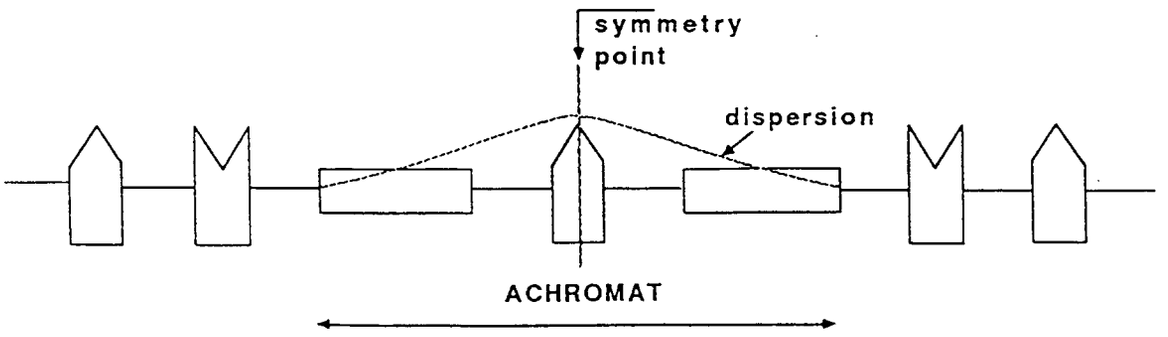


Fig. 4 Basic DFA structure

In this form, the structure is rather inflexible since the quadrupole does not provide focusing in both planes. Therefore, defocusing quadrupoles can be added upstream and downstream of the focusing quadrupole to restore focusing in both planes and to provide more flexibility in the adjustment of transverse dimensions (i.e. the ESRF with four quadrupoles or ELETTRA with three quadrupoles). This is the so-called expanded Chasman-Green achromat. The optical functions of these two lattices are shown in Figs. 5 and 6.

The emittance of a lattice built up of such cells has been computed by various authors [18, 19]. Since the dispersion and its derivative are zero at the entrance of the bending magnet, the emittance is given by:

$$\varepsilon_x = \frac{C_q \gamma^2}{J_x} \theta^3 \left(\gamma_0 \frac{L}{20} + \frac{\beta_0}{3L} - \alpha_0 \frac{L}{4} \right) \tag{11}$$

where β_0, γ_0, α_0 are the Twiss parameters at the entrance of the bending magnet, L is the length of the magnet and θ is the bending angle.

The emittance can be minimized by a proper adjustment of the optical parameters at the entrance of the bending magnet. The minimum value is achieved when the minimum of the horizontal beta function within the dipole occurs at a distance s = 3/8 L from the beginning of the magnet and the value of the minimum betatron function is :

$$\beta = L \frac{\sqrt{3}}{8\sqrt{5}} \tag{12}$$

It yields an emittance of:

$$\varepsilon_x = \frac{C_q \gamma^2}{J_x} \theta^3 \frac{1}{4\sqrt{15}} \tag{13}$$

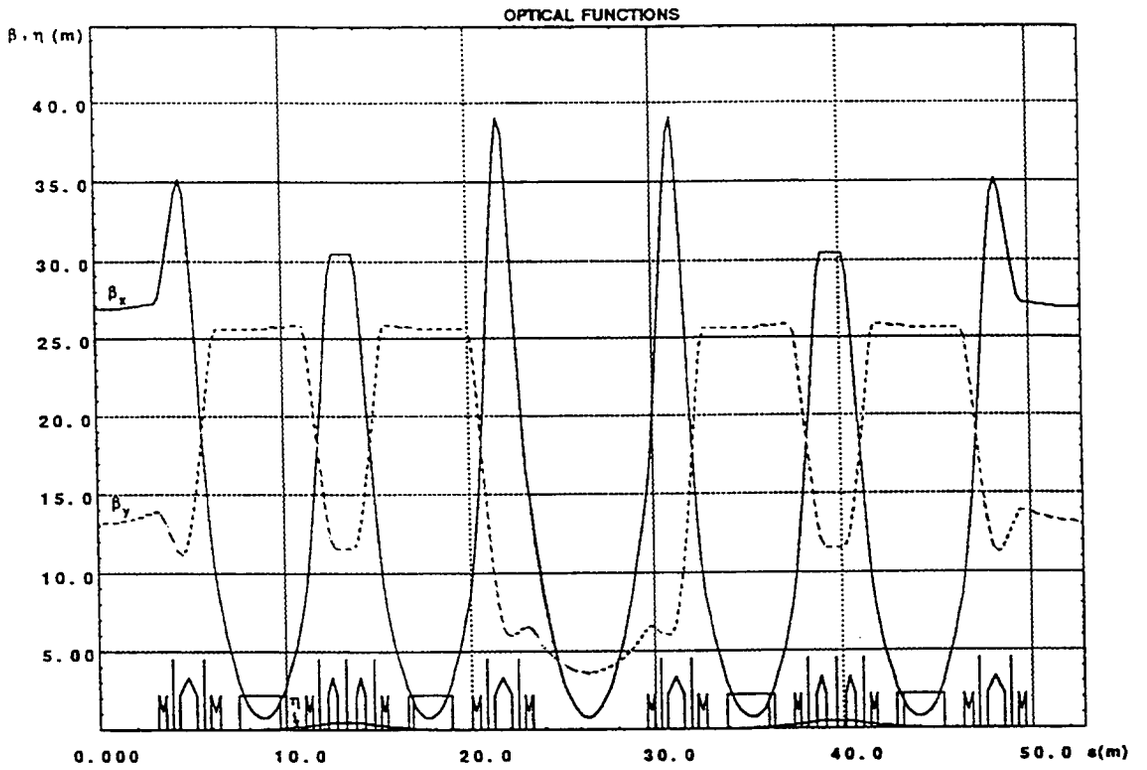


Fig. 5 ESRF lattice functions

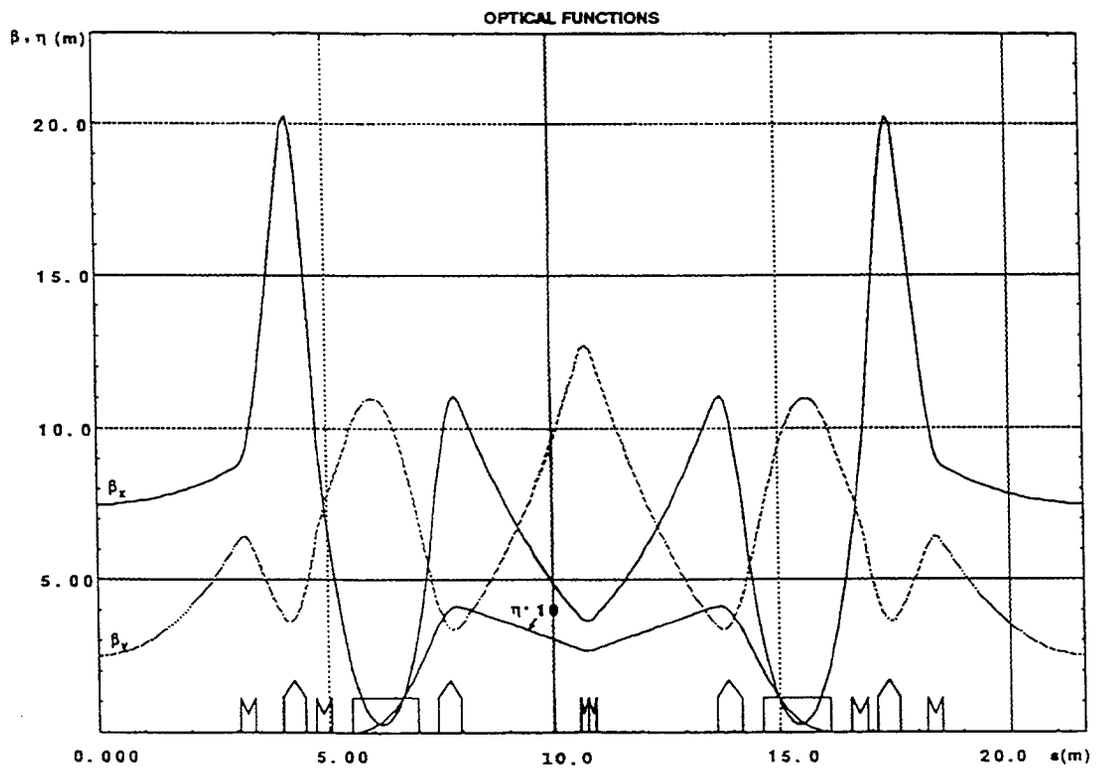


Fig. 6 ELETTRA lattice functions

It is difficult to reach the minimum emittance because the very low β value in the bending magnet (β_{min} = 19 cm for L = 2 m) generates unacceptably high β values in the focusing structure outside the bending magnet and leads to large chromaticities. Therefore, actual lattices result from a compromise with the needs of chromaticity correction and a more realistic estimate would be to increase the emittance by a factor two to three over the ideal minimum value.

The advantage of the DFA structure is to provide a large dispersion between the two bending magnets. As a consequence, sextupole strengths required for chromaticity correction, which are inversely proportional to the dispersion, are reduced. This is the reason why the expanded Chasman-Green is more suitable for large rings where the dispersion in the achromat becomes small. On the other hand, a clear disadvantage of such lattices is the fixed phase advance across the achromat which is approximately π. This places severe constraints on the horizontal betatron tune and limits the flexibility of the lattice since the β values in the insertions are coupled to the choice of tune.

3.3 Triplet achromat lattice

This lattice can be made very compact since there are no quadrupoles in the insertion straight sections, as shown in Fig. 7.

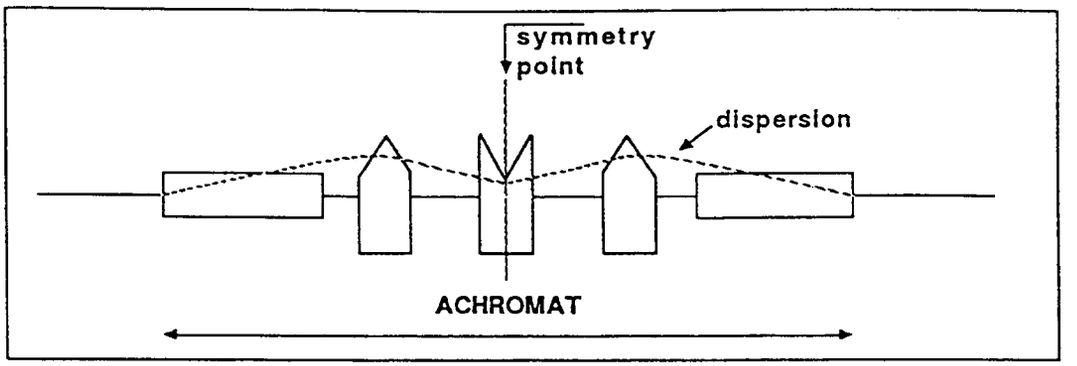


Fig. 7 Basic TAL structure

Using the same derivation as in the previous section, one can express the minimum emittance of the lattice in the form:

$$\varepsilon_x = \frac{C_q \gamma^2}{J_x} \theta^3 \frac{2}{3} \left(\frac{\beta}{L}\right)_{opt} \tag{14}$$

using the optimum value of β in the middle of an insertion section of length $2 L_i$ given by:

$$\left(\frac{\beta}{L}\right)^2_{opt} = \frac{3}{4}\left[\frac{1}{5} + \frac{L_i}{L} + \frac{4}{3}\left(\frac{L_i}{L}\right)^2\right] \tag{15}$$

The main disadvantage of the lattice is that the emittance depends on the value of the horizontal betatron function in the insertion region. Figure 8 shows the ACO lattice functions as an example of a triplet achromat structure.

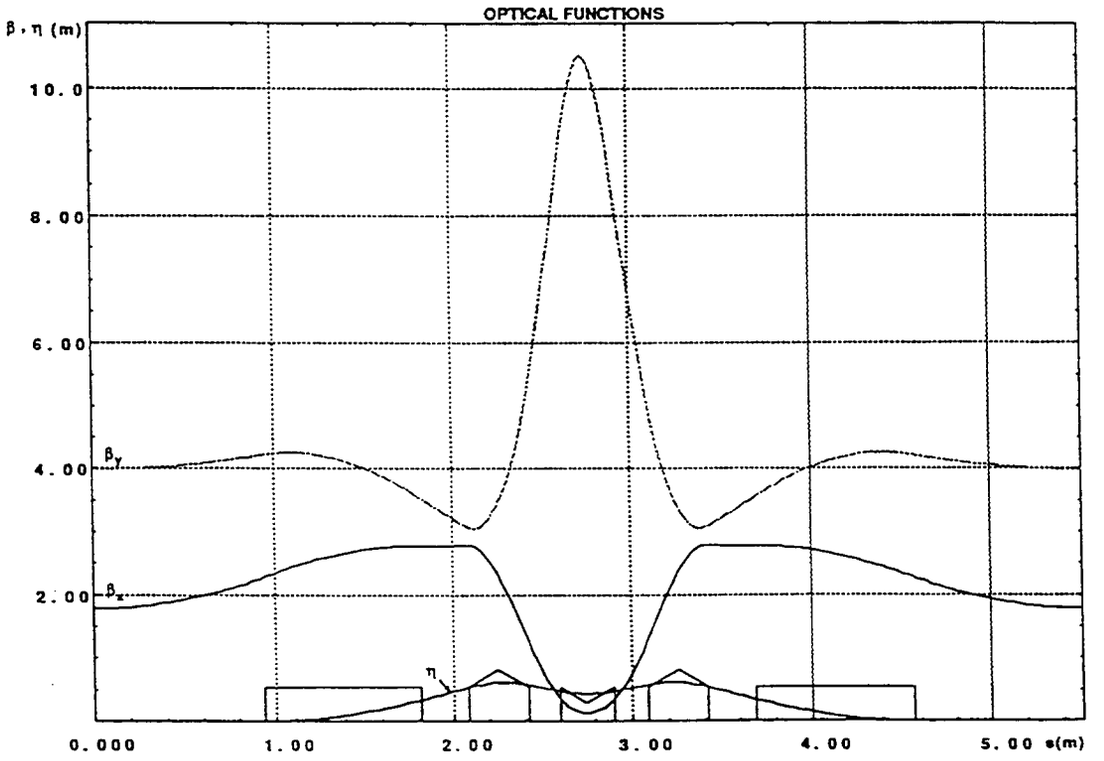


Fig. 8 ACO lattice functions

3.4 Triple bend achromat

A unit cell of the basic TBA lattice with one horizontally focusing quadrupole in-between the bending magnets is shown in Fig. 9. The addition of an extra bending magnet within the achromat gives more flexibility for adjusting the phase advance across the achromat. The phase advance can be varied from π to 2π by changing the positions of the achromat quadrupoles and/or the length of the central magnet compared to the outer magnets [20]. A more practical way to obtain a "movable" quadrupole with fixed magnet geometry is to include a second family of quadrupoles in the achromat.

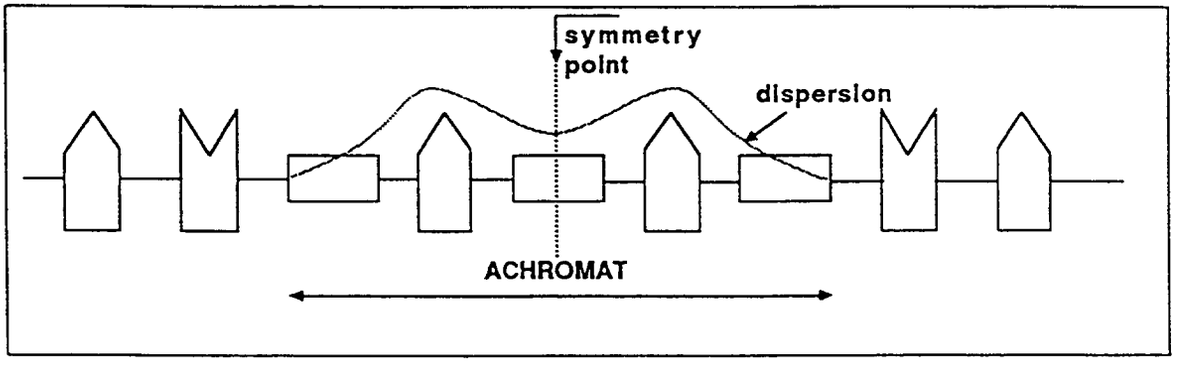


Fig. 9 Basic TBA structure

The emittance of the TBA lattice is obtained by summing the values of the driving integral $I = \int H\, ds$ over the two types of magnets. It can be expressed in the following form:

$$\varepsilon_x = \frac{C_q \gamma^2}{J_x} \frac{\frac{l_1}{\rho_1^3} \frac{2 l_0}{\rho_0^3}}{\frac{\theta_1}{\rho_1} + \frac{2 \theta_0}{\rho_0}} \tag{16}$$

where the index 0 refers to the outer magnet (radius ρ_0 and deflection angle θ_0) and the index 1 to the inner magnet.

Assuming equal bending magnet deflection, the minimum emittance is given by:

$$\varepsilon_x = \frac{C_q \gamma^2}{J_x} \theta^3 \frac{7}{36\sqrt{15}} \tag{17}$$

It is obtained for β values in the outer bending magnets identical to those of the double focusing achromat structure and for $\beta_x = L/\sqrt{15}$ in the inner magnet. Even lower emittance values can be obtained by optimizing the bending angle distribution. The optimum yields a structure in which the centre deflection angle is 1.5 times larger than the outer deflection angle.

To control the vertical beta function in the bending magnets, vertical focusing can be introduced either in the form of lumped quadrupoles adjacent to the bending magnet or by introducing a gradient into the bending magnet fields, as proposed by Vignola [21]. This also changes the damping partition numbers in such a way as to reduce the emittance. Such a solution with combined function magnets has been adopted in the ALS design (Fig. 10).

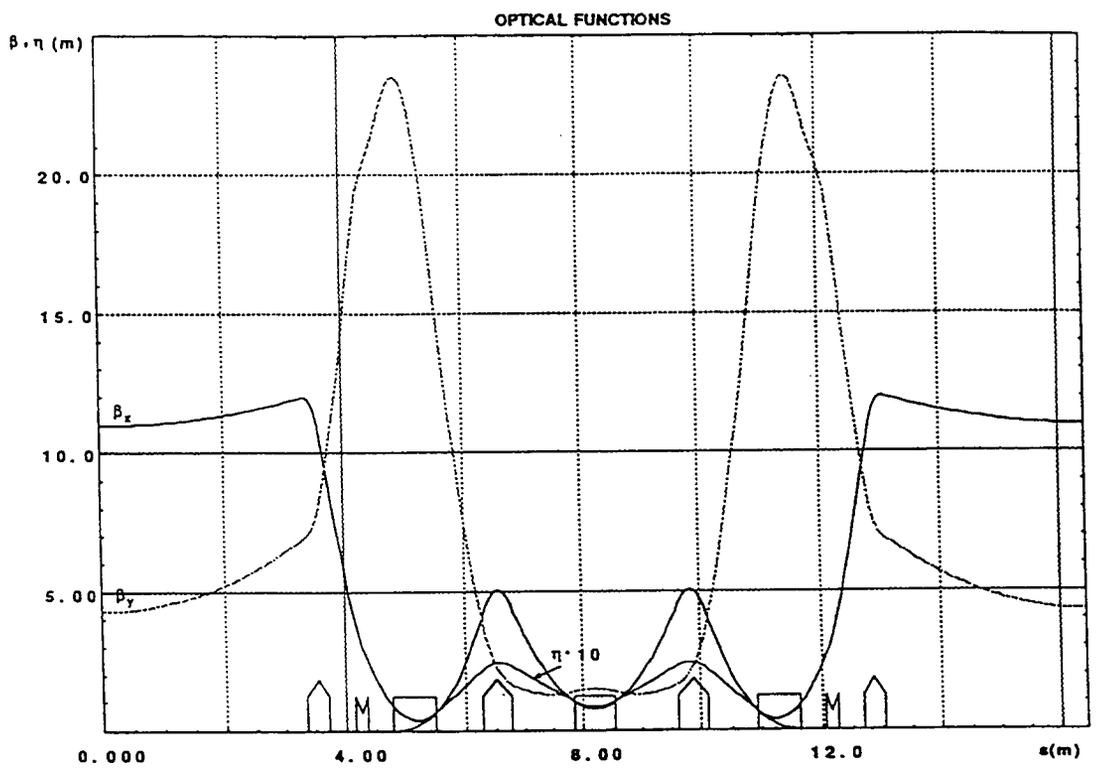


Fig. 10 ALS lattice functions

However, to keep the emittance small, the dispersion in the inner magnet must be reduced to small values which leads to large correcting sextupoles. When compared to the case of a

double focusing achromat lattice with comparable emittance, these sextupoles need to be about two or three times stronger. Such structures are more suitable for small rings where the deflection angle per bending magnet, and correspondingly the dispersion, is larger.

3.5 FODO structures

As used in storage rings constructed for high energy physics, a FODO lattice consists of a sequence of alternating focusing and defocusing quadrupoles separated by bending magnets. The emittance of a regular FODO structure scales as [22]:

$$\varepsilon_x = 4\frac{C_q \gamma^2}{J_x}\theta^3 F(\mu_c) \tag{18}$$

where $F(\mu_c)$ is a function of the betatron phase advance per cell μ_c.

The variation of $F(\mu_c)$ is plotted in Fig. 11. The minimum is reached at about $\mu_c = 135°$ and is rather flat in the range $100° \leq \mu_c \leq 160°$. The large phase advance required to achieve a small emittance implies very strong focusing.

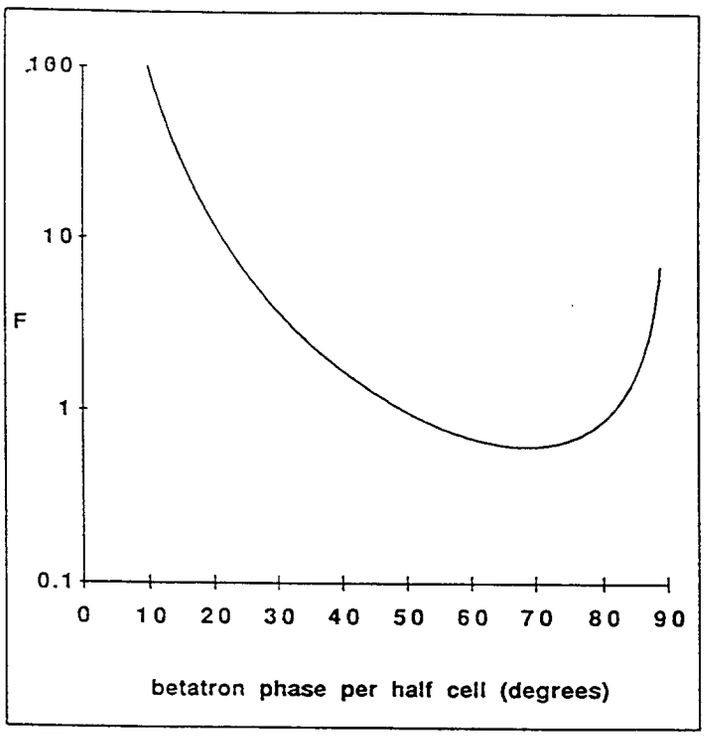


Fig. 11 Emittance behaviour as a function of the betatron phase per half cell

The FODO achromat is composed of a regular FODO structure followed by a dispersion suppressor, as shown in Fig. 12.

The overall emittance is produced by the contribution from the nominal FODO cell bending magnets and the dispersion suppressor bending magnet. It can be expressed as [23]:

$$\varepsilon_x = \frac{C_q \gamma^2}{J_x} \frac{\frac{\eta_0 l_0}{\rho_0^3} + \frac{2 l_1}{\rho^3}}{\frac{\eta_0 \theta_0}{\rho_0} + \frac{2 \theta_1}{\rho}} \tag{19}$$

where n_0 is the number of FODO half cells per achromat and $\theta = L / \rho$.

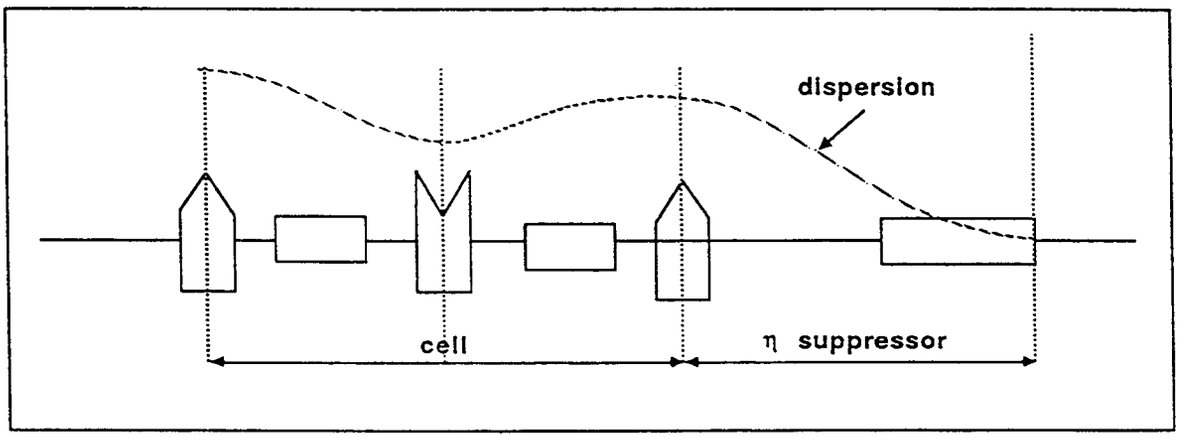


Fig. 12 Basic FODO achromat

Once the cell geometry and the phase advance are chosen, the emittance can be minimized by optimizing the length of the dispersion suppressor magnet and the bending angle distribution. Figure 13 (taken from [23]) shows the variation of emittance with these parameters.

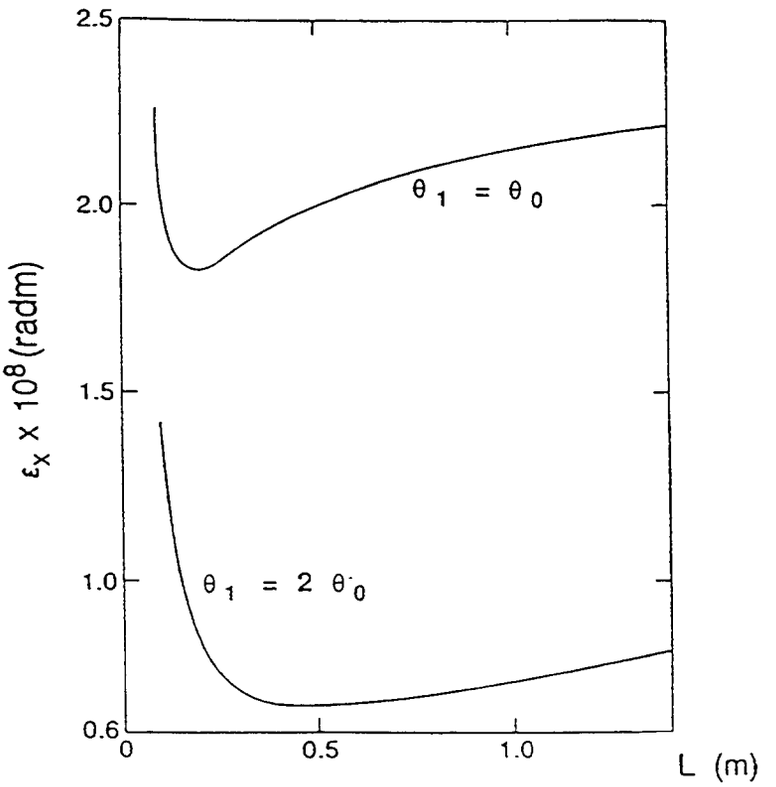


Fig. 13 Variation of emittance with length of dispersion suppressor

The FODO has the advantage of large flexibility. Also, since the average betatron function is reduced, the vacuum requirements can be relaxed. On the other hand, it is characterized by an extremely dense packing of magnetic elements which could lead to layout and engineering problems. Also it suffers from intrinsic problems similar to those of the triplet bend achromat, because of the small dispersion which is required to achieve a low emittance and leads to large sextupole strengths. The lattice functions of a FODO type structure are given in Fig. 14.

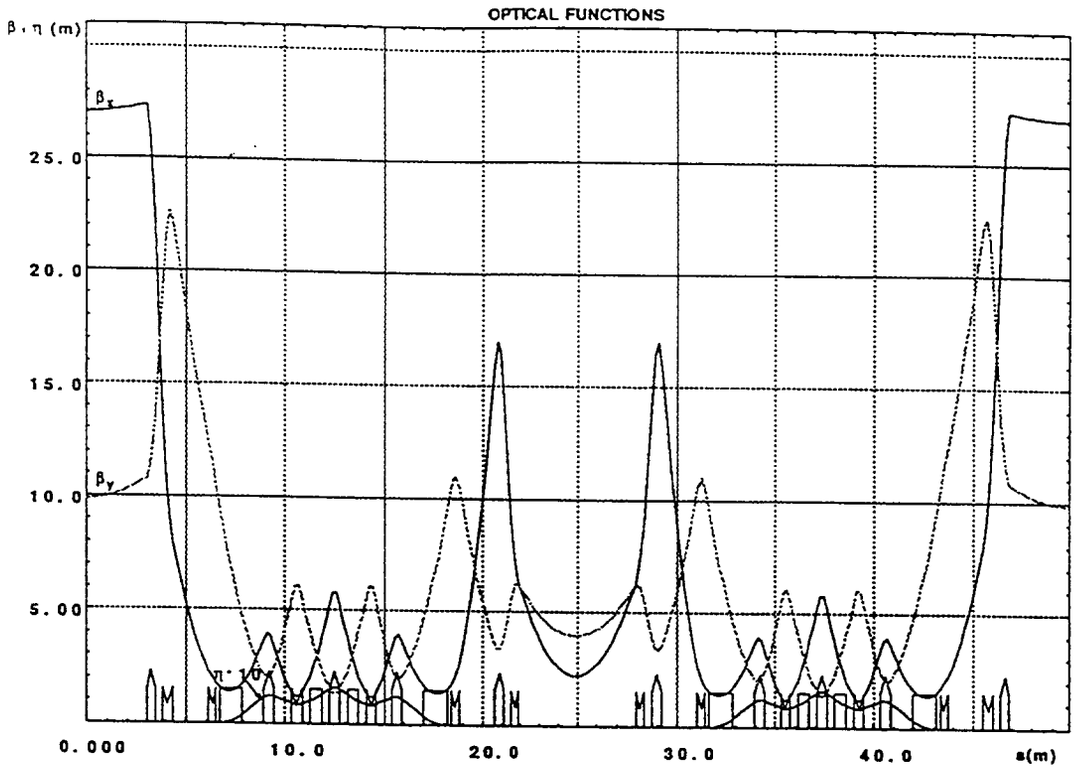


Fig. 14 Lattice functions of a FODO type structure

3.6 Comparison of the lattice performance

It can be seen from the previous sections that the beam emittance of all these lattices scales as $E^2\ \theta^3$ where E is the beam energy and θ the bending angle. This means that, in order to achieve a low emittance, the bending angle must be decreased and a large number of cells is needed. In order to make a comparison of the various types of lattices, it is useful to rewrite the emittance in the following form:

$$\mathcal{E}_x = \frac{E^2}{J_x N_d^3} F\,(\mu_x, \textit{lattice}) \tag{20}$$

where N_d is the number of bending magnets and E the beam energy (expressed in GeV).

Table 1 lists the minimum values of F which are achievable for the various lattices and gives an example of each lattice case. F is varying over a range of roughly a factor ten from the FODO lattice down to the DFA lattice. But because of the cubic dependence on N_d, even a factor ten can be compensated by a factor of about two in N_d. Thus, to get the same emittance with a FODO, one has to double the number of bending magnets in the ring. Design considerations imply in most cases that the emittance is larger than the minimum value.

Table 1
Minimum values of F

	F_{min}	Examples	Energy (GeV)	N_d	F
DFA	$2.35\ 10^{-5}$	ESRF	6	64	$5.02\ 10^{-5}$
TAL	$2.45\ 10^{-5}$(*)	ACO	0.536	8	
TBA	$1.83\ 10^{-5}$	ALS (**)	1.5	36	$1.14\ 10^{-4}$
FODO(***)	$4.52\ 10^{-4}$				

(*) to be multiplied by $(\beta / L)_{opt}$
(**) with gradient in the dipoles
(***) regular cells without dispersion suppressor

The emittance is an important issue when choosing a lattice. However criteria for evaluation of lattices include other considerations, such as chromatic correction and dynamic aperture, sensitivity to errors, flexibility, beam lifetime.... Some of these problems will be discussed in Section 3.8.

3.7 New ideas for ultra-low emittance lattices

The lattices considered at present achieve low emittance by minimizing the dispersion function and the horizontal betatron function in the bending magnets. This design philosophy encounters its natural limits in the increasing difficulty to correct the chromatic errors due to the high focusing, especially as more than half of the chromaticity is created by quadrupoles in the dispersion-free sections.

To avoid these limitations, a new model has been recently proposed [24]. The design strategy consists in compensating the chromatic errors where they are generated. This is possible by maintaining dispersion throughout the lattice and integrating sextupole fields into the focusing elements. Nevertheless, in order to achieve a small emittance, the dispersion is minimized by stretching the bending over practically the total circumference and applying an extremely high focusing.

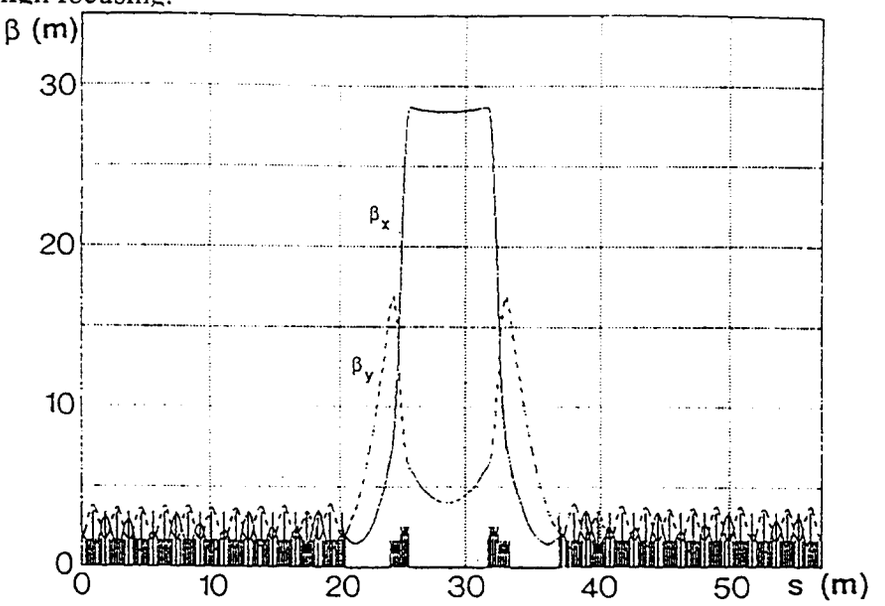


Fig. 15 Betatron functions of the ultra low emittance model

Obviously, this design philosophy is partly spoilt by the requirement of dispersion-free straight sections for insertion devices. Figure 15 shows the optical functions of a lattice composed of regular cells and a dispersion suppressor region. The structure is based on the use of multipole magnets with strong quadrupole fields integrated and with a low dipole field superimposed. An emittance of 0.7 nanometer-radians is achieved for a 6 GeV ring. The structure also provides one-order-of-magnitude larger momentum acceptance than present designs. This kind of lattice might be very attractive for synchrotrons and damping rings.

3.8 Problems associated with low emittance lattices

3.8.1 Chromaticity correction and dynamic aperture

Achieving such low emittances means that the storage ring will be highly sensitive to non-linear effects. One of the main problems raised is the correction of chromaticity. The chromaticity is the tune shift experienced by individual particles due to momentum errors $\Delta p/p$. It is defined as:

$$\xi_{x,y} = \frac{\Delta \upsilon_{x,y}}{(\Delta p/p)} = -\frac{1}{4\pi}\int \beta_{x,y}(s)\, K_{x,y}(s)\, ds \qquad (21)$$

where $K_{x,y}$ is the quadrupole gradient.

The strong focusing required to obtain a low emittance implies strong gradients and also large values of betatron functions so that the chromaticity tends to be large. As a consequence, off-momentum particles will cross resonances and may be lost. In addition, to prevent instabilities and the so-called "head-tail" effect, the chromaticity must be corrected or adjusted to a slightly positive value. This is performed by means of sextupoles located in the dispersive region. The momentum error produces an orbit shift which generates a quadrupole field at each sextupole location so that the correcting effect is given by:

$$\xi_s = \frac{1}{4\pi}\int \beta(s)\, \eta(s)\, m(s) ds \qquad (22)$$

where $m(s)$ is the sextupole gradient.

As a matter of fact, these sextupoles need to be very strong since the dispersion generated by the bending magnets is rather small in all lattices. Unfortunately, these sextupoles introduce various kinds of chromatic and geometric aberrations, such as betatron amplitude-dependent and momentum-dependent tune shifts, change of the betatron functions and dispersion with momentum. These aberrations limit the maximum stable amplitudes of oscillations. The dynamic aperture, defined as the boundary of stable motion in the x-y plane, may be considerably smaller than the physical aperture. Adequate dynamic aperture is an important figure of merit. The transverse aperture must be large enough to accommodate the oscillations of the injected beam and Coulomb scattered particles. Without this, the injection efficiency would be very poor and the beam lifetime very short. A large chromatic aperture is required to improve the lifetime due to Touschek and bremsstrahlung effects.

In order to obtain reasonably large dynamic apertures and minimize all detrimental effects, complicated arrangements of sextupoles have to be found for each lattice. In particular, additional sextupoles located in the dispersion-free straight sections are used to enlarge the dynamic aperture and reduce the tune shifts with amplitude. With regard to dynamic aperture, all lattices can be brought to comparable performance as shown in Fig. 16 for the ESRF.

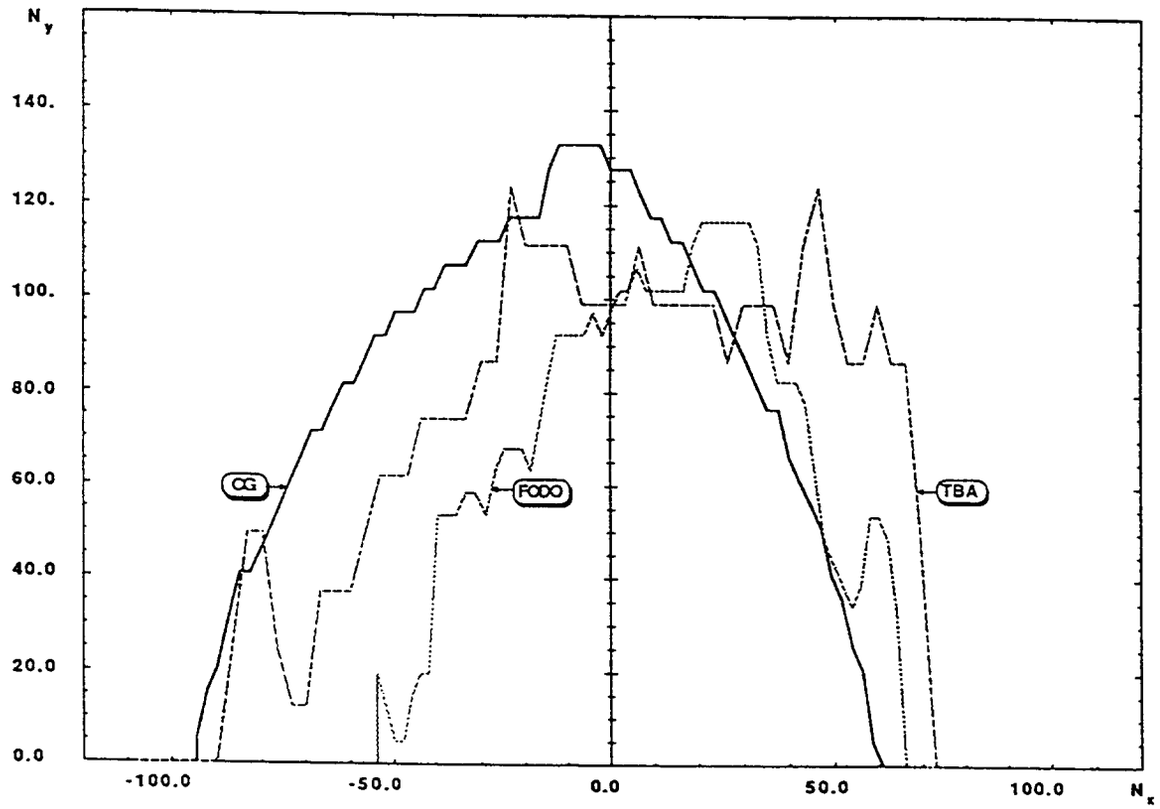


Fig. 16 Dynamic aperture of CG, TBA, FODO lattices for the ESRF

3.8.2 Sensitivity to errors

Errors of all kinds (magnet misalignments, systematic and random field errors) will significantly reduce the dynamic aperture. Because of the strong focusing, the amplification of quadrupole and sextupole misalignments is very large and generates unacceptable distortions of the linear optics, reduction of dynamic aperture, beam emittance blow-up or even unstable machines. The severity of this problem is illustrated in Fig. 17 which shows that a significant reduction of dynamic aperture is induced by very small quadrupole positioning errors (in that case the effects of a misalignment of 0.025 mm have been computed for different sample machines).

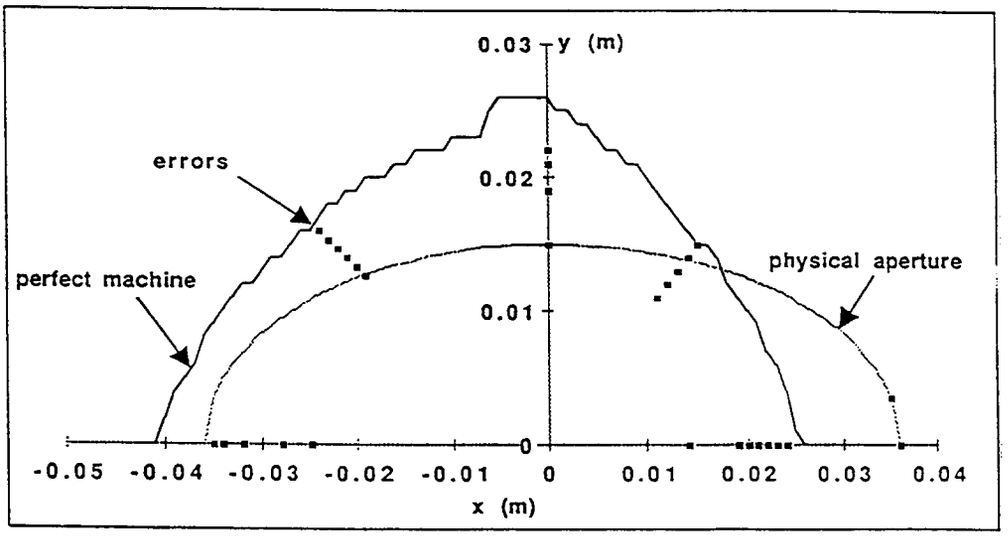


Fig. 17 Effects of quadrupole positioning errors on dynamic aperture

As a consequence, state-of-the-art alignment and construction capabilities (1/10 mm positioning errors for quadrupoles and sextupoles, tilts of magnetic elements of a few 10^{-4} rad) are required. Since a quite perfect closed orbit (i.e. 0.1 to 0.2 mm rms residual distortions) is necessary to achieve correct performance, sophisticated closed orbit correction schemes, including dynamic feedback systems to control beam stability, must be implemented.

3.8.3 Orbit stability

Orbit stability has more severe requirements than for other accelerator applications because of the very small vertical beam size at the source point (a typical figure is $\sigma_y < 100$ μm at undulator locations). Any change of the beam position with time will lead to an enlargement of the source and a macroscopic emittance growth. To avoid spoiling the source emittance, the beam centre of mass must be kept within a few μm and a few 1/10 μrad all around the machine.

The tolerances on the stability of the beam position concern both the long-term stability, therefore deal with ground settlement, and the short-term motion induced by vibrations transmitted through the ground.

As already pointed out, the closed orbit is extremely sensitive to quadrupole centre displacements. Therefore, the buildings, infrastructure, slab and magnet supports must be carefully designed to minimize transmission of vibrations to the magnetic elements. Figure 18 shows an example of tolerances on vibration amplitudes measured at the location of magnetic elements.

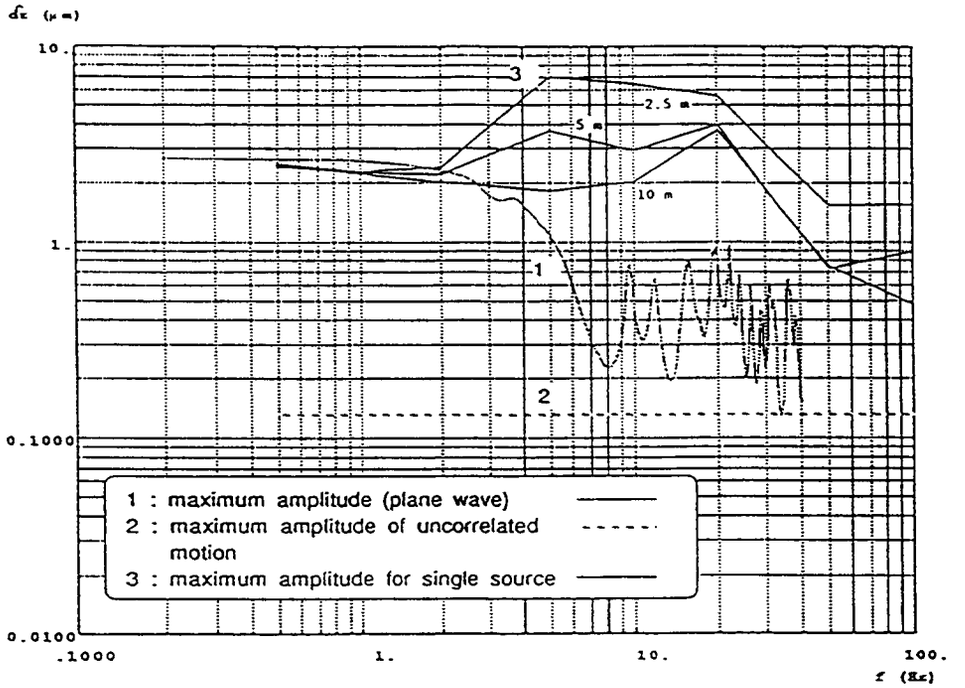


Fig. 18 Allowable amplitude of vibrations

The standard closed orbit position monitors are not able to guarantee the requested stability. Therefore, it is necessary to use feedback systems based on the detection of the photon beam and local steering magnets.

3.8.4 Small momentum compaction

Low emittance lattices tend to have a very low momentum compaction factor a, because they usually have low dispersion. Typical values are :

$\alpha \sim 3.0\ 10^{-4}$	for APS, ESRF
$\alpha \sim 1.5\ 10^{-3}$	for ALS, ELETTRA

When α becomes very small, non-linear contributions become important and can lead to longitudinal instabilities. The low value of the momentum compaction also affects bunch length and very short bunches can be obtained.

* * *

REFERENCES

[1] M. Sands, The Physics of Electron Storage Rings, SLAC-121 (1970)

[2] H. Bruck, Accélérateurs Circulaires de Particules (PUF, Paris, 1966)

[3] H. Wiedemann, An Ultra-Low Emittance Mode for PEP Using Damping Wigglers, Nucl. Instr. Meth. A266 (1988)

[4] R. Chasman and K. Green, Preliminary Design of a Dedicated Synchrotron Radiation Facility, IEEE Trans. Nucl. Sci., NS-22, 1765 (1975)

[5] L. Blumberg et al., National Synchrotron Light Source VUV Storage Ring, IEEE Trans. Nucl. Sci., NS-26, 3842 (1979)

[6] ESRF Foundation Phase Report (Feb. 1987)

[7] 6 GeV Conceptual Design Report (Feb 1986), ANL-86-8

[8] Design Study for the Trieste Synchrotron Light Source (Feb. 1987), LNF-87/6

[9] H. Zyngier et al., The VUV Radiation Source Super-Aco, IEEE Trans. Nucl. Sci., NS-32, 3371 (1985)

[10] Proc. of the 5th Int. Conf. on High Energy Accelerators, Frascati (1965)

[11] E. Rowe et al., Status of the Aladdin Project, IEEE Trans. Nucl. Sci., NS-28, 3145 (1985)

[12] D. Einfeld and G. Mülhaupt, Choice of the Principal Parameters and Lattice of BESSY, Nucl. Instr. Meth. 172 (1980)

[13] 1-2 GeV Synchrotron Radiation Source, Conceptual Design Report (July 1986), LBL Pub 5172

[14] R. Maier et al., BESSY 2, a Synchrotron Light Source of the Third Generation, IEEE Particle Accelerator Conference, 422 (1987)

[15] SRRC Status Report (Jan. 1988)

[16] G.E. Fisher et al., A 1.2 GeV Damping Ring Complex for the Stanford Collider, 12th Int. Conf. on High Energy Accelerators, Batavia (1983)

[17] L. Emery et al., The 1.2 GeV High Brightness Photon Source at the Stanford Photon Research Laboratory, IEEE Particle Accelerator Conference, 1496 (1987)

[18] M. Sommer, Optimisation de l'Emittance d'une Machine pour Rayonnement Synchrotron, DCI /NI/20/81 (1981)

[19] H. Wiedemann, Linear Theory of the ESRP Lattice, ESRP-IRM/9/83 (1983)

[20] A. Jackson, A Comparison of Chasman-Green and Triple-Bend Achromat Lattices, Particle Accelerators, Vol 22, p 111 (1987)

[21] G. Vignola, The Use of Gradient Magnets in Low Emittance Electron Storage Rings, Nucl. Instr. Meth, A 246 (1986)

[22] H. Wiedemann, Brightness of Synchrotron Radiation from Electron Storage Rings, Nucl. Instr. Meth., 172 (1980)

[23] A. Wrulich, Study of FODO Structures for a Synchrotron Light Source, Particle Accelerators, Vol 22, p 257 (1988)

[24] W.D. Klotz and G. Mülhaupt, A Novel Low Emittance Lattice for High Brilliance Electron Beams, submitted to Nucl. Instr. Meth.

SMALL RING LATTICE PROBLEMS

E.J.N. Wilson
CERN, Geneva, Switzerland

Abstract
This is a review of the special problems in particle dynamics, lattice design, and magnet construction and measurement which should be considered in the design of small synchrotrons and storage rings. The CERN Antiproton Accumulator is used as an illustration.

1. INTRODUCTION

This report sets out the practical problems which face the designer of a small storage ring: the CERN Antiproton Accumulator. The designers of other small rings for electrons and high intensity proton synchrotrons will encounter similar problems. Their solutions may differ according to the application but the example will be instructive.

Synchrotrons built for the highest energy electrons and protons are built of hundreds and even thousands of small aperture magnets several meters in length. There are usually several hundred regular periods and enough space in the six or eight insertions to string together a series of purpose-built sections for injection, ejection, dispersion correction and low-beta matching. In a small ring the number of magnets and periods is severely restricted and one must often exercise considerable ingenuity to arrive at a design which satisfies all the requirements necessary to arrive at the desired performance.

The ends of the long high-energy-machine magnets constitute only a small fraction of the integrated field seen by a circulating beam and the fact that the fields in the ends are three-dimensional can usually be ignored. Indeed many of the computer programs developed for the design of large rings contain approximations which treat the magnets as pure, two-dimensional fields and ignore the small curvature of the central trajectory through the end field. Such approximations must be reviewed critically by the designer of a small ring where the magnet aperture can become comparable to the length and where much of the magnet's effect comes from a fringe field in which the particle is deflected with a radius of curvature comparable to the magnet's length.

In this report we first discuss how one may satisfy a number of design constraints in a small ring like the Antiproton Accumulator, how measurement and correction of the end field can be combined with Q measurements on the finished machine to correct effects introduced by short magnets of large aperture, and finally we consider a class of end effects which are not normally included in even the most rigorous lattice programs.

2. THE EXAMPLE – AN ANTIPROTON ACCUMULATOR

In the photograph taken through a "fish-eye" lens, we see the large bending dipoles and focusing quadrupoles of the Antiproton Accumulator which has a mean radius of 25 m. The objects to be seen in Fig. 1, wrapped in shiny aluminium bake-out jackets and installed in between some of the magnets are pick-ups and kickers which are mainly used for stochastic cooling and do not concern us for the purposes of this talk.

In the bottom right-hand corner of the plan view shown in Fig. 2, we see the target where antiprotons are produced and a short transport line which brings them to a point, at 12 o'clock on the "dial" of the ring, where they are bent by a septum magnet to join the circulating beam.

A little further around, at about 2 o'clock on the dial, is a kicker magnet which inflects them onto the injection orbit. At 10 o'clock is an ejection kicker which uses the same septum to extract the beam which has been accumulated and stored.

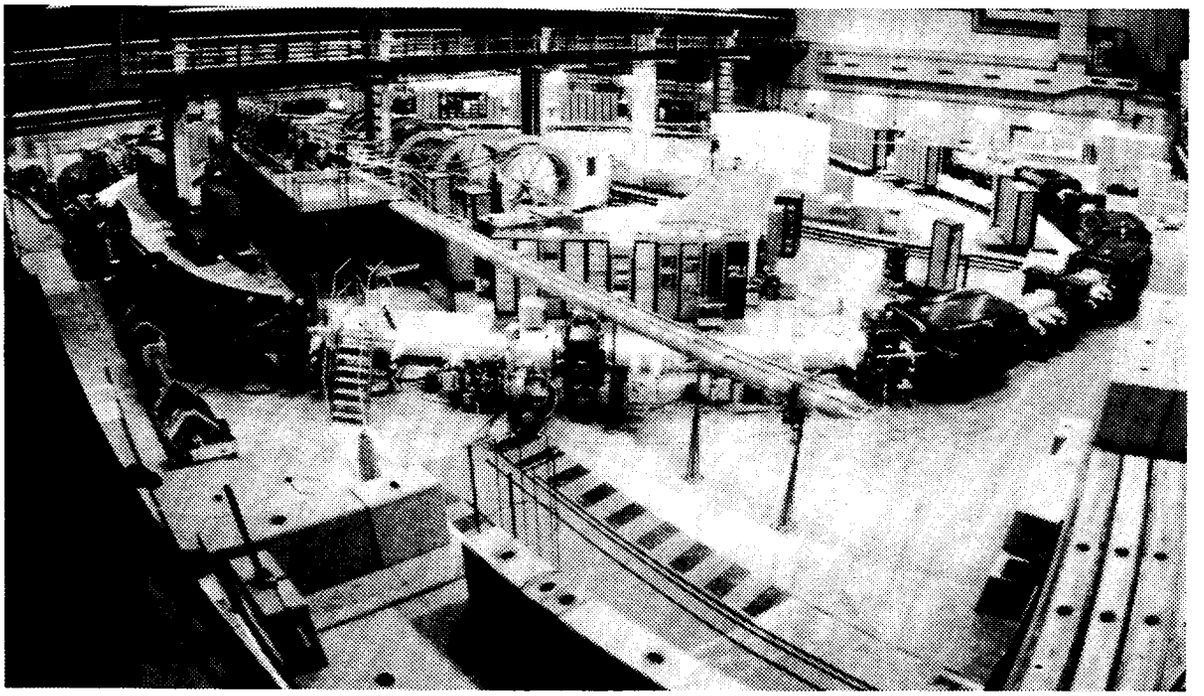


Fig. 1 A fish-eye view of the Antiproton Accumulator

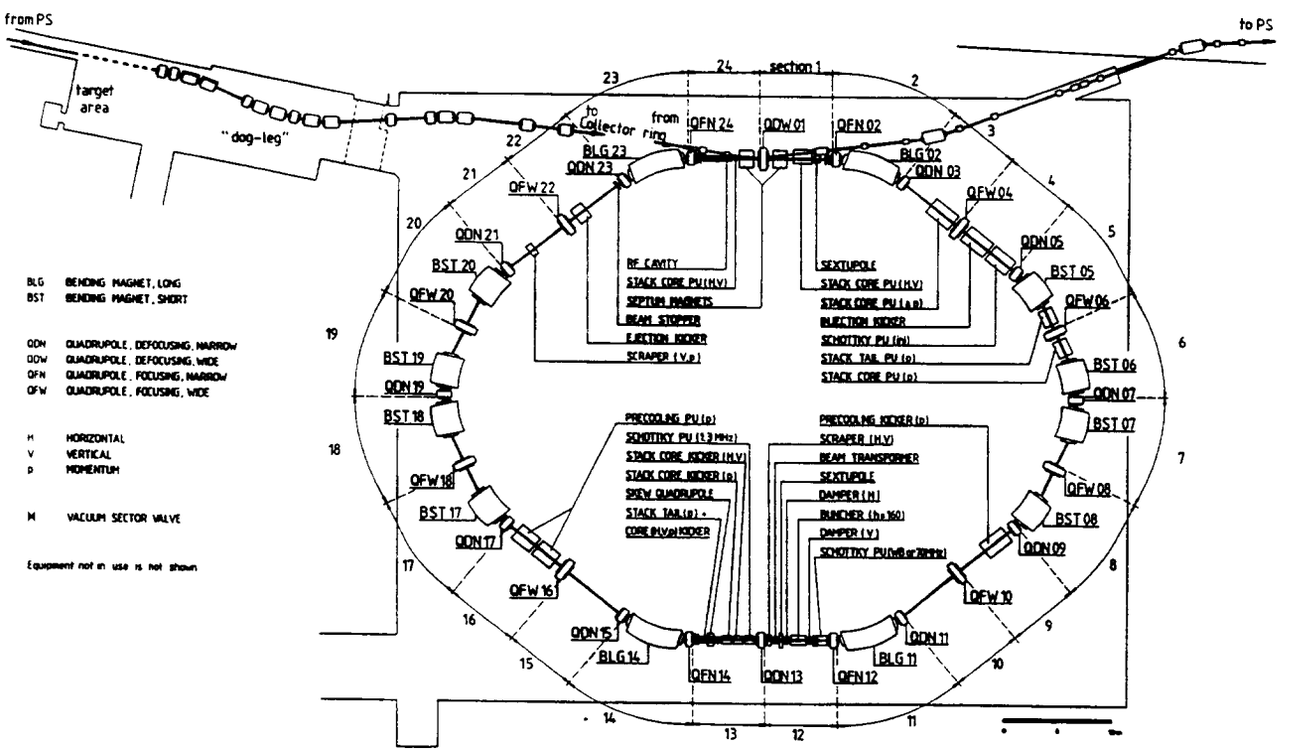


Fig. 2 General layout of the Antiproton Accumulator

One of the constraints on the design is that these kickers should be located at a particular horizontal betatron phase advance from the septum. This should be rather close to an odd multiple of 90 degrees. We shall see that such injection and ejection details often have quite strong influence on the lattice design and should not be left until after the major parameters have been frozen.

One of the problems which arises naturally when the ring is small is that there is not enough room for all the components and this leads to a great shortage of space between the magnets. In Fig. 3 we see a typical gap between quadrupole and bending magnet. We must be careful to ensure that their fields do not interfere or, if they do, that we know from measurements with the neighbouring magnet in place, how the effective length of the magnet and quadrupole are affected by the other's presence. We can see in Fig. 4 some early measurements of quadrupole gradient along a line parallel to the axis which were made in the 1950's [1], that a steel plate acting as a mirror to simulate the presence of another magnet has a significant effect on the effective length and central field gradient.

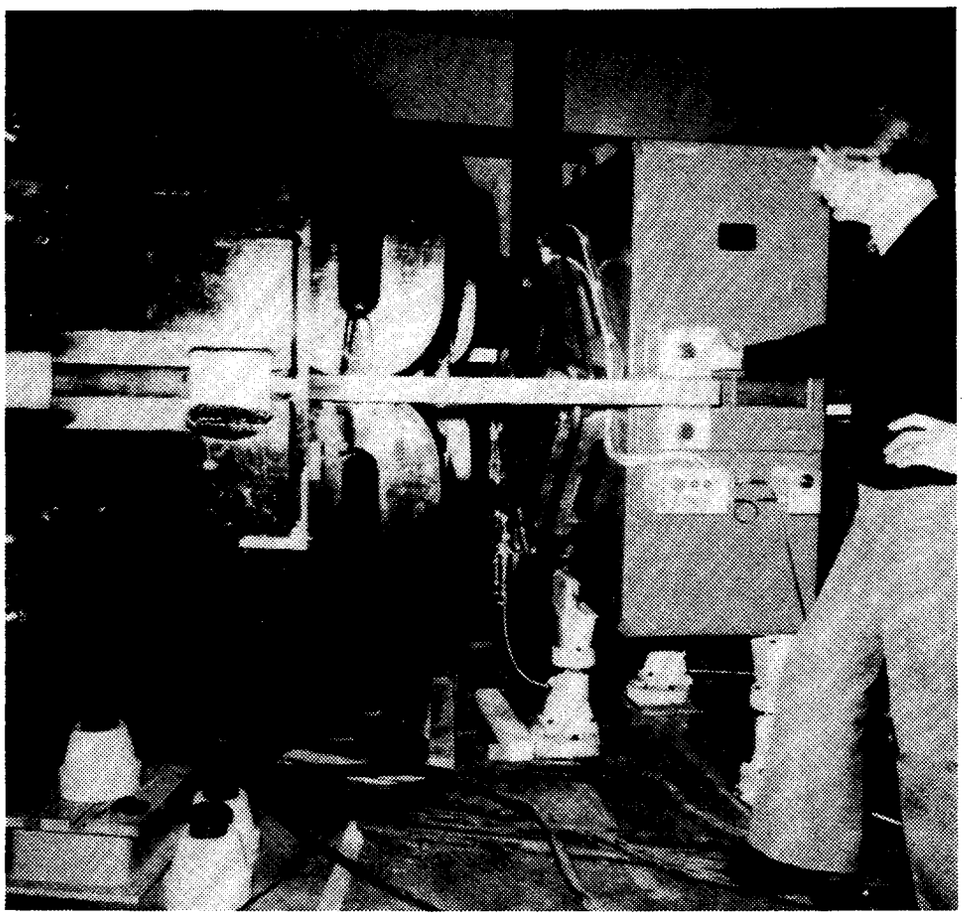


Fig. 3 Gap between a dipole and a quadrupole in the AA

In arriving at a suitable lattice design for this ring we first chose the energy to be 3.5 GeV since this is the energy at which the antiprotons are most abundant and we tried to make the radius 25 m, so that its circumference is exactly one quarter of that of the PS. The bunches of protons which produce the antiprotons originate in the PS and it is this machine which will later have to re-accelerate the antiprotons.

It is natural to first try to design around a FODO pattern of quadrupole lenses since this is the simplest from many points of view. Of course, designers of electron rings for synchrotron radiation usually choose another kind of lattice, but the problems we shall discuss tend to be common to any lattice configuration.

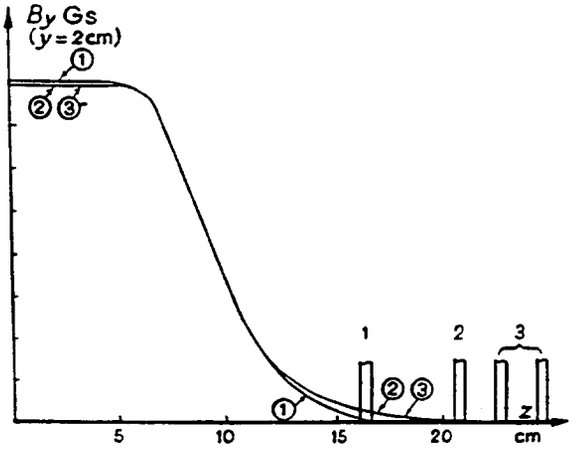


Fig. 4 End field shape for a quadrupole with a steel mirror plate

This particular ring has a specific requirement imposed by the cooling system which demands that the spread of revolution frequencies, which stems from the momentum spread and which is determined by

$$\eta = \frac{1}{\gamma^2} - \frac{1}{\gamma_{tr}^2} = \frac{p}{f}\frac{df}{dp} , \tag{1}$$

is within rather close limits. This fixes γ_{tr}. Other machine designers will encounter different constraints on this parameter. In high intensity boosters it is often thought a good idea not to include γ_{tr} in the energy range of the synchrotron. In electron rings, although one is far above transition, the momentum compaction which is linked to the energy damping time, places a similar constraint on the designer.

The second term in the expression for η is equal to the momentum compaction function divided by bending radius. It is determined by the change in circumference with momentum about the equilibrium momentum and is just the average value of the dispersion function $D(s)$ around the ring:

$$\frac{1}{\gamma_{tr}^2} = \alpha = \frac{1}{2\pi R}\int \frac{D(s)}{\rho(s)}\, ds \tag{2}$$

We shall see that γ_{tr} is controlled by the choice of Q value and indeed, for a proton machine transition, is roughly equal to Q. The Q value is in turn closely tied to one quarter of the number of FODO periods since one tries to choose a betatron phase advance of 90 degrees or perhaps 60 degrees but rarely outside this range. Since the desired γ_{tr} for the AA was 2.4, the Q was chosen to be 2.3 and therefore the number of periods had to be approximately 10.

If the number of periods is too large the space will be chopped up into too many small pieces; there will be too many components and too many wasteful ends to them. Furthermore, the AA ring has a natural symmetry of four and we are left with only 8, 12 and 16 as possible numbers of periods. When we come to examine the relative merits of the few remaining options we find that $N = 8$, would suggest a smaller Q value and hence β, which is roughly R/Q, becomes uncomfortably large. When we considered the acceptance necessary to collect enough antiprotons the apertures of the magnets are large even with $N = 12$. The larger β of $N = 8$ would make matters worse. On the other hand 16 periods would result in too many small components.

Once the period number is fixed, the position of all the quadrupoles is determined and the Q value is fixed within narrow limits. The length of the quadrupoles is also determined by the peak field one may allow on their pole pieces which must inscribe an ellipse or rectangle sufficient to accommodate the beam. The lattice functions and the required emittance and momentum spread (100 π mm. mrad and $\pm$ 3% in the case of the AA) are now pretty well defined and it is merely a matter of juggling to make the exact numbers from a lattice program consistent.

We still have to decide on the layout of bending magnets within the ring and the positions of injection, ejection and (in our example) the cooling systems. The kickers must be 90 degrees (or some odd multiple of this) from the septum and this determines which of the half periods they lie in. The AA ring has a rather special requirement that there will be two "beams" circulating which differ only in momentum as shown in Fig. 5, one is the "stack" where antiprotons are accumulated and the other the injected beam which must be physically separated from the stack at the injection kicker. This magnet, which encloses only the injected beam, has a movable ferrite shutter which closes the magnet aperture on the stack side to prevent the stack from being disturbed by the firing of the kicker. The two beams must be separated by enough space for this shutter.

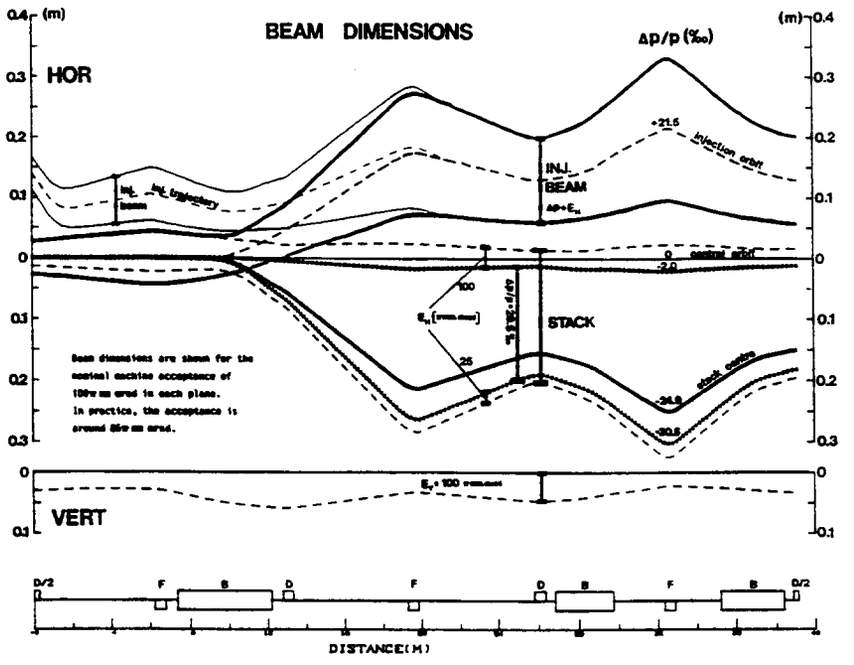


Fig. 5 Beam dimensions for one quadrant of the AA

Elsewhere both beams must pass through a narrow cooling pick-up at 12 and 6 o'clock and there must be a steep rise in the dispersion function between the septum and kicker as can be seen in Fig. 6.

This rapid change in dispersion is achieved by choosing to make the lattice symmetric about a D quadrupole at 6 o'clock. We also make use of the horizontal defocusing properties of the end field of a dipole whose faces are tilted horizontally to make an angle with the beam. Both these features help to raise the dispersion as fast as possible immediately after the first dipole encountered by the incoming beam and then cause it to roll over into a high sustained value in the rest of the super period [2]. The result is to be seen in Fig. 6. This rather peculiar feature is mentioned here only to complete the AA story. Other machines will no doubt have their own special reasons for choosing a particular symmetry or a particular kind of dipole end design.

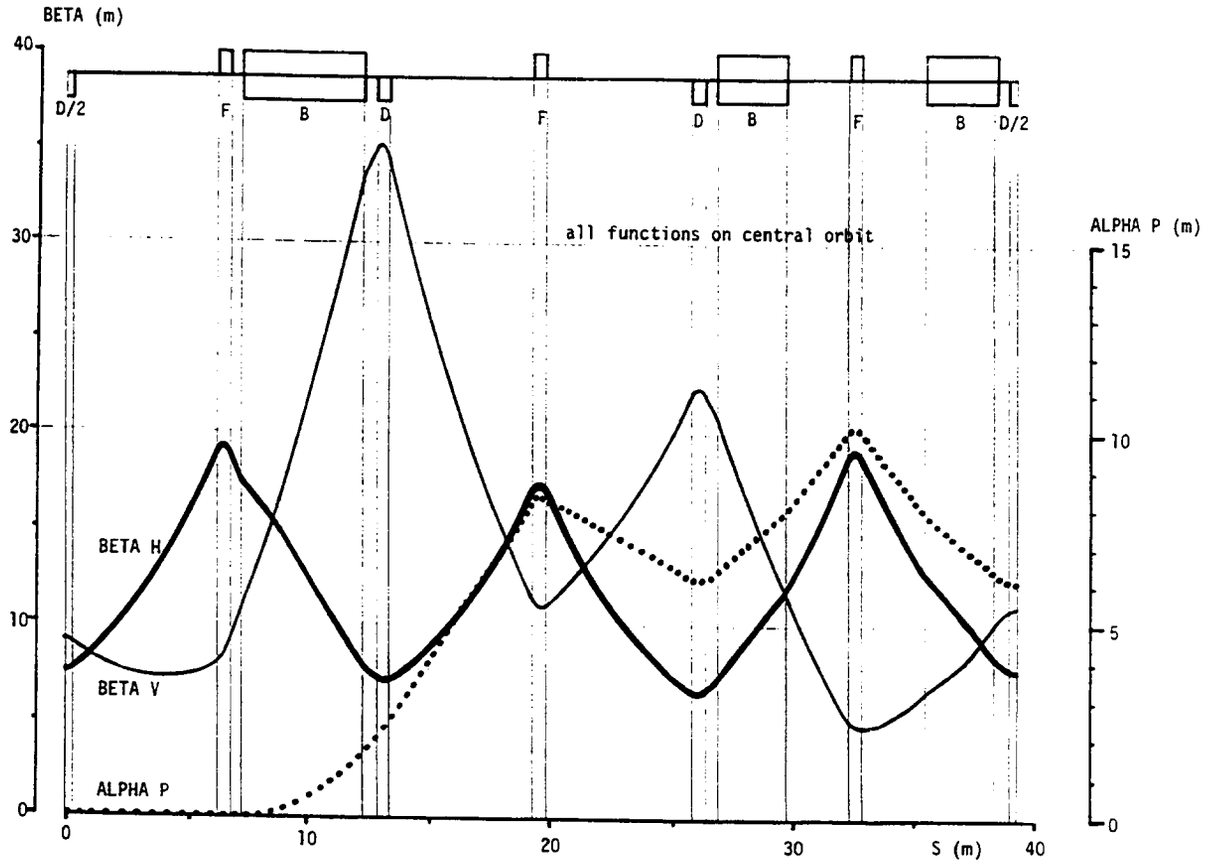


Fig. 6 Beta functions and dispersion for one quadrant of the AA

Once the lattice designer has reached this advanced stage in fixing the design it is high time that he checks with the designers of all the other components of the ring that there is enough room for them to be installed. Some of these will be best placed where the dispersion is zero. In an electron ring this is true of RF cavities and in our example it is the betatron cooling kickers that have this preference. There may be other components, momentum cooling pick-ups in the AA but momentum scrapers in synchrotrons, which need to be where the dispersion is greatest. All these must have enough free length for their function and still leave room for diagnostic pick-ups.

3. MAGNET DESIGN, MEASUREMENT AND CORRECTION

The magnets of small rings tend to be short when measured in numbers of gaps or aperture radii. The end fields of such magnets must be carefully considered because no longer will the shape of the poles in the body of the magnet alone fix the field purity. This is particularly so in the case of the AA magnets which are not only short but have huge apertures in order to accept enough beam Fig. 7. Each of the dipoles of a small ring with only a few periods will bend through an appreciable fraction of a circle. In the case of the AA the average entry angle is more than 20 degrees.

We must decide whether we should make the magnet curved so that the end faces are normal to the entering and exiting beam or whether the magnets will be just simply stacked from parallel laminations so that their ends are parallel and present, in this case, an angle of

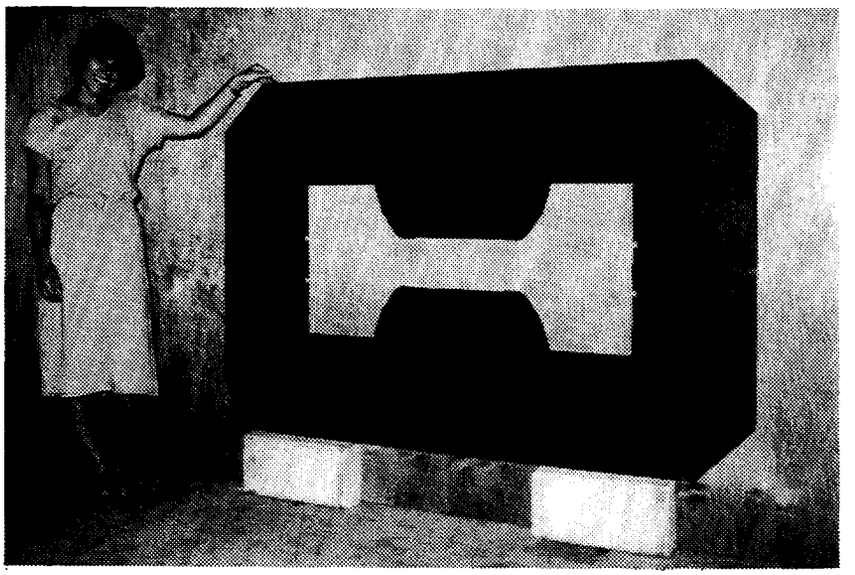


Fig. 7 One lamination of an AA dipole magnet

20 degrees to the beam. The focusing properties of the two kinds of magnet are quite different as are their transport matrices in the horizontal and vertical planes. The transport matrices for a magnet which is curved to ensure that the beam enters and leaves normally are given by:

$$M_H = \begin{pmatrix} \cos\theta & \rho\sin\theta & \rho(1-\cos\theta) \\ -\dfrac{\sin\theta}{\rho} & \cos\theta & \sin\theta \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

$$M_V = \begin{pmatrix} 1 & \rho\theta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

We can see that the (2,1) element of the horizontal matrix, which expresses the focusing power, is finite. In contrast, in the vertical plane there is no focusing action. On the other hand you can change this by tilting the end faces by angle ε as shown in Fig. 8. To first approximation, the effect of this tilt may be expressed by thin-lens matrices at each end:

$$M_H = \begin{pmatrix} 1 & 0 & 0 \\ \dfrac{\tan\varepsilon}{\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

$$M_V = \begin{pmatrix} 1 & 1 & 0 \\ \dfrac{-\tan\varepsilon}{\rho} & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

For the special case of the parallel ended magnet $\varepsilon = \theta/2$ the (2,1) elements become $\pm(1/\rho)\tan(\theta/2)$. If you multiply the matrices together you will find that the focusing power in

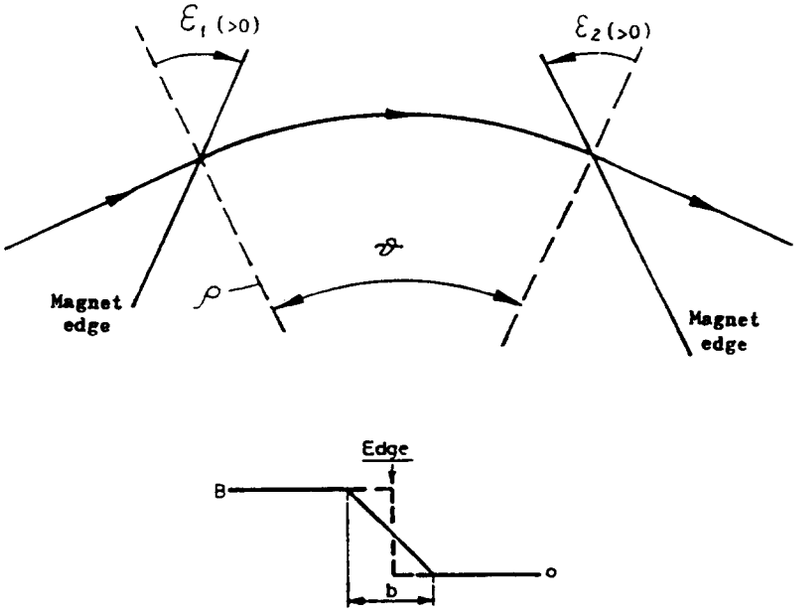


Fig. 8 Geometry of a bending magnet which is not a sector

the horizontal plane is cancelled by these thin lenses and instead, the magnet focuses vertically. Either way one cannot avoid considerable beating of the beta function in one or the other of the transverse phase planes. We see from Fig. 6 that we chose to let this happen in the vertical plane of the AA, where apertures were not so huge, rather than adopt complicated special quadrupole arrangements to match it out.

The optical properties are becoming quite precise at this stage and one should recalculate Fig. 5 to check aperture dimensions and design the vacuum chamber. We can see that some of the magnets have to house a beam 60 cm wide. This is a very extreme case which illustrates admirably some of the effects which one must be aware of in small rings even if, when calculated for smaller apertures, they prove not to be as important as in the AA.

4. MULTIPOLES IN THE MAGNET DESIGN

One of the difficulties we encounter in an analysis of beam dynamics for such a machine is that there are nonlinear terms in the focusing which have to be corrected. These are important for they modify the chromaticity or introduce a variation with momentum in the dispersion of the machine. One must compensate the chromaticity rather precisely in a storage ring and this may be done in a large ring with sets of sextupole magnets. In a small ring there is often too little space for special correction magnets but multipole correction fields may be incorporated in the magnet design. This may either be done by shaping the pole pieces or by shimming their ends with lumps of steel bolted onto the pole pieces. For example we may design the main quadrupoles with poles having a sextupole asymmetry. This causes a variation of gradient so that it is no longer the same over the whole width but has a slope. It is stronger on the outside of the machine and therefore compensates the variation of Q with momentum: the chromaticity. The corrections labelled SF1 in the plan of one quadrant of the AA, Fig. 9 are of this type. The left right symmetry of the quadrupole is broken to do this. At the same time any curvature in the variation of Q as a function of momentum may be corrected.

We can see in Fig. 10 how the integrated gradient of the quadrupole, designed with a magnet field mesh program, has a linear variation with horizontal displacement plus a quadratic term to match the chromatic curvature. The linear term is sextupolar and the curvature is octupolar (labelled OF, and OD in Fig. 9). The same figure shows correction SF2 and SD which are sextupole terms applied by end shims to the dipoles.

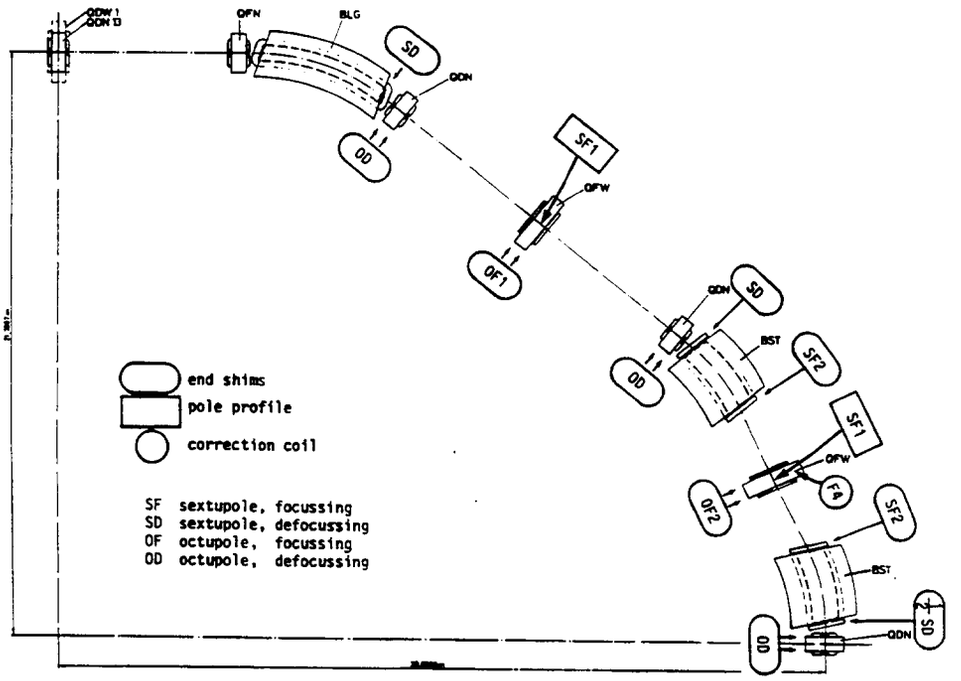


Fig. 9 Multipole corrections applied to one quadrant of the AA

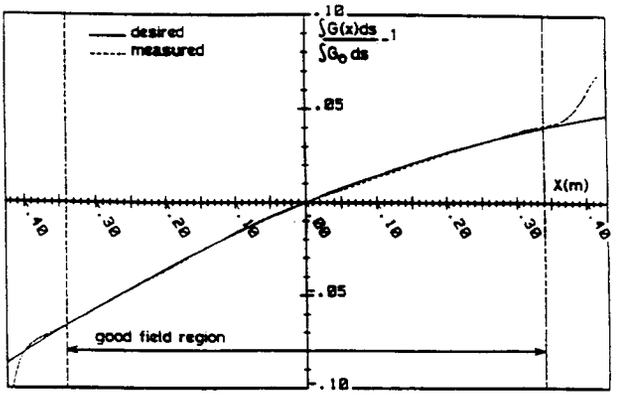


Fig. 10 Comparison between required and measured quadrupole gradient

In the AA we designed all these corrections in at the beginning but also provided the means to make adjustments to the radial field variation afterwards with packs of washers mounted on studs protruding from the pole ends. In Fig. 11 we see a wide quadrupole magnet being measured and the studs with their washers are clearly seen. The long coil integrates the gradient along a paraxial line. By controlling the number of washers on each stud we can shape the integral of the gradient as a function of horizontal position without using any power or taking up any of the circumference.

In order to calibrate the effect of washers in the 16 different stud positions seen in Fig. 12 a pack of six washers was placed on each stud in turn and the change in integrated gradient compared with an unshimmed magnet measurement [3].

The results of this calibration are shown Fig. 13 where each curve shows the change in gradient due to one pack of six on each of the washer positions. A program was written to combine combinations of washers to produce any desired change in field gradient.

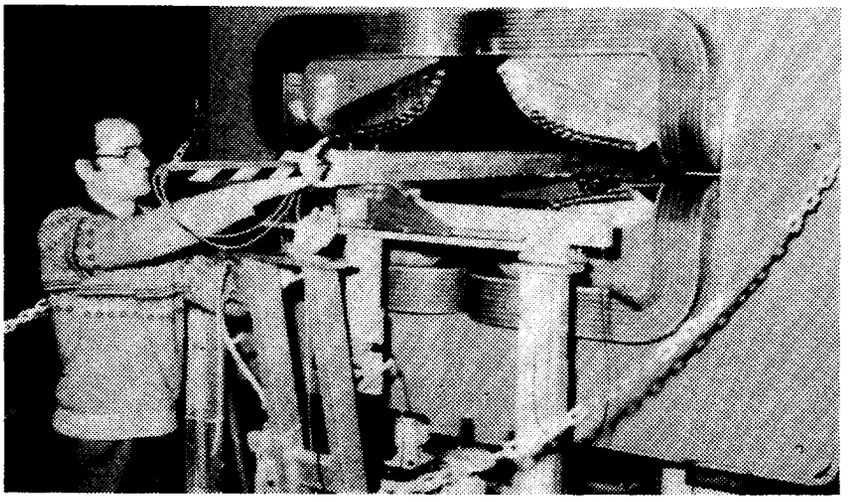


Fig. 11 Measurements being made on an AA wide quadrupole

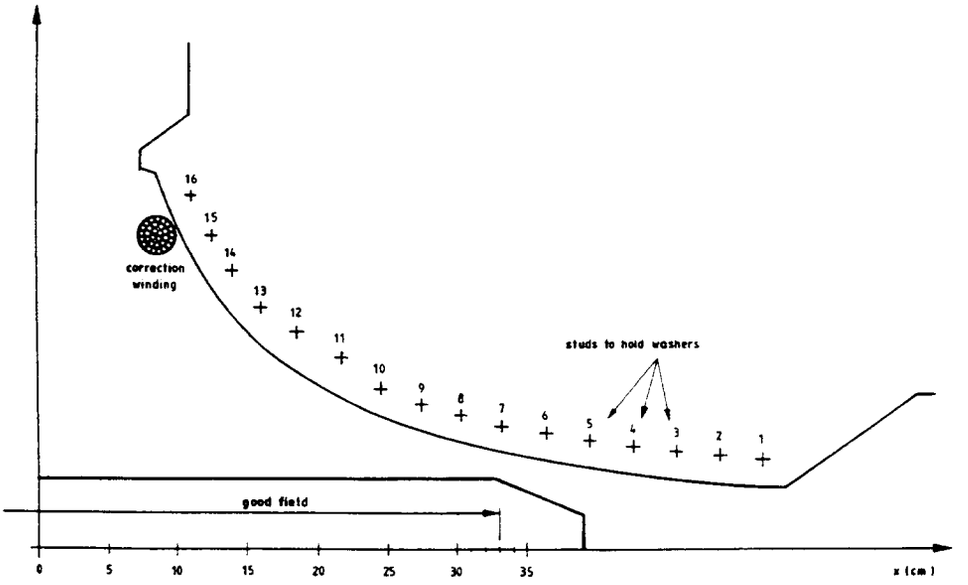


Fig. 12 Stud positions and correcting winding on an AA wide quadrupole

It proved perfectly practical to shim field shapes empirically on the basis of Q measurements with these washers but the process took several hours. To modify field shapes on-line and with the beam circulating we installed single-turn correction windings which can be seen in Fig. 12 mounted on the pole. Their effect was also calibrated by field measurement on a prototype and the results for a range of currents can be seen in Fig. 14. Of course such correction windings lead to power supply complications in a pulsed machine.

Now we come to consider how successful all this shimming was in correcting the Q variation with momentum. We see in Fig. 15 the residual variation in Q when the AA was first switched on [4]. Although every care had been taken there is still a Q variation in the horizontal plane which is larger than the space between the one-third and one-quarter integer resonances.

Adjustments to the end shims were calculated and applied to reduce this by almost an order of magnitude (Fig. 16) and when the same points are plotted on a Q_h, Q_v diagram (Fig. 17) we find that all sum resonances up to 11th order are avoided. The correction was applied in two iterations.

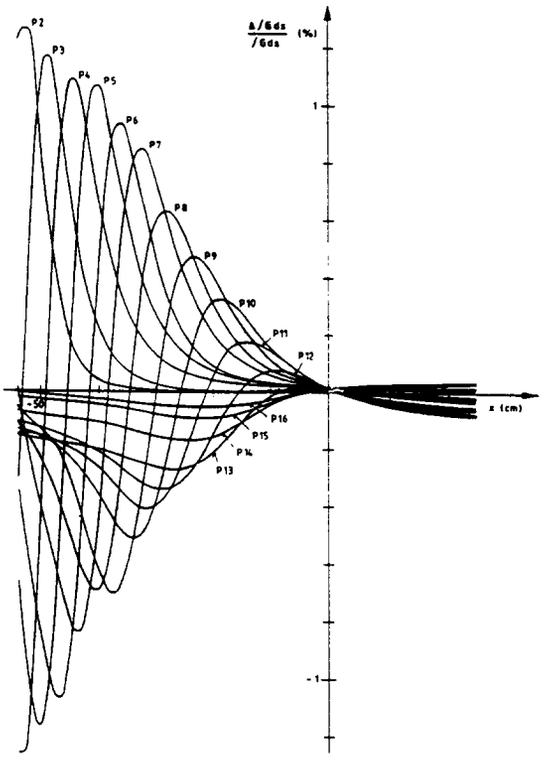


Fig. 13 The effect of a six-pack of washers on each stud position

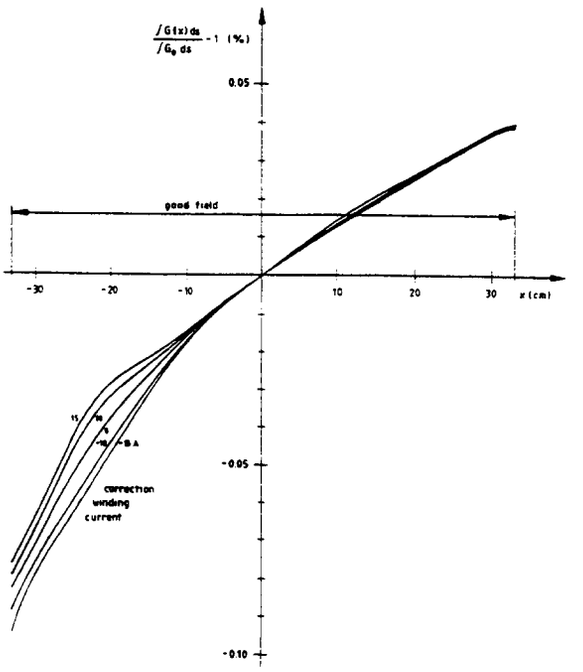


Fig. 14 Effect of current in the correction winding of an AA quadrupole

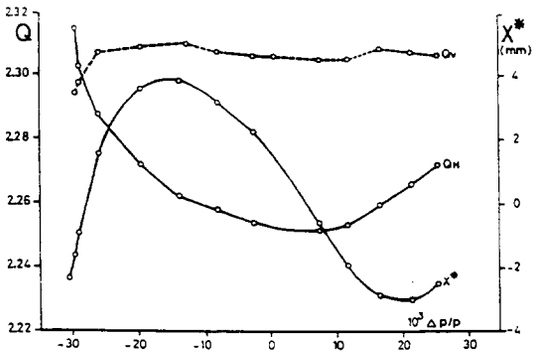


Fig. 15 Variation of Q and mid-straight dispersion before correction

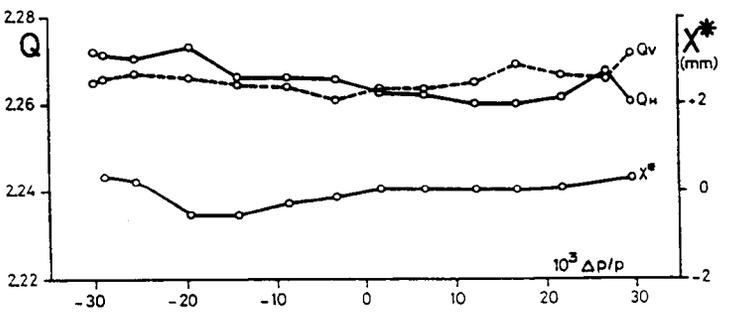


Fig. 16 Variation of Q and mid-straight dispersion after correction

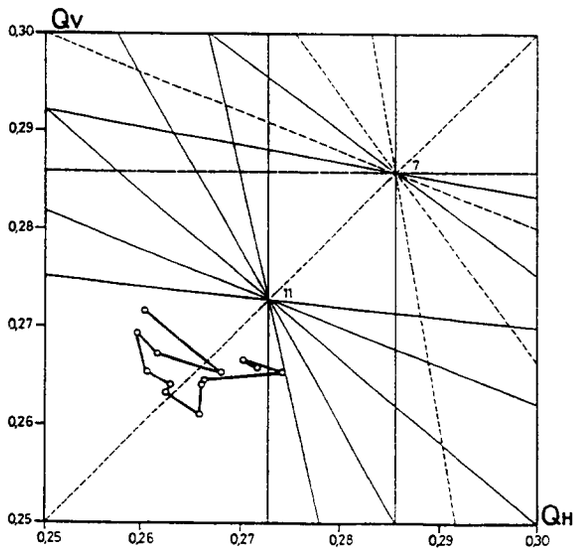


Fig. 17 The working line of the AA after correction

5. TWO-DIMENSIONAL FIELD INTEGRALS

And now we should turn to some of the theory behind the description of magnets and how this must be modified when the magnets are only a few gap-dimensions long. If a magnet's ends are identical, the field in the plane of symmetry mid-way between the ends will be "two-dimensional". Symmetry dictates that there can be no axial component, B_y. Laplace's equation for the scalar magnetic potential, ϕ:

$$\frac{\partial^2\phi(x,y,z)}{\partial x^2}+\frac{\partial^2\phi(x,y,z)}{\partial y^2}+\frac{\partial^2\phi(x,y,z)}{\partial z^2}=0 \qquad (7)$$

reduces to just the outer two terms [5]. The two-dimensional equation has the very attractive property that the solution is a harmonic series

$$\phi=\phi_n r^n \sin(n\theta)\ . \qquad (8)$$

Each term in this series corresponds to a magnet with a different number of poles. The index n of the nth term is just half the number of poles. Pure multipole would have produced the nth term in the series. It is also associated with the order of the nonlinear resonance which this multipole can produce. For example if n is 2 we obtain the field produced by a quadrupole. The potential must reverse as we describe a circle of constant radius to have two positive and two negative excursions corresponding to the four poles. Of course, in this case, the motion is linear and any resonant condition has order 2.

This is a useful simplification when we consider the dynamics of a beam passing through a long magnet where most of the field has this two-dimensional property. We can even preserve this simplification when we include the end fields of a short magnet provided the deflection in the magnet is small, as is the case in high energy synchrotrons with large radius. We can then approximate the particle's equilibrium orbit to a straight line paraxial to the centre line of the magnet. The integral of the potential along such a line, in other words the average potential, has the same two-dimensional properties. This may be proved by integrating the three terms of the Laplace equation along such a line. The middle term becomes:

$$\int_{-y_0}^{y_0} \frac{\partial^2 \phi(x,y,z)}{\partial y^2}\, dy = \left[\frac{\partial \phi(x,y,z)}{\partial y}\right]_{y_0}^{y_0} \tag{9}$$

Clearly if y_0 is taken arbitrarily far away from the ends the expression will be zero at each limit. Each of the other two terms becomes a simple average of the transverse second derivatives of potential. For example:

$$\int_{-y_0}^{y_0} \frac{\partial^2 \phi(x,y,z)}{\partial z^2}\, dy = \frac{\partial^2}{\partial z^2}\left[\frac{1}{2y_0}\int_{-y_0}^{y_0} \phi(x,y,z)dy\right] = \frac{\partial^2 \Phi}{\partial z^2} \tag{10}$$

and the two-dimensional Laplace Equation as well as its trigonometrical solutions apply exactly to the averages of the potential along paraxial lines. The average transverse fields seen by the particles are just the derivatives of the average potential and are the familiar trigonometric series with n being half the number of poles in the multipole associated with each term.

It is of course the vector potential, $\boldsymbol{A}$, which is more commonly used in the study of beam dynamics since it describes the effect of magnetic field in the Hamiltonian of the motion. In the two-dimensional case A_x and A_z are zero and the axial component obeys Laplace's equation in two dimensions producing a harmonic series solution for the field expansion.

The two-dimensional solution of A_y may also be expressed in Cartesian co-ordinates as a polynomial:

$$A_y = \sum_{1}^{\infty} A_n f_n(x,z) \tag{11}$$

where:

$$f_n(x,z) = (x+iz)^n \ . \tag{12}$$

For a quadrupole this is:

$$f_n(x,z) = \left(x^2 - z^2\right) + i(2xz) \ . \tag{13}$$

We can derive the transverse field components:

$$B_z = \frac{\partial A_y}{\partial x} = \sum_{1}^{\infty} nA_n x^{(n-1)} \tag{14}$$

when $z = 0$. We can deduce that the imaginary terms in the polynomial correspond to the skew orientation of multipoles while the real term in our example could represent a normal lattice quadrupole. A real cubic term would be a sextupole with a vertical plane of symmetry while the imaginary cubic terms stem from a skew sextupole, one that is, which has been rotated about its axis by 1/12 of a revolution.

We find, therefore that there is a simple association between the terms in the Cartesian expansion and the order of multipole, just as there was in the polar case. We can even go further to associate each term with a term in the Taylor expansion of the field about the axis of the magnet:

$$B_z = \sum_1^\infty \frac{1}{(n-1)!} \frac{dB^{(n-1)_z}}{dx^{(n-1)}} x^{n-1} \tag{15}$$

and thus a quadrupole produces a gradient, a sextupole, a second derivative and higher multipoles produce successively higher derivatives.

All of this applies equally to particles whose path may be approximated by paraxial rays and to measurements made with long paraxial search coils which extend clear of the magnetic field at each end. One can often identify the multipole content of the magnet simply by inspecting the transverse dependence of the field measured by such a coil. A quadratic variation indicates a sextupole and a cubic points to an octupole error term.

Just to complete this two-dimensional treatment, which we recall is an approximation valid only for paraxial trajectories, let us write down the simple Hamiltonian:

$$H = \frac{p_x^2}{2} + \frac{p_z^2}{2} - \left(\frac{e}{p}\right) A_s \ . \tag{16}$$

The vector potential term can, in the paraxial approximation, be written:

$$A_s \approx \sum_1^\infty A_n f_n(x,z) \tag{17}$$

and we can conveniently associate each multipole with a term in the Hamiltonian. So for our example of a quadrupole we obtain for the motion in the x direction:

$$H = \frac{p_x^2}{2} + \frac{kx^2}{2} \tag{18}$$

where k is the normalized gradient $B/'(B\rho)$. Application of Hamilton's equations leads rapidly to the familiar Hill's equation of motion. This is using a sledgehammer to crack a nut but when we come to introduce higher order terms which are nonlinear the power of the Hamiltonian produces an overall simplification.

6. CURVILINEAR COORDINATES

The above treatment is all well and good for large synchrotrons but breaks down seriously if the dipoles of the ring bend particles significantly from a paraxial path. This is much more likely to be the case in a small ring with a few focusing periods. In such a case we may consider a model which I will refer to as the "sliced-loaf" model in which the end field is broken up into elementary slices each of which may be represented as one of the multipole shaped fields of the two-dimensional model. We would expect such a model to be capable of telling us, for instance, how bad is the effect of correcting an error in the field due to the pole profile by putting shims on the end of the magnet. The two fields would not quite compensate for a particle that moved its position with respect to the magnet axis as it entered the fringe field.

To make this model work we must be very careful with the curvature of the trajectory. Bengtsson [6] has made an elegant analysis of this using the metric tensor which defines the relation between differential changes in the curvilinear system to the rectilinear Cartesian system:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -(1+hx)^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{19}$$

where h is the curvature of $1/\rho$. Using the four-vector co-ordinates of special relativity he arrives at a differential equation of motion:

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\lambda}\frac{dx^\nu}{d\tau}\cdot\frac{dx^\lambda}{d\tau} = \frac{e}{m_0c^2}F^\mu_\nu\frac{dx^\nu}{d\tau} \tag{20}$$

where F^μ_ν is the electromagnetic field tensor. This is really just an up-market version of:

$$\frac{d^2x}{dt^2} = \frac{e}{m_0}v\times B \tag{21}$$

and we can see that the middle extra term contains all the information on the transformation to curvilinear co-ordinates. He derives explicit equations of motion from this expressed in terms of the two-dimensional harmonic field coefficients. Later he derives a Hamiltonian in terms of the normalized quadrupole and sextupole strengths, k and m, and including the effect of dispersion, D, on a particle with a momentum defect $\delta = dp / p$:

$$\begin{aligned} H = &\frac{1}{2}\left\{h^2 - k - \delta\left[(m+2hk)D + h^2 - k\right]\right\}x^2 \\ &+\frac{1}{2}\left\{k + \delta\left[(m+hk)D - k + h'D'\right]\right\}z^2 \\ &-\frac{1}{6}(m+2hk)x^3 + \frac{1}{2}(m+hk)xz^2 \\ &+\frac{1}{2}(1+\delta hD)p_x^2 + \frac{1}{2}(1+\delta hD)p_z^2 \\ &+\delta hD'\,xp_x + \frac{1}{2}h'\,z^2p_x + \frac{1}{2}hxp_x^2 + \frac{1}{2}hxp_z^2 \end{aligned} \tag{22}$$

Inspecting this Hamiltonian we first should note the appearance of the focusing strength, k, alone in the first two lines. This is just like the simple two-dimensional Hamiltonian from which one may derive linear betatron motion. We can see that the focusing will be modified by the first term h^2. This term will be familiar to those who remember combined function machines. The square brackets in the first two lines contain terms which affect the off-momentum behaviour. The simple focusing terms k and h^2 appear again here to express chromaticity. The sextupole strength, m, also appears as a coefficient of δD and this must describe the effect of sextupoles in modifying the chromaticity. Terms which contain the product, hk, must be peculiar to combined function lattices where curvature and focusing occur in the same element but we should not ignore the product $h'D'$ in the second term which tells us that the shape of the curved orbit can modify vertical chromaticity.

Turning to the third line we have two terms which, in a separated function lattice in which $hk = 0$, just describe the effect of sextupoles on the betatron motion. These terms occur in our simple two-dimensional description. One can derive the strength of nonlinear resonances from

these terms. The fourth and fifth lines contain momentum dependent terms in which the curvature, h, is present and the last three terms which modify the betatron motion for on momentum particles where there is finite curvature and, presumably can contribute to nonlinear resonance width.

If we follow the analysis of the "sliced-bread" model of Bengtsson and which is also embodied in computer programs like ORBIT [7], MAD [8], we go a long way towards an exact description of the end field of magnets. The lattice functions, dispersion and chromaticity will turn out close to reality. However there remains one more effect which is less well known and which to the best of the author's knowledge is correctly embodied in only one computer program MIRKO [9]. We shall call it Electron Microscope Distortion since it is well known in that field.

7. EXCURSIONS WITHIN THE END FIELD

Suppose we return to the model of paraxial trajectories and ignore curvature for the moment. We have shown that the integrated field is two-dimensional but we know that locally this is not the case. The paraxial particle may see, say, a left-hand deflecting field from the three-dimensional nature of the field as it enters the end field region but our paraxial theorem tells us this is exactly cancelled by an equal and opposite right-hand deflection somewhere else in the end field. Now suppose the particle is following a line which is not paraxial because the beta function is varying in the end field or perhaps it is an off-momentum particle whose displacement follows the dispersion function and is not therefore paraxial. The two perturbations will now be different because the paraxial theorem relies on the particle staying the same distance from the axis through the end field and this is no longer the case. In order to understand such effects we must have a model for the end field which expresses how the multipole coefficients vary as a function of distance along the beam axis. There is a complete polynomial expression for the three-dimensional end field to be found in a text by Glaser on electron optics [10] and we reproduce this polynomial for the scalar potential as he wrote it.

$$\begin{aligned}\phi_m(x,y,z) = \Phi_m - Gx - Hy - \frac{1}{4}\left(\Phi_m'' - \Delta\right)x^2 + Qxy - \frac{1}{4}\left(\Phi_m'' + \Delta\right)y^2 \\ + \frac{1}{3}\left(\frac{1}{2}G'' + G_1\right)x^3 - H_1x^2y - G_1xy^2 + \frac{1}{3}\left(\frac{1}{2}H'' + H_1\right)y^3 \\ + \left(\frac{1}{64}\Phi_m^{(4)} - \frac{1}{48}\Delta'' + \Delta_1\right)x^4 - \left(\frac{1}{12}Q'' - 4Q_1\right)x^3y \\ + \left(\frac{1}{32}\Phi_m^{(4)} - 6\Delta_1\right)x^2y^2 - \left(\frac{1}{12}Q'' + 4Q_1\right)xy^3 \\ + \left(\frac{1}{64}\Phi^{(4)} - \frac{1}{48}\Delta'' + \Delta_1\right)y^4 \ .\end{aligned} \tag{23}$$

The coefficients G, H, Q refer to vertical and horizontal dipole and quadrupole fields respectively. Other coefficients contain the axial derivatives of these quantities. We can still identify most of the multipole coefficients although in Glaser's notation we must remember that x and y are transverse and z is axial.

In Table 1 we can see the same terms sorted according to the familiar multipoles and shown in the various columns are the expression for the fields and their gradients. Near the bottom of the table are two lines which we name "end quadratic" and "end cubic" which do not fit into our multipole description. The "end cubic" looks at first sight like an octupole but when its field gradient or focusing effect is plotted as a function of the transverse (x,y) in Fig. 18 we

see it is like a hammock while an octupole Fig. 19 is a saddle shape. Anyone who has tried to sleep in a saddle or ride in a hammock will avow to the different topology of these functions.

Table 1
Terms in Glaser's expansion according to multipole symmetry

Name	ϕ_m	B_y	B_x	dB_y / dx	dB_x / dy
Solenoid	$\phi_m - \phi'' r^2/4 + \phi^{(IV)} r^4/64$	$-\phi^{(IV)}(x^2y + y^3)/16$	$-\phi^{(IV)}(y^2x + x^3)/16$	$-\phi^{(IV)}_{xy}/8$	$-\phi^{(IV)}_{xy}/8$
H bending	$-Hy$	H	0	0	0
V bending	$-Gx$	0	0	0	0
Quadrupole	Qxy	$-Qx$	$-Qy$	$-Q$	$-Q$
Skew quadrupole	$\Delta(x^2 - y^2)/4$	$\Delta y/2$	$-\Delta x/2$	0	0
Sextupole	$H_1(y^3 - 3x^2y)/3$	$H_1(x^2 - y^2)$	$2H_1xy$	$2H_1x$	$2H_1x$
Skew sextupole	$G_1(y^3 - 3xy^2)/3$	$2G_1xy$	$G_1(y^2 - x^2)$	$2G_1y$	$2G_1y$
Octupole	$\Delta_1(x^4 - 6x^2y^2 + y^4)$	$4\Delta_1(3x^2 - y^3)$	$4\Delta_1(3xy^2 - x^3)$	$24\Delta_1xy$	$24\Delta_1xy$
Skew octupole	$4Q_1(xy^3 - x^3y)$	$4Q_1(x^3 - 3xy^2)$	$4Q_1(3x^2y - y^3)$	$12Q_1(x^2 - y^2)$	$12Q_1(x^2 - y^2)$
End quadratic	$(G''x^3 - H''y^3)/6$	$H''y^2/2$	$-G''x^2/2$	0	0
End cubic	$-Q''(x^3y + xy^3)/12$	$Q''(x^3 + 3xy^2)/12$	$Q''(3x^2y + y^2)/12$	$Q''(x^2 + y^2)/4$	$Q''(x^2 + y^2)/4$

If we look carefully at the symmetry of the end field cubic we find that, while it has the symmetry of a quadrupole in θ its radial dependence is that of an octupole. It is a characteristic of the end field of a quadrupole and proportional to the slope of the main gradient term, Q. Such terms must be included explicitly in any simulation of the end field shape although their effect in one plane can be thought of as a simple multipole.

Because Q' reverses sign as we pass through the fringe field the paraxial integral of these terms is zero but they have an effect which is proportional to the slope of the betatron function, α, and, in the case of an off-momentum particle, to the slope of the dispersion function [11].

At the time that the AA ring was designed this effect had not been discovered and it is still not incorporated in the standard lattice programs. When it was simulated with the MIRKO program, it exactly matched the curvature in Q versus momentum (Fig. 15) before empirical shim correction was applied. Simulation has also shown that it can drive fourth-order resonances. Clearly both of these effects become significant only in small rings of short magnets where the emittances and momentum spread are large.

8. CONCLUSIONS

In this review of the special features of small rings we have re-examined some of the approximations which are normally built into the analysis and the computer programs which describe transverse motion in today's large synchrotrons. The accelerator designer should be particularly careful of these approximations when the ring he is designing has magnets which bend through an angle of a few degrees or when the magnet length, measured in number of gap widths, is small.

The practical problem of having so many ends of quadrupoles, dipoles and other equipment in a small ring yet finding enough room for the components themselves is always severe but can be made less difficult if corrections are built into the magnet design and tuned by modifications to the ends of poles. One may even save the space for closed orbit dipoles by installing remotely controlled jacks to support the quadrupoles. This solution proved perfectly successful in the design of the AA and later the Antiproton Collector ring.

Finally, although there has been no time in this review to apply these lessons to small electron ring, many of the considerations are equally valid [13]. However the reader can readily appreciate that some of the correction methods may not be possible in a fast-cycling electron ring and the acceptances may not be large enough for the finer points of the end-field shape to be an important consideration. Nevertheless, the scenario for the design of the AA is a good vantage point to scan the horizon of possible pitfalls.

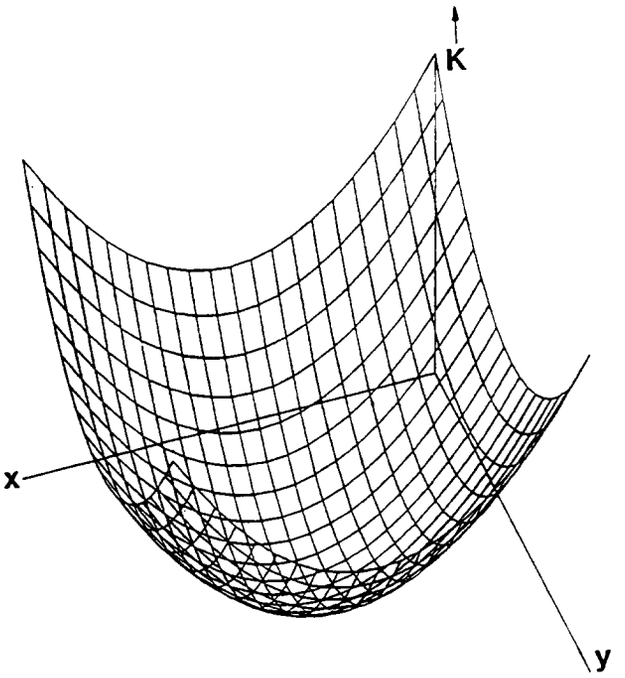


Fig. 18 End quadratic potential

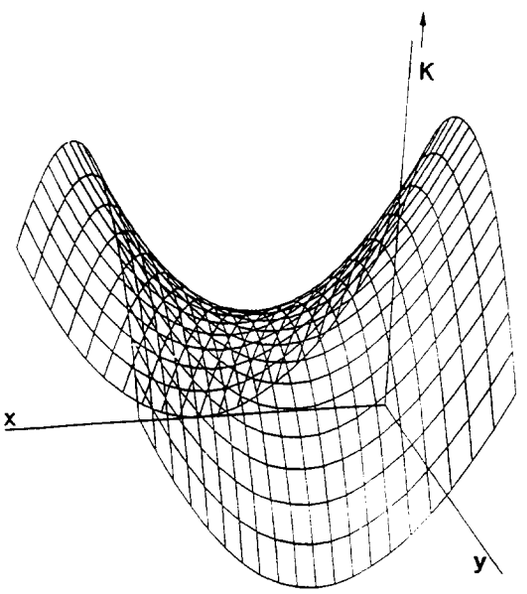


Fig. 19 Octupole potential

REFERENCES

[1] Grivet and Septier, Les Lentilles Quadrupolaire Magnetiques, CERN 58-25 (1958).

[2] B. Autin, Dispersion Suppression with Missing Magnets in a FODO Structure Application to the CERN Antiproton Accumulator, Proceedings of the 1979 Particle Accelerator Conference, San Francisco, IEEE Transactions on Nuclear Science Vol. NS-268, No. 3, pp. 3493-5 (1979).

[3] R. Brown, T. Dorenbos, M. Frauchiger, C.D. Johnson and L. Petty, An Automated Measuring System for the Antiproton Accumulator Magnets, Proceedings of the 7th International Conference on Magnet Technology, Karlsruhe, IEEE (1981).

[4] B. Autin, R. Billinge, R. Brown, G. Carron, C. Johnson, E. Jones, H. Koziol, C. Leeman, T.R. Sherwood, S. van der Meer and E.J.N. Wilson, Beam Optics Studies on the Antiproton Accumulator, Proceedings of the 1981 Particle Accelerator Conference, Washington, IEEE Transactions on Nuclear Science, Vol. NS-28, No. 3, pp, 2055-7 (1981).

[5] W.C. Elmore and M.W. Garret, Measurement of Two-Dimensional Fields, Part I: Theory Review of Scientific Instruments, Vol. 25, pp. 482-483 (1954).

[6] J. Bengtsson, Non-Linear Transverse Dynamics for Storage Rings with Applications to the Low-Energy Antiproton Ring (LEAR) at CERN, CERN 88-05 (1958).

[7] B. Autin and M. Bell, ORBIT – A computer Program for Circular Machines, Users Guide (available from authors, CERN, Geneva).

[8] F.C. Iselin and J. Niederer, The MAD Program – Users Reference Manual CERN LEP-TH 88-38 (1988).

[9] B. Franczak, MIRKO – An Interactive Program for Beam Lines and Synchrotron Computing in Accelerator Design and Operation, Lecture Notes in Physics, No. 215, p. 170, pub. Springer Verlag (1974).

[10] W. Glaser, Elektronen & hyph und Ionenoptik Handbuch der Physik, Band XXXIII, Springer Verlag (1956), p. 166.

[11] E.J.N. Wilson, Strange Multipoles in Magnet Ends, CERN PS/ACOL/Note 84-3 (1984).

[12] P. Krejcik, Non-Linear Quadrupole End-Field Effect in the CERN Antiproton Accumulator, Proceedings of the 1987 IEEE Particle Accelerator Conference, Washington, IEEE Catalog No. 87C#2387-9, pp. 1278-9.

[13] J.P. Delahaye and A. Krusche, The Lattice Design of the LEP Electron Positron Accumulator (EPA), Proceedings of the 1983 Particle Accelerator Conference, Santa Fe, IEEE Transaction on Nuclear Science, Vol. NS-30, No. 4 (1983), p. 2050.

HAMILTONIAN TREATMENT OF SYNCHROBETATRON RESONANCES

G. Guignard
CERN, Geneva, Switzerland

Abstract

Different mechanisms may couple the equations of betatron motion and synchrotron motion through the physical quantities of accelerators such as chromaticity and dispersion. They generate synchrobetatron resonances (SBR) that can be seen as sidebands of the pure betatron resonances. The importance of having convenient formulae to analyse their damaging effects, to make numerical evaluations, and to find possible corrections is emphasized. It is shown how the Hamiltonian formalism can be used to reach these goals. Direct applications to design studies or functioning accelerators are given, together with some observations.

1 INTRODUCTION

In circular accelerators and colliders, coupling between synchrotron (longitudinal) and betatron (transverse) motions gives rise to nonlinear resonances, termed synchrobetatron resonances (SBR). It generates a web of sidebands or satellites beside the more familiar betatron resonances associated with electromagnetic fields [1]. Different mechanisms may be at the origin of this coupling.

One possible source of coupling is linked to time variations of the betatron tunes, due to the simultaneous presence of momentum oscillations and finite chromaticity or tune dependence with energy. The treatment, based on unpublished work carried out by the author [2] during the design of the proton–antiproton facility in the Super Proton Synchrotron (SPS) at CERN, is given in detail for one-dimensional transverse oscillations in the presence of nonlinear electromagnetic fields, caused by magnetic multipoles or beam–beam forces. The corresponding field potential (see the Appendix) can be written, after an expansion to order N

$$H_1 = \sum_N \frac{1}{N!} K^{(N-1)} x^N \ , \tag{1}$$

with

$$K^{(N-1)} = \frac{R^2}{|B\rho|} \frac{\partial^{(N-1)} B_z}{\partial x^{(N-1)}} \qquad \text{for } 2N\text{–pole components,}$$

and

$$K^{(N-1)} = \frac{1+\beta^2}{\beta c} \frac{R^2}{|B\rho|} \frac{\partial^N \Phi}{\partial x^N} \qquad \text{for beam–beam potential } \Phi.$$

The stability of the motion is discussed analytically, and emphasis is given on how to use a few simple results for rapid numerical evaluations and practical applications.

Another source of synchrobetatron coupling finds its origin in the existence of finite dispersion at the RF accelerating cavities. The local change of energy induces a jump of

the equilibrium orbit, and consequently additional betatron excursions, while the variation of the orbit length provokes phase shifts of the beam with respect to the acceleration voltage. The summary which is presented below is based on the elegant Hamiltonian treatment of this mechanism published by T. Suzuki [3], and revised more recently by R. Baartman [4]. This treatment offers all the advantages linked to the Hamilton mechanics, i.e. the preservation of the symplectic conditions and the availability of many general results already established [1]. Analytical outcome is given in the form of a few formulae that make numerical evaluations easier, and provide useful clues on how to compensate or keep under control the effects of SBR caused by dispersion in cavities. Examples of application are given in order to illustrate the validity of the theory, and its usefulness in a design period to study the resonance effects and their compensations. Observations made in a running accelerator are also summarized.

For completeness, let us mention yet another mechanism, though not treated in this article, which is related to the longitudinal wakefields generated by a beam interacting electromagnetically with the cavities. Starting from the equations of the coupled synchrobetatron motion rather than from the Hamilton function, Ref. [5] gives analytic expressions of the vertical displacement and its derivative after n turns, for one particle in a gaussian bunch and a single RF cavity.

2 TUNE MODULATION AND SYNCHROBETATRON RESONANCES

2.1 Time varying betatron tune

In this section, we shall deal with the one-dimensional betatron motion and explain how variations of the transverse tunes related to synchrotron oscillations (see the Appendix) may induce SBR. The perturbation treatment is applied, and the motion and its stability are investigated analytically, near the resonances.

The reasons for a time variation of betatron tunes are linked to a series of factors recalled hereafter:

- As it is well known, the magnetic deflection associated with a constant external field B_0 is inversely proportional to the energy (or momentum) of the particle.
- As a direct consequence of the first factor, the focusing strength of the quadrupoles depends on the actual energy of the particles inside the beam.
- The RF acceleration system generates energy oscillations (synchrotron motion) of the particles with a wave number per turn (synchrotron tune) Q_s.
- As a result of the two preceding points, i.e. energy oscillations and focusing variation with momentum, it follows that any of the two transverse tunes Q (Q_x or Q_z) may in general oscillate at a frequency given by the synchrotron tune Q_s.
- The Q-oscillation is not necessarily linear, since nonlinear magnetic fields (such as sextupoles or higher multipoles) generate a nonlinear dependence of Q on the momentum p or momentum deviation $\Delta p/p_0$.

Figure 1 gives an illustration of an arbitrary nonlinear chromatic dependence of one transverse tune Q with the momentum deviation. The linear variation around the

origin $Q = Q_0$, and $\Delta p = 0$ corresponds to the finite slope of the curve $Q(\Delta p/p_0)$, and is characterized by the so-called chromaticity in the accelerator jargon,

$$\xi = \frac{dQ}{d(\Delta p/p_0)} = Q' \,. \tag{2}$$

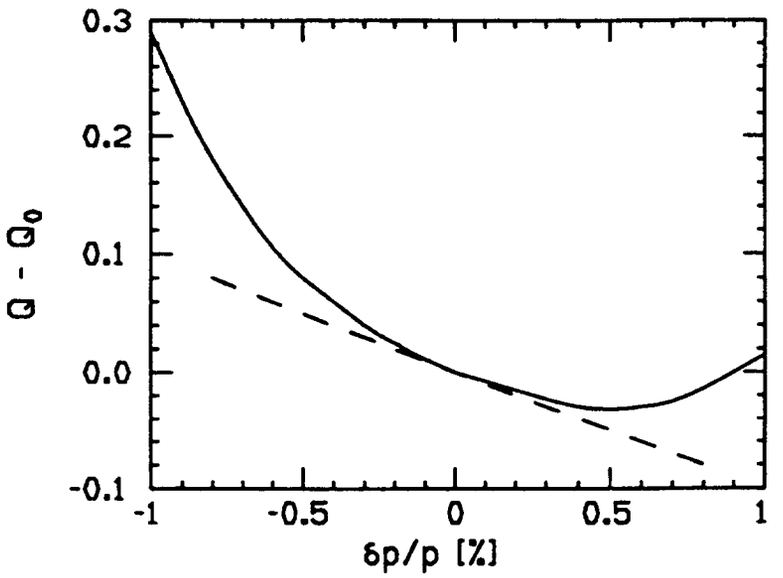


Fig. 1: Example of chromatic dependence of the tune Q and of the slope at the origin that defines the chromaticity.

If the particle momentum oscillates and $\Delta p/p_0$ varies with time, then the betatron tune Q depends on the angular variable θ via the function $Q(\Delta p/p_0)$. Consequently, one likes to replace in the Floquet function (A.10) and the Hamiltonian (A.14) of the Appendix, the product $Q\theta$ with Q constant by the integrated phase variation

$$Q\theta \rightarrow \int_0^\theta Q(\theta) d\theta \,. \tag{3}$$

With this change in the perturbing Hamiltonian U (see the Appendix), the effects of the time varying betatron tune can be studied.

2.2 Modified Hamiltonian and synchrobetatron resonances

Using the notation as in the Appendix, the Hamiltonian of a perturbation due to electromagnetic fields can be written for one degree of freedom [1]

$$U(a, \bar{a}, \theta) = \sum_N \sum_{\substack{j,k \\ j+k=N}} \sum_p h^{(N)}_{jk-p}\, a^j \bar{a}^k \exp\left[i(j-k) \int_0^\theta Q(\theta) d\theta \right] e^{-ip\theta} \,, \tag{4}$$

where a is the transverse complex amplitude of the motion and $h^{(N)}$ is by definition the Fourier component p of the perturbation of order N

$$h^{(N)}_{jk-p} = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{\sqrt{\beta_x}^{(j+k)}}{(j+k)!} K^{(N-1)}(\theta) \exp\left[i(j-k)(\mu_x - Q\theta) + ip\theta \right] \,. \tag{5}$$

For small (near linear) synchrotron oscillations around the stable fixed point of the RF bucket (Appendix), one can write after linearization

$$\frac{\Delta p}{p_0} = A_p \cos\, Q_s\theta\,. \tag{6}$$

Consequently, for $\Delta p/p_0 \ll 1$, one obtains via the chromaticity (2)

$$Q(\theta) = Q_0 + \Delta Q \cos\, Q_s\theta$$

and

$$\int_0^\theta Q(\theta)d\theta \;=\; Q_0\theta + \frac{\Delta Q}{Q_s}\sin Q_s\theta\,, \tag{7}$$

with $\Delta Q = \xi A_p$, where A_p is the peak amplitude of the momentum variation. Q_s is the synchrotron tune and Q_0 the transverse tune on momentum.

When introducing (7) into the equation (4) of U, the phase term gives

$$\begin{aligned}
&\exp\left\{i(j-k)\left[Q_0\theta + \frac{\Delta Q}{Q_s}\sin Q_s\theta\right]\right\} = \\
&= \cos\left\{(j-k)\left[Q_0\theta + \frac{\Delta Q}{Q_s}\sin Q_s\theta\right]\right\} + i\sin\left\{(j-k)\left[Q_0\theta + \frac{\Delta Q}{Q_s}\sin Q_s\theta\right]\right\}\,.
\end{aligned} \tag{8}$$

Replacing $(j-k)$ by n for simplification and developing these trigonometric functions provides the following relations

$$\begin{aligned}
&\cos\left\{\underline{(j-k)}\left[Q_0\theta + \frac{\Delta Q}{Q_s}\sin\, Q_s\theta\right]\right\} = \cos\,\left\{\underline{n}\left[Q_0\theta + \frac{\Delta Q}{Q_s}\sin\, Q_s\theta\right]\right\} = \\
&= \cos\, nQ_0\theta\,\underline{\cos\left(\frac{n\Delta Q}{Q_s}\sin\, Q_s\theta\right)} - \sin\, nQ_0\theta\,\underline{\sin\left(\frac{n\Delta Q}{Q_s}\sin\, Q_s\theta\right)}\,, \\
&\sin\left\{n\left[Q_0\theta + \frac{\Delta Q}{Q_s}\sin\, Q_s\theta\right]\right\} = \\
&= \sin\, nQ_0\theta\,\underline{\cos\left(\frac{n\Delta Q}{Q_s}\sin\, Q_s\theta\right)} + \cos\, nQ_0\theta\,\underline{\sin\left(\frac{n\Delta Q}{Q_s}\sin\, Q_s\theta\right)}\,,
\end{aligned} \tag{9}$$

where the arguments of the underlined trigonometric functions are periodic functions with a period $2\pi/Q_s$. Auspiciously, trigonometric functions of sines and cosines have closed-form Fourier's expansions involving Bessel's functions of the first kind [6]

$$\begin{aligned}
\cos\left(\frac{n\Delta Q}{Q_s}\sin\, Q_s\theta\right) &= J_0\left(\frac{n\Delta Q}{Q_s}\right) + 2\sum_{\ell=1}^{\infty} J_{2\ell}\left(\frac{n\Delta Q}{Q_s}\right)\cos\, 2\ell Q_s\theta \\
\sin\left(\frac{n\Delta Q}{Q_s}\sin\, Q_s\theta\right) &= 2\sum_{\ell=0}^{\infty} J_{2\ell+1}\left(\frac{n\Delta Q}{Q_s}\right)\sin\,(2\ell+1)Q_s\theta\,.
\end{aligned} \tag{10}$$

The use of Eqs. (10) and of the basic functional relations for the sine and cosine products that figure in the expressions (9) gives the entire phase term of the Hamiltonian U (4)

$$\exp[i\,n\int_0^\theta Q(\theta)d\theta] = J_0\,\exp[i\,nQ_0\theta] \;+\; \sum_{\substack{\ell=-\infty\\ \ell\neq 0}}^{+\infty} J_{2\ell}\,\exp[i(nQ_0+2\ell Q_s)\theta] + \\ + \sum_{\ell=-\infty}^{+\infty} J_{2\ell+1}\,\exp\{i[nQ_0+(2\ell+1)Q_s]\theta\}\;, \quad (11)$$

where the argument of the Bessel functions is $n\Delta Q/Q_s$.

According to the principles recalled in the Appendix, the new Hamiltonian of the perturbation can be written as follows,

$$U = \sum_N \sum_{j,k,-p} h^{(N)}_{jk-p}\,a^j\bar{a}^k \sum_{q=-\infty}^{\infty} J_q\Big[(j-k)\frac{\Delta Q}{Q_s}\Big]\exp\{i[(j-k)Q_0+qQ_s-p]\theta\}\;, \quad (12)$$

in which the index q replaces the indices 2ℓ or $(2\ell+1)$ of Eq. (11).

In order to push further the analysis of the SBR due to a time varying tune, it is necessary to introduce here a basic assumption: the low frequency part of the function U gives the important variations of the amplitude a of the perturbed motion. This requires extracting from U the pure low-frequency terms obtained by equalling to zero the square-bracket of the phase term in (12); there are two possible ways of satisfying this condition:

1) $j-k=0$ and $q=p=0$, simultaneously.

   The corresponding terms, sometimes called stabilizing coefficients, are the frequency shifts with amplitude that arise in the presence of nonlinear forces.

2) $j-k=N$, and the following relation holds

$$NQ_0+qQ_s-p=0\;. \quad (13)$$

   Since $j+k=N$ by construction, this implies together with $j-k=N$ that $j=N$ and $k=0$. The relation (13) is the characteristic of a so-called synchrobetatron resonance of order N, while the integer q is the side-band number, and p the harmonic number of the driving force.

   When wishing to study the motion near one of these resonances of order N (at a side-band q), one has to consider the following 'single-resonance' Hamiltonian,

$$U = \sum_\nu h^{(2\nu)}_{\nu\nu 0}\,(a\bar{a})^\nu + \kappa\,a^N\,\cos\,\psi_1\;. \quad (14)$$

   which can be rewritten with different variables, as defined below

$$U = \sum_\nu h^{(2\nu)}_{\nu\nu 0}\,r^{2\nu} + 2|\kappa|\,r^N\,\cos\,\psi\;. \quad (15)$$

where,

$h^{(2\nu)}_{\nu\nu 0}$ are the stabilizing coefficients with $j = k = \nu$ associated with $2 \times 2\nu$-pole magnetic components,

$\kappa = J_q[N(\Delta Q/Q_s)]h^{(N)}_{N0-p}$ is the excitation coefficient or driving term of the resonance [with $h^{(N)}$ defined by Eq. (5)],

and the complex amplitude a of the betatron motion has been replaced by a real amplitude, and a phase φ according to

$$\begin{cases} a = r\, e^{i\varphi} & r, \varphi \in \mathbb{R} \\ \bar{a} = r\, e^{-i\varphi} & r^2 = a\bar{a} = R\epsilon/2\pi \end{cases} , \tag{16}$$

if ϵ is the emittance associated with a particle trajectory.

The phase ψ_1 gives the 'distance' from the SBR at given tunes and harmonic number, and ψ includes in addition the phase φ of the motion,

$$\begin{aligned} \psi_1 &= (NQ_0 + qQ_s - p)\theta \\ \psi &= \psi_1 + N\varphi\,. \end{aligned} \tag{17}$$

The Hamiltonian U (14) in the presence of the perturbation, i.e. the time variation of the betatron tune due to synchrotron oscillations, is the starting point to study separately each SBR defined by Eq. (13).

2.3 Analysis of the motion near resonance

As explained in the Appendix, the canonical equations associated with U (15) give the equations of the perturbed motion,

$$\begin{aligned} \frac{dr^2}{d\theta} &= -\frac{\partial U}{\partial \varphi} = 2|\kappa| N r^N \sin\,\psi \\ \frac{d\varphi}{d\theta} &= \frac{\partial U}{\partial r^2} = \sum_{\nu=1}^{\infty} h^{(2\nu)}_{\nu\nu 0}\, \nu\, r^{2(\nu-1)} + N|\kappa| r^{(N-2)} \cos\,\psi\,, \end{aligned} \tag{18}$$

and the first relation can be simplified to

$$\frac{dr}{d\theta} = N|\kappa| r^{(N-1)} \sin\psi\,. \tag{19}$$

At the resonance, where the 'distance' (13) vanishes and $\psi_1 = 0$, the equations of the motion become autonomous, i.e. independent of θ. Many properties of autonomous equations can be derived from the fixed points (in the phase space) that are defined by

$$\frac{dr}{d\theta}\,(r_0, \varphi_0) = \frac{d\varphi}{d\theta}\,(r_0, \varphi_0) = 0\,, \tag{20}$$

where (r_0, φ_0) are the coordinates of these points. For a SBR of order N, there are different possible solutions to the conditions (20).

1) The trivial solution, $r = 0$ (any φ), corresponds to the origin of the phase space.

2) Non-trivial solutions result from the first condition (20) $r' = 0$, and the expression (19) taken at $\psi_1 = 0$,

$$N|\kappa|r^{(N-1)} \sin N\varphi = 0 \,. \tag{21}$$

They correspond to particular phases given by

$$N\varphi_0 = m\pi \quad \text{or} \quad \varphi_0 = \frac{m\pi}{N} \quad m = 0, 1, 2, 3, ..., (2N-1) \,, \tag{22}$$

which define $2N$ directions in the phase space. Then, the second condition (20) that $\varphi' = 0$ gives the following relation

$$\sum_{\nu} h^{(2\nu)}_{\nu\nu 0}\, \nu\, r^{2(\nu-1)} + N\,|\kappa|\, r^{(N-2)}(-1)^m = 0 \,, \tag{23}$$

in which $\cos\psi = \cos N\varphi$ has been replaced by $(-1)^m$ because of the particular values taken by φ_0 (22). Equation (23) is a polynomial in r that provides the amplitudes r_0 of the $2N$ fixed points (r_0, φ_0) as solutions. In the phase space, the normalized coordinates of the fixed points are simply given by

$$\begin{aligned} x^{(n)} &= r_0 \cos\varphi_0 \\ p_x^{(n)} &= r_0 \sin\varphi_0 \,. \end{aligned} \tag{24}$$

Figure 2 gives an illustration of how the phase space and the invariants of the betatron motion near one single resonance defined by $N = 4$ and $q = 0$ look. It is possible to recognize the eight privileged directions corresponding to the eight non-trivial fixed points expected from the analysis made above. The trivial solution is of course visible at the origin, around which the motion is locally stable.

Looking at the non-trivial fixed points (with finite r_0), it is fairly obvious that there are two different kinds (Fig. 2):

- There are Stable Fixed Points at phases φ_0 equal to $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$, with elliptic motion around them. Particles sweep closed invariants corresponding to oscillation around the point (r_0, φ_0).

- There are however Unstable Fixed Points at phases φ_0 equal to 0, $\pi/2$, π, and $3\pi/2$, with hyperbolic motion going away from (r_0, φ_0) to possibly large amplitudes. Particles moving along these invariants may either circle the whole stable island area (as in Fig. 2) or diverge to infinity, depending on the initial conditions and the perturbation strength.

This duality of the fixed points can be established analytically from the linearization of the motion around (r_0, φ_0), assuming

$$\begin{aligned} r &= r_0 + \Delta r \\ \varphi &= \varphi_0 + \delta\varphi \,. \end{aligned} \tag{25}$$

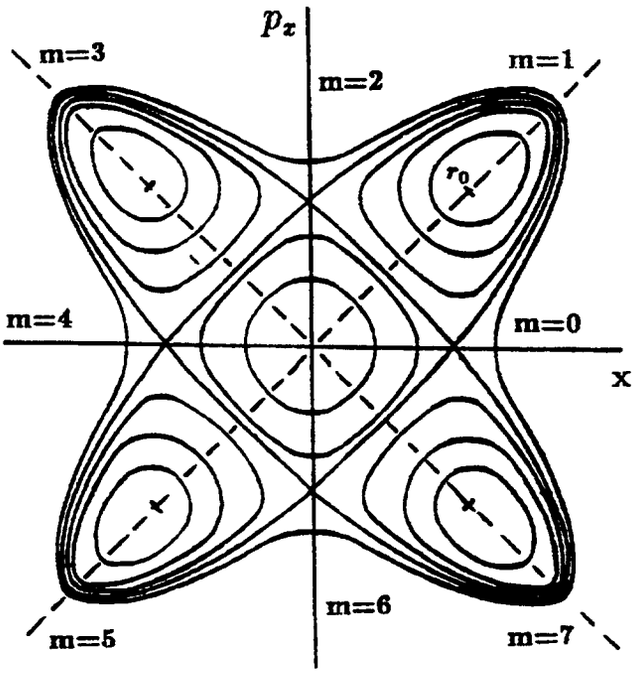


Fig. 2: Invariants in the betatron phase space near a fourth-order resonance, with the corresponding fixed points.

Developing the equations of motion (18) and (19) around one arbitrary fixed point, using (25), and keeping the terms of first order in Δr and $\Delta\varphi$, we can derive new equations valid for small-amplitude motion near (r_0, φ_0)

$$\begin{aligned} \Delta r' &= N^2\,|\kappa|\,r_0^{(N-1)}(-1)^m\Delta\varphi \\ \Delta\varphi' &= \left[\sum_{\nu=2}^{\infty} 2\nu(\nu-1)h_{\nu\nu 0}^{(2\nu)}\,r_0^{(2\nu-3)} + N(N-2)|\kappa|\,r_0^{(N-3)}(-1)^m\right]\Delta r\,, \end{aligned} \tag{26}$$

with $\varphi_0 = m\pi/N$, and therefore $\cos N\varphi_0 = (-1)^m$. In order to make the discussion about the stability of the fixed point easier, it is helpful to assume that among all the stabilizing coefficients $h^{(2\nu)}$, the one which dominates is of lowest order $\nu = 2$ (i.e. the octupole term), and that the driving term κ of the resonance is small such as $\kappa^2 \ll 1$. With these assumptions, the second equations (26) can be simplified by neglecting the term with κ and keeping only the term $\nu = 2$ of the sum, and the two first-order equations (26) can then be combined in a single second-order differential equation as follows,

$$\Delta\varphi'' = \left[4N^2h_{220}^{(4)}|\kappa|\,r_0^N\right](-1)^m\Delta\varphi\,. \tag{27}$$

This is the key equation describing small excursions around a fixed point, now characterized by its amplitude r_0 and its 'direction' given by m [Eq. (22)]. The condition of stability for these small motions is obviously

$$\begin{aligned} & C_f(-1)^m < 0 \\ \text{with} \quad & C_f = 4N^2h_{220}^{(4)}|\kappa|\,r_0^N\,. \end{aligned} \tag{28}$$

Hence, the kind of fixed points observed depends on the sign of $h^{(4)}_{220}$ (octupole coefficient):

$$\begin{aligned} h^{(4)}_{220} > 0 \quad & \text{Stable Fixed Points for} \quad m = 1, 3, 5, 7, \ldots \\ & \text{Unstable Fixed Points for} \quad m = 0, 2, 4, 6, \ldots \\ h^{(4)}_{220} < 0 \quad & \text{Stable Fixed Points for} \quad m = 0, 2, 4, 6, \ldots \\ & \text{Unstable Fixed Points for} \quad m = 1, 3, 5, 7, \ldots \end{aligned}$$

For stable fixed points, the small excursions given by (27) are periodic oscillations of frequency

$$Q_f = 2N\sqrt{|h^{(4)}_{220}|\,|\kappa| r_0^N}\,. \tag{29}$$

The four stability islands of Fig. 2 (for $N = 4$) are delimited by separatrices passing by and crossing at the four unstable fixed points (m even). The corresponding phase space configuration can therefore be represented schematically by these separatrices that are defined as the outermost invariants touching the nearby unstable fixed points. They can be determined numerically in specific cases by searching for the extreme 'closed trajectory' around each stable fixed point: as a consequence, the picture in the phase space will be reduced to four (or N) island contours defined with a precision that depends on the accuracy of the numerical calculations. Such families of N unstable fixed points, and N elliptic invariants (island contours) around the stable fixed points exist, and can be drawn for each side-band q ($q \geq 1$) of a synchrobetatron resonance of order N. The width of the separatrix or of the island contour and the frequency spacing between different SBR side-bands (according to the expressions (17) of ψ and ψ_1) are given respectively by

$$\begin{aligned} W_f &= 4Q_f \\ S_f &= \frac{Q_s}{N}\,. \end{aligned} \tag{30}$$

Consequently, the elliptic islands of the sidebands of one SBR of particular order N will overlap when $W_f > S_f$, and be separated when $W_f < S_f$. Taking into account the presence of stochastic layers of finite width along the separatrices, the elliptic islands do not overlap, if and only if

$$W_f < \frac{2}{\pi} S_f \qquad \rightarrow \qquad Q_f < \frac{1}{2\pi}\,\frac{Q_s}{N}\,. \tag{31}$$

This condition of stability is known under the name of Chirikov's criterion, after its author [7]. In the particular case of a time-varying betatron tune and a dominant stabilizing octupole, the criterion of stability (31) can be rewritten by using both Eq. (29) and the definition of κ following Eq. (15),

$$4\pi N^2 \sqrt{|h^{(4)}_{220}| J_q\left(\frac{N\Delta Q}{Q_s}\right) h^{(N)}_{N0-p}\, r_0^N} < Q_s \tag{32}$$

with

$$h^{(N)}_{N0-p} = \left(\frac{2\pi}{R}\right)^{N/2-1} \frac{\Delta e}{2N^2 \epsilon_b^{(N/2-1)}}\,. \tag{33}$$

Equation (33) is a reminder of the relation between $h^{(N)}$ and the total bandwidth Δe of Ref. [1], ϵ_b being the transverse beam emittance [see Eq. (16)].

The passage from a stable situation to an unstable situation in the presence of a time-varying tune developing synchrobetatron sidebands is illustrated for $N = 4$ in three successive figures (Figs. 3, 4, 5) resulting from numerical solutions of the equations (18) for φ, and (19) for r, with increasing resonance strength κ (in arbitrary units);

i) Figure 3 describes a stable case. The four island contours are drawn successively from inside to outside for every sideband between $q = 1$ (mostly inside) and $q = 6$ (mostly outside). With $\kappa = 0.05$, the resonance strength is weak enough for the sideband islands to remain separated; the particles trapped in these islands stay there, and their motion is therefore stable.

ii) Figure 4 shows a case where the excursion of the particles with initially small amplitudes may increase. The same contours are drawn for sideband islands from $q = 1$ to $q = 6$, but now with $\kappa = 0.20$; the first four families of islands overlap while the following ones are still separated. Particles with small amplitudes may travel from island to island and far enough to reach the largest radius of the separatrix of the $q = 4$ sideband. The particles remain confined within this finite amplitude, but there is a possible amplitude growth for particles near the centre of the beam.

iii) Figure 5 illustrates a case of instability for $\kappa = 0.75$. The island contours are now drawn for the sidebands $q = 1$ to $q = 5$, and then all the islands overlap. Furthermore, islands corresponding to higher q values would also overlap with the ones which are represented. This means that particles near the phase space origin can wander from island to island (sideband q to sideband $q + 1$), and diffuse to infinite amplitudes. Therefore, the betatron motion is unstable and particles may be lost in the direction indicated in Fig. 5.

The fact that the separatrices of each sideband of Figs. 3, 4, and 5 do not exactly join on the main horizontal and vertical axis as they should is simply due to the accuracy and step size of the numerical search for the outermost invariants.

The analytical treatment described above, and applied to the SBR mechanism illustrated in Figs. 3 to 5, provides the approximate but convenient formulae (32) and (33), which can be used to establish the stability or instability of the motion under given conditions, and in the presence of a time-varying betatron tune. In this paper, this time variation is supposed to come from the momentum oscillations caused by synchrotron motion in conjunction with finite chromaticity. The same analysis is however directly valid for other sources of tune modulation such as power supply ripple, provided the synchrotron tune Q_s is replaced, from equation (6) onwards, by the wave number Q_m of the modulation. The term 'synchrobetatron resonances' should then be replaced by something like 'modulation betatron resonances', involving also sidebands and possible instabilities.

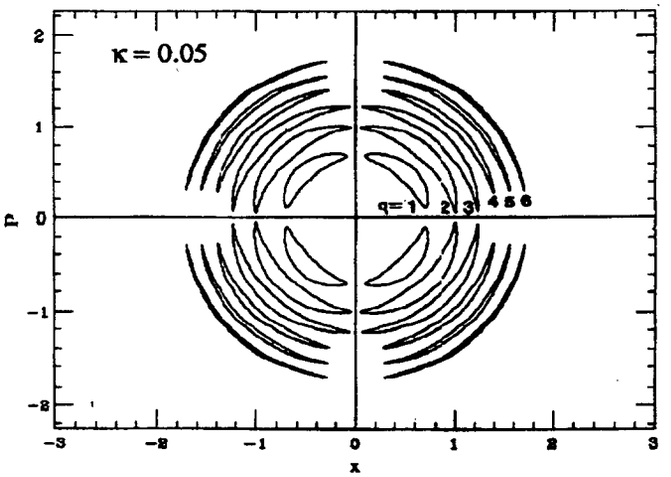


Fig. 3: Typical island contours of six synchrotron sidebands of a fourth-order resonance, in a stable case.

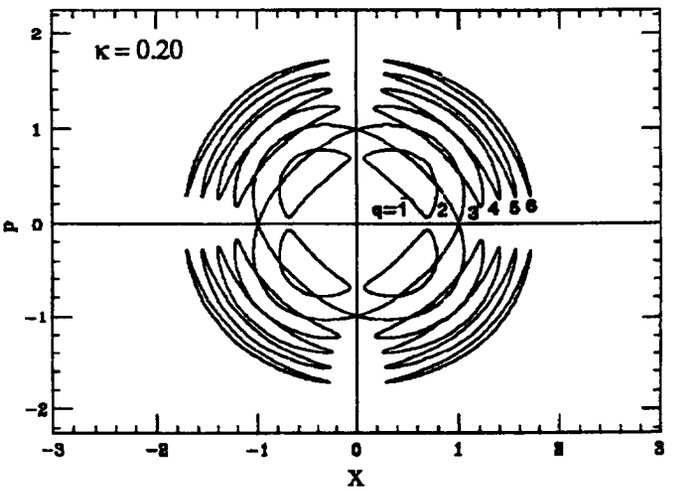


Fig. 4: Typical island contours of six synchrotron sidebands of a fourth-order resonance, with limited amplitude growth.

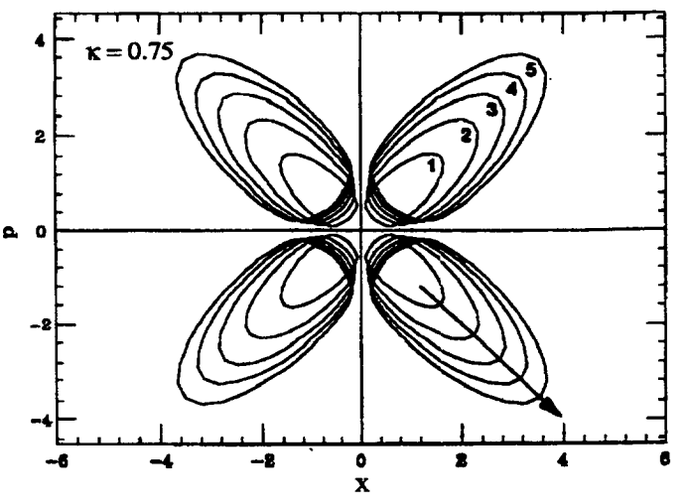


Fig. 5: Typical island contours of five synchrotron sidebands of a fourth-order resonance, in an unstable case.

2.4 Practical application to accelerators

During the design study of the proton–antiproton colliding beam facility based on modifications and additions to the Super Proton Synchrotron (SPS) at CERN, the theoretical model presented above for SBR was developed [2] with a view to analysing some of the effects caused by beam–beam interactions of colliding bunches. The design [8] was based on a $p\bar{p}$ colliding facility with 6 bunches per beam, two low-beta insertions, and for a flat-top energy of 270 GeV. The bunch length was taken as 0.63 m, the emittances were 0.08 and 0.04 μrad m in the two planes, respectively, the assumed chromaticity ξ was ~ 0.1, the energy spread amounted to 0.7‰, and the fractional parts of the betatron tunes were chosen near 0.88 for the study.

In this case, the dominant forces driving resonances are the beam–beam forces which were estimated first. In other words, the excitation coefficients $h^{(N)}_{jk-p}$ (5) and the total bandwidths Δe of resonances of order 2 to 10 have been calculated for beam intensities corresponding to a beam–beam tune shift of 0.003 per crossing, and for beam off-centering of ~ 0.035 mm. Some of the results are briefly summarized in Table 1.

Table 1
Bandwidth of Sp$\bar{p}$S beam–beam resonances

Order N	Range of Δe values		
5	$0.17 \cdot 10^{-3}$	to	$0.17 \cdot 10^{-2}$
6	$0.90 \cdot 10^{-4}$	to	$1.08 \cdot 10^{-3}$
7	$0.50 \cdot 10^{-4}$	to	$0.50 \cdot 10^{-3}$
8	$0.40 \cdot 10^{-5}$	to	$2.00 \cdot 10^{-4}$
9	$1.10 \cdot 10^{-6}$	to	$2.50 \cdot 10^{-5}$
10	$0.50 \cdot 10^{-6}$	to	$7.20 \cdot 10^{-6}$

The next step consisted in looking at the SBR effects and the stability criterion (32) in order to decide which resonances can effectively be tolerated within the beam. The detuning or self-stabilizing effect of the strong octupole component of the beam–beam field has to be taken into account; the corresponding stabilizing coefficient is proportional to the linear beam–beam tune shift ΔQ_{bb} [2],

$$h^{(4)}_{220} \cong \frac{3\pi}{4} \frac{\Delta Q_{bb}}{R\epsilon_b} , \tag{34}$$

using the notation of the previous section, the coefficient of proportionality depending on the beam aspect ratio (close to 1 in the present case). Expressing the criterion of stability at an amplitude r_0 given by the beam emittance ϵ_b, Eq. (32) can be rewritten as follows

$$\pi N \sqrt{3\Delta Q_{bb}\, \Delta e\, J_q \left(\frac{N\Delta Q}{Q_s} \right)} < Q_s , \tag{35}$$

for dominant beam–beam forces characterized by ΔQ_{bb}, and the bandwidths Δe (Table 1). Using $\Delta Q_{bb} = 0.02$, the Δe-values of Table 1, an amplitude modulation ΔQ of 0.002, and a synchrotron tune Q_s of 0.005, the left term of the relation (35) has been evaluated for each order N of the beam–beam perturbation. Testing on the one hand the corresponding criterion (35) for each synchrobetatron satellite q of the N-order resonance, it is possible

to determine the number n_s of satellite islands that overlap, while for higher q-values they remain separated (as in Fig. 4). On the other hand, superposing the phase-space circle corresponding to the emittance of a gaussian beam onto an island pattern similar to the one in Fig. 4, the calculation of the number n_t of satellite islands that lie inside the r.m.s. beam emittance can be carried out. Comparing these two numbers gives an indication of the probability of a steady beam growth (at a resonance N) as particles migrate from one side band to another. Both n_s and n_t are given in Table 2 for several resonances N in the case of the $p\bar{p}$ collider design at CERN [8] (with the quoted assumptions).

The results in Table 2 show that resonances of order greater than $N = 10$ produce no growth, since no satellites overlap. A small amplitude growth detunes the particles from their resonance condition (because of $\Delta Q_{bb} \neq 0$). For resonances from 10th order down, the beam–beam tune spread is equal or greater than the satellite spacing, and several islands (of low q-values) overlap within the emittance limits. For orders N below 7, even all the satellites lying inside the emittance overlap. For all these resonances with N below 10, blow-ups can be expected, and growth rates can be estimated from the resonance multicrossing theory [1]. Consequently, at the time of this study it was recommended to avoid that particles cross resonances of order lower than and equal to 10, by choosing judiciously the betatron tunes of beams which collide (e.g. below an integer where resonances are sparse).

Table 2
Number of resonances that overlap, in the Sp$\bar{p}$S.

Order N	n_s (see text)	n_t (see text)
5	10	10
6	12	12
7	10	14
8	12	16
9	10	18
10	8	20
11	0	22

This conclusion was found to be in remarkably good agreement with the observations of beam–beam effects subsequently made when the Sp$\bar{p}$S hadron collider was running. The antiproton beam lifetime was indeed seen to critically depend on the working point (set of betatron tunes); when the beam–beam tune shift (spread) induced the working tune diamond to touch the 10th order resonance, the lifetime fell dramatically by a factor of approximately 20 with three bunches per beam, and even more with six bunches per beam [9]. Decreasing the sensitivity of the emittance growth to the tune adjustment while running with six bunches per beam required beam separation at the unwanted crossing points, in order to reduce the total beam–beam tune shift (spread), and keep the beam away from all resonances of order lower than 13 (which comes directly from the fact that, in the working region of the tune diagram, the higher order mode next to the 10th order one is of order 13). This experience gives a positive indication about the usefulness in the design phase of the analytical model described above and suggests that, during the study of future circular colliders such as the Large Hadron Collider (LHC) at CERN, similar analyses of SBR and/or modulation betatron resonances could provide interesting results.

3 DISPERSION IN RF CAVITIES AND SYNCHROBETATRON RESONANCES

3.1 Hamiltonian formalism in the presence of dispersion

As mentioned in the introduction, this is a summary of fairly recent theories [3, 4] on SBR driven by dispersion at RF cavity locations. Since they have mostly been published in journals, the following reminder will be brief with some emphasis on the key points. We will begin directly with the expression of the Hamiltonian, for the analysis referred to is based on this particular formalism. As it is well known, and as it can also be retrieved from the equations of motion (A.8) and (A.17) (betatron and synchrotron motion, respectively) given in the Appendix, the pertinent Hamiltonian for the study of synchrobetatron coupling is as follows:

$$\begin{aligned} H &= \frac{Rp_0}{2}\left(Kx^2+\frac{p_x^2}{p_{x_0}^2}\right)-\frac{1}{2}\frac{h^2}{Rp_0}\left(\frac{D_x}{\rho}-\frac{1}{\gamma^2}\right)W^2 \\ &\quad -\frac{R}{h\beta c}\sum_i V_i e\delta(\theta-\theta_i)(\cos\psi+\psi\sin\varphi_s)\ , \end{aligned} \tag{36}$$

in which one has to replace the η of Eq. (A.17) by the actual function $-(D/\rho-1/\gamma^2)$, since η is the one turn average of it, and the RF voltage $\hat{V}$ by $2\pi\sum_i V_i\delta(\theta-\theta_i)$, since in an accelerator there are a certain number of discrete cavities at fixed angular coordinates θ_i (or positions s_i). The definitions of the quantities entering Eq. (36) are given in the Appendix, but it is useful to recall here the meaning of W (different from the usual definition)

$$W=-\frac{\Delta E}{\omega_{rf}} \quad \text{with} \quad \Delta E=E-E_s\ . \tag{37}$$

The phase ψ is equal to the RF phase φ in the absence of the perturbation owing to finite horizontal dispersion, but is modified when $D_x \neq 0$ at cavities because of path length differences and consequent phase variations

$$\psi=\varphi-\frac{h}{R}\left(D_x\frac{p_x}{p_0}-D_x'x\right)\ . \tag{38}$$

The mechanism of the SBR clarifies the form (38) of ψ and is of the twofold type:

i) The acceleration Δp does not change the instantaneous position and angle of the particle, but the equilibrium orbit, with respect to which the betatron oscillation takes place, suddenly changes with Δp, and an additional betatron excursion is excited.

ii) The betatron oscillations lead to a change of the orbit length per revolution and thus to a variation of the beam position with respect to the phase of the acceleration voltage. Hence, betatron excursions affect the synchrotron motion.

This double mechanism can be treated as a perturbation, following the symplectic theory recalled in the Appendix. The Hamiltonian H (36) can indeed be written in the required form $H_0 + H_1$ (A.1),

$$H_0 = \frac{Rp_0}{2}\left(Kx^2 + \frac{p_x^2}{p_0^2}\right) - \frac{1}{2}\frac{h^2}{Rp_0}\left(\frac{D_x}{\rho} - \frac{1}{\gamma^2}\right)W^2 - \frac{R}{h\beta c}(\cos\varphi + \varphi\sin\varphi_s)\sum_i eV_i\delta(\theta - \theta_i)$$

$$H_1 = -\frac{\sin\varphi - \sin\varphi_s}{\beta c}\sum_i eV_i\delta(\theta - \theta_i)\left(D_{x_i}\frac{p_x}{p_0} - D'_{x_i}x\right) , \tag{39}$$

according to (38), and separating the contributions of φ and D_x.

To single out each nonlinear term of power q in $\Delta\varphi = \varphi - \varphi_s$ (see Appendix), H_1 can be expanded by using

$$\begin{aligned}\sin\varphi - \sin\varphi_s &= \cos\varphi_s\Delta\varphi - \frac{\sin\varphi_s}{2}\Delta\varphi^2 - \frac{\cos\varphi_s}{3!}\Delta\varphi^3 + \ldots \\ &\stackrel{\text{def}}{=} \sum_q \frac{\alpha_q}{q!}(\Delta\varphi)^q ,\end{aligned} \tag{40}$$

which leads to the form

$$H_1 = \sum_q \frac{\alpha_q(\Delta\varphi)^q}{2\pi\, q!}\sum_i eV_i\left(-D_{x_i}\frac{p_x}{c\beta p_0} + D'_{x_i}\frac{x}{c\beta}\right) , \tag{41}$$

with

$$\begin{aligned}\alpha_q &= (-1)^{q/2}\sin\varphi_s \qquad \text{if } q \text{ even} , \\ \alpha_q &= (-1)^{(q-1)/2}\cos\varphi_s \quad \text{if } q \text{ odd} .\end{aligned}$$

Equation (41) shows that the perturbation considered will only give rise to SBR of the form $Q_x \pm qQ_s = p$ (only the linear term of betatron motion in H_0).

The difficulty at this point, correctly pointed out in Ref. [4], is that the longitudinal motion is not simple harmonic (Fig. 6) when the maximum stable phase $\varphi_{\max}$ is large and the terms $q > 1$ become important. This means that a consistent treatment requires not only the term $\sin\varphi - \sin\varphi_s$ of H_1 to be expanded, but also the term $\cos\varphi + \varphi\sin\varphi_s$ of H_0. The proposed way to do this consists in finding the solutions, formally written as

$$\varphi = \sum_q a_q \cos\, q(Q_s\theta + \psi) , \tag{42}$$

of the differential equation (see the Appendix)

$$\varphi'' + \frac{Q_{s0}^2}{\cos\,\varphi_s}(\sin\varphi - \sin\,\varphi_s) = 0 , \tag{43}$$

in which Q_s is in general different from the value Q_{s0} valid at the small amplitude, where the motion is almost a simple harmonic (Fig. 6),

$$Q_{s0}^2 = \frac{h\eta\hat{V}e}{2\pi\beta p_0 c}\cos\varphi_s . \tag{44}$$

It is very difficult to get closed expressions for a_q at $\varphi_s \neq 0$ and Ref. [4] proposes an approximation at $\varphi_s = 0$ and q odd (even orders do not occur), with $\varphi_{\max}$ as the amplitude of the phase modulation

$$a_q = \frac{8}{q}\frac{(2\lambda)^{q/2}}{2^q + \lambda^q} \tag{45}$$

$$\text{with} \quad \lambda = \frac{1 - \sqrt{\cos \varphi_{max}/2}}{1 + \sqrt{\cos \varphi_{max}/2}} \quad \text{and} \quad \frac{Q_s}{Q_{s0}} = \frac{1}{(1+\lambda)^2} \, .$$

The canonical angle is ψ and the associated action I_s is hidden in the coefficient a_q; the latter is proportional to the area of the closed unperturbed orbit in phase space.

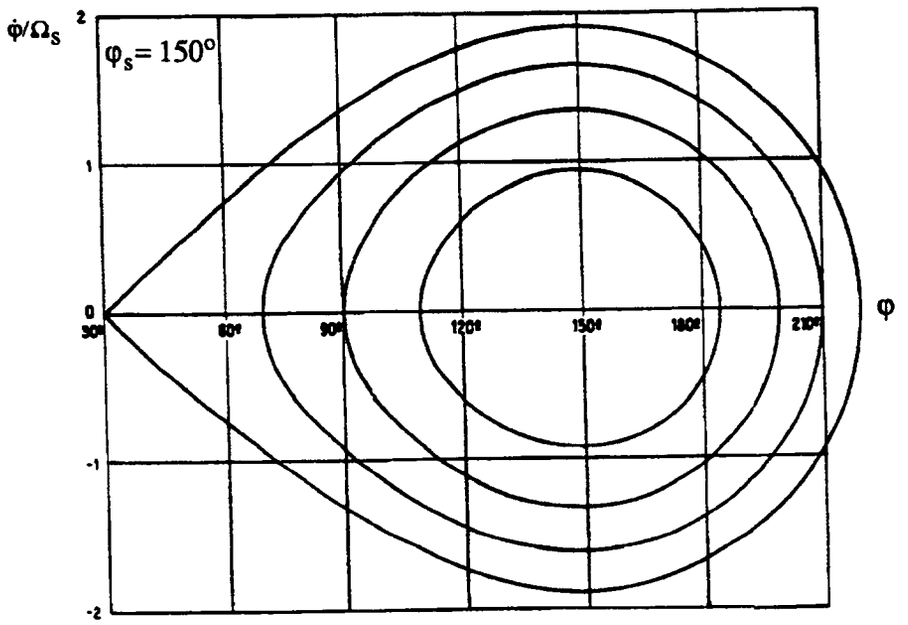


Fig. 6: Invariants of the longitudinal motion with $\varphi_s = 150°$, in the phase space defined by the coordinates φ and $\dot{\varphi}/\Omega_s$.

3.2 Analysis of the perturbed dynamics

In order to analyse further the perturbed motion, it is convenient to use the canonical action-angle variables for both the betatron and synchrotron motion. The link between these variables and physical quantities results from the forms of the equations and solutions of the unperturbed motions recalled in the Appendix (with $a_1 \sim \sqrt{I_x}\, \exp(i\varphi_x)$ and $a_1 \bar{a}_1 \sim I_x$, in the betatron case and similarly in the synchrotron case using I_s and ψ)

$$\begin{aligned}
x &= \left(\frac{2\beta_x I_x}{p_0}\right)^{1/2} \cos\left(\mu_x + \varphi_x\right) \\
p_x &= -\left(\frac{2p_0 I_x}{\beta_x}\right)^{1/2} \left[\alpha_x \cos\left(\mu_x + \varphi_x\right) + \sin(\mu_x + \varphi_x)\right] \\
W &= -\left(\frac{2cp_0\beta Q_s I_s}{h^2|\eta|\omega_0}\right)^{1/2} \sin\left(\pm Q_s\theta + \psi\right) \\
\Delta\varphi &= \left(\frac{2h^2|\eta|\omega_0 I_s}{cp_0\beta Q_s}\right)^{1/2} \cos\left(\pm Q_s\theta + \psi\right) \, .
\end{aligned} \tag{46}$$

Most of the symbols used in these equations are defined in the Appendix and ω_0 is the angular revolution frequency. The quantity η that changes sign across the transition is taken as its absolute value, but in the last two relations, the sign + must be retained below transition energy ($\eta > 0$) and the sign − above ($\eta < 0$).

The next step consists in finding the transformed Hamiltonians H_0 and H_1 for the new set of variables $\{I_x, \varphi_x, I_s, \psi\}$, and this is done by performing the transformations defined by (46). The same basic assumption is then introduced, as for the case of the time varying tune in Section 2.2: i.e. the low frequency part of H_1 gives the important contribution to the perturbation and the resonant terms are the only ones kept. The resonance condition can now be written (with positive integers q and p)

$$Q_x \pm q\,Q_s - p = 0\ , \tag{47}$$

and is of first order in Q_x since only the linear betatron motion is included. However, the sideband number q can be as large as desired, for the nonlinear terms of the synchrotron motion have been taken into account. Close to the resonance (47), i.e. $Q_x \pm qQ_s - p = \epsilon$, the canonical transformations give eventually,

$$\begin{aligned} H_0 &= \pm\frac{\epsilon}{q} I_s \qquad (48)\\ H_1 &= \frac{\cos\varphi_s}{4\pi\, c\beta}\left(\frac{Q_s}{Q_{s_0}}\right)^2 q^2 a_q \left(\frac{2I_x}{\beta_x p_0}\right)^{1/2} \\ &\quad \times[(D_{c,p} - F_{s,p})\sin(\varphi_x \pm q\psi) + (D_{s,p} + F_{c,p})\cos(\varphi_x \pm q\psi)]\ , \end{aligned}$$

in which the consistent development (42) of φ [4] has been included and the contribution of the finite horizontal dispersion is given by Fourier's expansions [3]

$$\begin{aligned} D_{s,p} &= \sum_i eV_i D_{x_i} \sin\, p\varphi_{x_i} \\ D_{c,p} &= \sum_i eV_i D_{x_i} \cos p\varphi_{x_i} \\ F_{s,p} &= \sum_i eV_i (D'_{x_i}\beta_{x_i} + D_{x_i}\alpha_{x_i}) \sin\, p\varphi_{x_i} \\ F_{c,p} &= \sum_i eV_i (D'_{x_i}\beta_{x_i} + D_{x_i}\alpha_{x_i}) \cos\, p\varphi_{x_i}\ . \end{aligned} \tag{49}$$

These expansions sum the contributions of all the discrete cavities positioned at $\theta = \theta_i$, and α_x is related to the derivative of β_x by $\beta'_x = -2\alpha_x$.

The perturbed motion is now entirely described by H_1 (48) and the perturbation theory (Appendix) tells us that

$$\frac{dI_x}{d\theta} = -\frac{\partial H}{\partial \varphi_x} \qquad \text{and} \qquad \frac{dI_s}{d\theta} = -\frac{\partial H}{\partial \psi}\ . \tag{50}$$

From the expression of H_1 and the form of the phase term involved, i.e. $\varphi_x \pm q\psi$, the following relation holds

$$\frac{dI_x}{d\theta} = \pm\,\frac{1}{q}\,\frac{dI_s}{d\theta}\ , \tag{51}$$

and after integration this gives the form of one invariant of the motion,

$$\begin{aligned} qI_x \pm I_s = \text{constant} \quad &\text{for } Q_x \pm qQ_s \quad \text{above transition}\ , \\ &\text{for } Q_x \mp qQ_s \quad \text{below transition}\ . \end{aligned} \tag{52}$$

The role of the sum and difference resonances are indeed interchanged above and below transition. According to (52) the amplitude growth is limited below transition for a difference resonance and above transition for a sum resonance. Knowing the initial emittances, Eq. (52) indicates where this limit is, though the growth in betatron amplitude can be important because I_s is usually much larger than I_x.

The first equation of motion (50) makes it possible to calculate the growth rate per revolution of the beam size. It results from the Hamiltonian (48),

$$\begin{aligned} I_x' &= -\frac{\cos\varphi_s}{4\pi\beta c}\left(\frac{Q_s}{Q_{s0}}\right)^2 q^2 a_q \left(\frac{2I_x}{\beta_x p_0}\right)^{1/2} \\ &\quad \times[(D_{c,p}-F_{s,p})^2+(D_{s,p}+F_{c,p})^2]^{1/2}\sin(\varphi_x \pm q\psi+\psi_0)\ , \end{aligned} \tag{53}$$

where ψ_0 is a constant phase, and the prime denotes differentiation with respect to θ. Since the change δI_x per revolution of I_x is simply,

$$\delta I_{x_{\rm rev}} = 2\pi I_x' \stackrel{\rm def}{=} \frac{p_0}{2}\,\delta\epsilon_x\ , \tag{54}$$

the maximum change per revolution of the emittance ϵ_x is obtained by combining Eqs. (53) and (54),

$$\delta(\epsilon_x)_{\max} = \frac{\cos\varphi_s}{\beta c p_0}\left(\frac{Q_s}{Q_{s0}}\right)^2 q^2|a_q|\sqrt{\frac{\epsilon_x}{\beta_x}}[(D_{c,p}-F_{s,p})^2+(D_{s,p}+F_{c,p})^2]^{1/2}\ . \tag{55}$$

Similarly, the second equation of motion (50) and the relation between I_s and the energy spread give the maximum growth of $\Delta E/E$ per revolution [3, 4],

$$\delta\left(\frac{\Delta E}{E}\right)_{\max} = \frac{\beta^3\cos\varphi_s}{2cp_0R}\left(\frac{Q_s}{Q_{s0}}\right)^2 q^3|a_q|\sqrt{\frac{\epsilon_x}{\beta_x}}\frac{Q_s}{|\eta|(\Delta E/E)}[(D_{c,p}-F_{s,p})^2+(D_{s,p}+F_{c,p})^2]^{1/2}. \tag{56}$$

These growth rates are given here at the resonance, but usual treatments of standard effects like multiple, fast-crossing of the resonance [1] also apply to the particular phenomenon discussed here.

3.3 Comparison with simulation and application

We shall first summarize briefly the discussion presented in Refs. [3, 4] about the two rings, PETRA and the TRIUMF booster, which were designed with finite horizontal dispersion in the accelerating cavities. Secondly, the observations made at LEP (Large Electron-Positron storage ring at CERN), a ring with zero dispersion at cavities by design, are also presented.

The PETRA ring of DESY (Hamburg) has been retained [3] to compare the Hamiltonian theory with numerical simulations. The most relevant parameters are an energy of 23 GeV, a synchronous phase of 38°, a synchrotron tune of 0.125, and a horizontal dispersion of 2 m. The latter value indicates that the SBR effects are probably going to be strong. Simulations were published in Ref. [10] and the rise times of the betatron amplitude for a $6\sigma_x$ beam size and a $6\sigma_E$ energy spread were computed analytically [4] on the basis of the development described in Section 3.2. Table 3 gives the results for a few sidebands and shows that the agreement between simulations and theory is quite good, including that for $q = 5$.

Table 3
Betatron amplitude rise-times in PETRA

Sideband q	Simulated (μsec)	Theoretical (μsec)
1	24	27
2	48	42
3	312	260
4	1020	1250
5	1065	1800

This comparison shows that we may have confidence in the Hamiltonian theory and the few closed expressions such as (55) and (56) which are convenient for rapid numerical evaluations. Therefore, working on the design of a 3 GeV booster ring of the TRIUMF Kaon Factory project, people used the theory to get some insight about synchrobetatron effects and their compensation. The dispersion in the cavities was large (3–4 m) because a lattice with high transition energy was desired [11]. The tune Q_x was 4.24, the superperiodicity of the machine five, and three of the five long straight sections were filled with cavities. The effects were then expected to be inherently strong at injection energy ($T = 458$ MeV) and for about 1 ms (or 1300 revolutions approximately) after injection, while Q_s was reaching a maximum of 0.04. The corresponding increase $\Delta I_x/I_x$ was calculated [3] using Eq. (53) and multiple resonance-crossing estimates [1]. This amplitude growth was found to reach a factor 13 for $q = 1$, 10% for $q = 2$ and 50% for $q = 3$. For higher sidebands with q values larger than 4, it remains however below 4%. This effect is consequently unacceptable up to $q = 3$, and there are basically only two possible remedies:

i) The most obvious is to make the dispersion and its derivative zero at the cavities.

ii) Another solution consists of using the phase relations (49) that depend on the positions of the cavities and the phase advances between them. The contribution of the dispersion may vanish if the cavities are placed symmetrically, and the integer part of the tune chosen to be different from the ring superperiodicity.

For the TRIUMF booster, one suggestion [12] was for instance to adopt a superperiodicity of 6, a threefold symmetry in the cavity positioning, and a tune above 5, for it is not a multiple of three.

LEP at CERN is a collider designed with zero dispersion and dispersion derivative at the cavities. Effects of SBR related to cavities are therefore only caused by spurious dispersion that was measured to be of the order of 20 to 40 cm. Consequently, the synchrobetatron coupling may appear as well in the vertical plane as in the horizontal one. Moreover, non-zero chromaticity may also contribute to the excitation of such resonances. Systematic investigations of SBR were carried out to study their possible role in intensity limitations, in relation with the spurious dispersion observed, and dispersion bumps generated in the RF cavities [13]. With tune integers equal to 70/76 and $Q_s \cong 0.086$, two different scans were performed. Firstly, the coherent tunes were measured as functions of the beam current by scraping the beam, initially at 612 μA. The incoherent tunes stay constant while the coherent ones increase, therefore crossing resonance satellites as the

current decreases. For instance, vertical bunch instability was observed when the vertical coherent tunes were slightly below the resonance $Q_y = 3Q_s$. Secondly, the vertical emittance was measured as a function of Q_y, at constant Q_x, and for a beam current of $\sim$ 200 μA. It shows sharp peaks at Q_y-values just below $3Q_s$ and $2Q_s$ (Fig. 7). As the incoherent vertical tunes are far away ($\sim$ 0.023) from SBR, these observed resonances have to be coherent in contrast to the single-particle description made above that concerns incoherent effects. An interesting fact is that the lifetime shows dips corresponding to the emittance peaks (shifted towards lower Q_s values, maybe because of the tune spread in the bunch). If this correlation is true, it could be a possible mechanism which limits the intensity in LEP; for higher currents, the split between coherent and incoherent tunes increases and the tune diagram-region free from SBR is reduced. For Q_y close to 0.2, a vertical beam pulsation at intervals of a few seconds was noticeable, and this phenomenon could perhaps be explained by the coupling resonance $Q_x - Q_y = Q_s$. Note that if Q_s was in fact slightly below 0.086, the observations would agree better with the resonance line positions. The remedies for the instability of the above-mentioned mechanism are the same as those given in previous sections for incoherent resonances. On the one hand, the residual dispersion has to be controlled using either local bumps or tilted quadrupoles, on the other hand, the betatron tune has to be optimized in order to avoid damaging SBR.

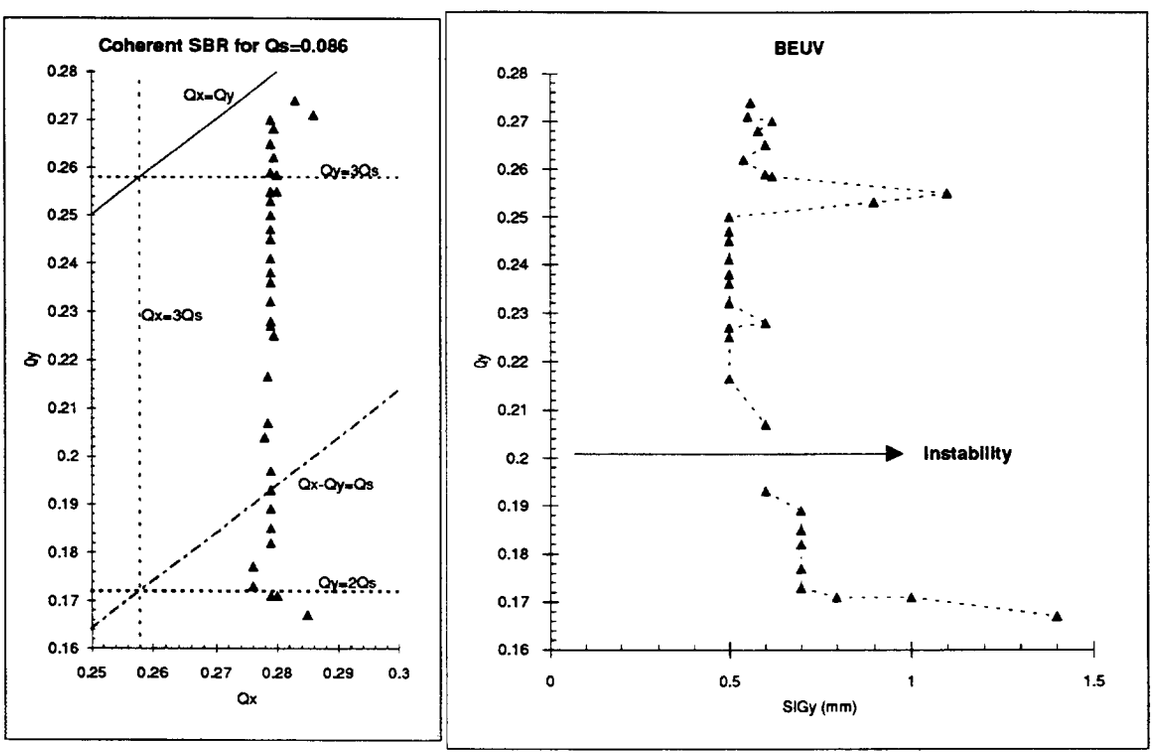


Fig. 7: Synchrobetatron resonances in the tune diagram and observed vertical emittance growth in LEP.

Beside these coherent resonances due to a residual dispersion different from zero in the cavities, SBR associated with the synchrotron motion have also been observed; they can be driven during energy ramping by tune and chromaticity excursions. Though SBR are present in LEP, measuring and correcting these excursions as well as controlling dispersion and tunes circumvent their effects, so that the beam intensity is rather limited by other mechanisms such as long range beam–beam forces and transverse mode coupling.

References

[1] G. Guignard, Nonlinear Dynamics, in 'Physics of Particle Accelerators', Eds. M. Month and M. Dienes, AIP Conf. Proc. **184**, Am. Inst. of Physics, New York (1989).

[2] G. Guignard, Isolated Beam-Beam Resonances, and Tune Modulation and Satellite Resonances, in Proc. of the Workshop on $p\bar{p}$ in the SPS, CERN report SPS–$p\bar{p}$–1 (1980).

[3] T. Suzuki, Synchrobetatron Resonance Driven by Dispersion in RF Cavities, Particle Accelerators, Vol. 18, (1985) pp. 115-128.

[4] R. Baartman, Synchrobetatron Resonance Drive by Dispersion in RF Cavities: a Revised Theory, Design Note of TRIUMPH, Vancouver, TRI-DN–89-K40 (1989).

[5] J. Hagel and T. Suzuki, Analytic Theory of the Coupled Synchrobetatron Motion in Electron Storage Rings, CERN Report, CERN/LEP–TH/88-57 (1988).

[6] I.S. Gradshteyn and I.M. Ryzhik, Table of Integrals, Series and Products, Academic Press Inc., New York and London (1973).

[7] B.V. Chirikov, Research Concerning the Theory of Nonlinear Resonance and Stochasticity, Nuclear Physics Inst., USSR Academy of Sciences, Report 267 (1971).

[8] CERN-SPS Division, Theoretical Aspects of Machine Design, Proc. of Workshop on $p\bar{p}$ in the SPS, CERN Report SPS–$p\bar{p}$–1 (1980).

[9] L.R. Evans, J. Gareyte, Beam-Beam and Single Beam Effects in the SPS Proton–Antiproton Collider, Proc. PAC83, Santa Fe, USA (1983).

[10] A. Piwinski and A. Wruhlich, DESY Report, DESY 76/07, Hamburg (1976).

[11] J. Botman, private communication.

[12] T. Suzuki, TRIUMF Report, TRI-DN–84-61, Vancouver (1984).

[13] J. Gareyte et al., Synchro-Betatron Resonances as an Intensity Limitation Mechanism, CERN, LEP Performance Note 57, July 1991.

APPENDIX A: SUMMARY OF PERTURBATION PRINCIPLES IN ACCELERATORS

A.1 Brief reminder of perturbation theory

The method of the perturbation of the constants in the formalism of the classical Hamiltonian treatment is described elsewhere [1] in some detail and applied to a different source of linear motion disturbance. However, since it will be used extensively in this paper for the synchrobetatron coupling, its main points are briefly recalled below for the sake of completeness.

The motion is defined as usual by the total Hamiltonian H, which is the sum of the Hamiltonian H_0 of the unperturbed motion, and of the perturbing Hamiltonian H_1, i.e.

$$H = H_0(p, q, \theta) + H_1(p, q, \theta) \ , \tag{A.1}$$

where θ is the independent variable and p, q are the $2N$ canonical variables in an N-dimensional space.

The perturbation treatment requires four basic steps:

i) To solve the canonical equations of the unperturbed motion described by H_0

$$\left.\begin{aligned} \dot{q}_\rho^{(0)} &= \frac{\partial H_0}{\partial p_\rho^{(0)}} \\ \dot{p}_\rho^{(0)} &= -\frac{\partial H_0}{\partial q_\rho^{(0)}} \end{aligned}\right\} \Longrightarrow \quad \begin{aligned} q_\rho^{(0)} &= q_\rho^{(0)}(a_j, \theta) \qquad \rho = 1, ..., N \\ p_\rho^{(0)} &= p_\rho^{(0)}(a_j, \theta) \qquad j = 1, ..., 2N \end{aligned} \tag{A.2}$$

where a_j are the $2N$ arbitrary constants along the unperturbed trajectories.

ii) As a basic principle, the perturbed motion in the space $q - p - \theta$ can be followed by a set of a_j-values that change with the variable θ and describe the effect of the perturbation H_1

$$q_\rho = q_\rho^{(0)}[a_j(\theta), \theta] \quad p_\rho = p_\rho^{(0)}[a_j(\theta), \theta] \ . \tag{A.3}$$

However, the solution of the perturbed motion keeps the form of the solution (A.2) of H_0 with time-dependent a_j. Since $a_j(\theta)$ become the new canonical variables, the Hamiltonian H_1 must be rewritten as a function of a_j and θ

$$H_1(p_\rho, q_\rho, \theta) = H_1[a_j(\theta), \theta] \stackrel{\text{def}}{=} U(a_j, \theta) \ . \tag{A.4}$$

The last equation is also a definition of the function U.

iii) To solve the equations of the unperturbed motion i) for the constants a_j

$$a_j = a_j(p_\rho, q_\rho, \theta) \ . \tag{A.5}$$

These functions are determined by the form of H_0 and form a system of coordinates in the phase space.

iv) Find out the expressions $a_j = f_j(\theta)$ by solving the differential equations that result from the presence of a perturbation H_1 [1],

$$\frac{da_j}{d\theta} = [a_j, H_1] = \sum_{m=1}^{2N} [a_j, a_m] \frac{\partial U}{\partial a_m} , \tag{A.6}$$

using the following definition of the Poisson bracket

$$[a_j, a_m] = \sum_{\rho=1}^{N} \left[\frac{\partial a_j}{\partial q_\rho} \frac{\partial a_m}{\partial p_\rho} - \frac{\partial a_j}{\partial p_\rho} \frac{\partial a_m}{\partial q_\rho} \right] . \tag{A.7}$$

The $2N$ first-order equations (A.6) give the explicit solution of the perturbed motion and so far everything is exact.

A.2 Specific case of single-particle motion

Let us first consider the transverse oscillations of the betatron motion which satisfy Hill's equation in the absence of perturbation, e.g. in the horizontal plane

$$x'' + K(\theta)x = 0 \quad \text{with} \quad K(\theta) = K(\theta + 2\pi) . \tag{A.8}$$

In order to show how to apply the perturbation theory in this case, when some perturbation is present, let us follow the four steps recalled in the preceding section:

i) The solution of the unperturbed motion results from the use of Floquet's theorem which states that any Hill's equation of the form (A.8) has a general solution which looks like

$$\begin{aligned} x^{(0)}(\theta) &= a_1 u(\theta) e^{iQ_x\theta} + \bar{a}_1 \bar{u}(\theta) e^{-iQ_x\theta} \\ p_x^{(0)}(\theta) &= a_1 (u' + iQ_x u) e^{iQ_x\theta} + \bar{a}_1 (\bar{u}' - iQ_x \bar{u}) e^{-iQ_x\theta} , \end{aligned} \tag{A.9}$$

where $x^{(0)}$, $p_x^{(0)}$ stand for $q_1^{(0)}, p_1^{(0)}$ of the previous section, a_1 and $\bar{a}_1$ are complex constants equivalent at the a_j s of Section A.1, and $u(\theta)$ is a complex periodic function with the same period as the focusing function $K(\theta)$,

$$u(\theta) = \sqrt{\frac{\beta_x(\theta)}{2R}} \exp[i(\mu_x - Q_x\theta)] , \tag{A.10}$$

in which $\beta_x(\theta)$ is the familiar horizontal betatron amplitude. The horizontal phase advance $\mu_x(\theta)$ and the horizontal wave number Q_x (number of betatron oscillations per turn or tune) are defined by

$$\mu_x(\theta) = \int_0^\theta \frac{R d\xi}{\beta_x(\xi)} , \quad Q_x = \frac{\mu_x(2\pi)}{2\pi} . \tag{A.11}$$

In the case of two-dimensional motion (horizontal and vertical), the same description applies two times with the following equivalence in the notation:

$$\begin{aligned} u(\theta) &\longleftrightarrow v(\theta) \\ \beta_x(\theta), \mu_x(\theta), Q_x &\longleftrightarrow \beta_z(\theta), \mu_z(\theta), Q_z \\ a_1, \bar{a}_1 &\longleftrightarrow a_2, \bar{a}_2 \\ x^{(0)}, p_x^{(0)} &\longleftrightarrow z^{(0)}, p_z^{(0)} . \end{aligned} \tag{A.12}$$

This provides the description of the form (A.2) for the betatron oscillations with $N = 2$ and with four arbitrary constants.

ii) In the presence of a perturbation in the two-dimensional betatron dynamics, the perturbed motion is described accordingly to Eq. (A.3), by

$$\begin{aligned} x &= x^{(0)}\,[a_1(\theta), \bar{a}_1(\theta), \theta] \\ p_x &= p_x^{(0)}\,[a_1(\theta), \bar{a}_1(\theta), \theta] \\ z &= z^{(0)}\,[a_2(\theta), \bar{a}_2(\theta), \theta] \\ p_z &= p_z^{(0)}\,[a_2(\theta), \bar{a}_2(\theta), \theta]\,. \end{aligned} \tag{A.13}$$

where $a_1, \bar{a}_1, a_2$ and $\bar{a}_2$ are then supposed to vary with θ and to become the new variables $(a_1 \sim \sqrt{I_x}\exp(i\varphi_x), a_2 \sim \sqrt{I_z}\exp(i\varphi_z)]$.

The Hamiltonian H_1 has to be rewritten as a function of the constants a and this can be done formally if the function H_1 satisfies some assumptions, though the explicit form of U depends on the problem treated. In general, the form of U is subordinated to the following properties:

a) H_1 is made of homogeneous polynomials of degree N in the coordinates x, z and momentum conjugates p_x, p_z.

b) The solutions (A.13) are linear functions of the four constants $a_1, \bar{a}_1, a_2, \bar{a}_2$ and contain oscillatory terms with frequencies Q_x and Q_z, by virtue of the form of equation (A.9).

c) For circular accelerators and storage rings, the perturbation is obviously periodic in θ with period 2π.

Introducing the linear equations (A.9) in a polynomial of degree N generates a sum of terms with the same degree and all possible combinations of powers for the four constants a, the coefficients of these terms being termed $h(\theta)$. The periodicity of the perturbation suggests that the coefficients $h(\theta)$ be developed into Fourier's series. All this provides the following form for the Hamiltonian $U = H_1$,

$$\begin{aligned} U(a_1, \bar{a}_1, a_2, \bar{a}_2, \theta) = \sum_{N} \sum_{\substack{j,k,\ell,m=0 \\ j+k+\ell+m=N}}^{N} \sum_{q=-\infty}^{\infty} h^{(N)}_{jk\ell mq}(\theta) a_1^j \bar{a}_1^k a_2^\ell \bar{a}_2^m \\ \times \exp\{i[(j-k)Q_x + (\ell - m)Q_z + q]\theta\}\,, \end{aligned} \tag{A.14}$$

All the coefficients h depend on the Floquet functions u and v by virtue of (A.9) and on the perturbation strengths that multiply linearly each polynomial of degree N.

iii) The equations (A.9) (and their equivalent ones in the vertical plane) can be solved for obtaining expressions of a_1 and $\bar{a}_1$ as functions of x and p_x,

$$\begin{aligned} a_1 &= i[(\bar{u}' - iQ_x\bar{u})x - \bar{u}p_x]e^{-iQ_x\theta} \\ \bar{a}_1 &= -i[(u' + iQ_x u)x - up_x]e^{iQ_x\theta}\,, \end{aligned} \tag{A.15}$$

while similar equations hold for $a_2, \bar{a}_2$ as functions of z, p_z, v and Q_z.

iv) Using the equations (A.6), (A.7), and (A.15) (plus their equivalent ones in the vertical plane), as well as the Floquet functions u (A.10) and v, the differential equations can be given for the a_j's explicitly and simplify remarkably,

$$\begin{aligned} \frac{d\bar{a}_1}{d\theta} &= -i\frac{\partial U}{\partial a_1} \qquad & \frac{da_1}{d\theta} &= i\frac{\partial U}{\partial \bar{a}_1} \\ \frac{d\bar{a}_2}{d\theta} &= -i\frac{\partial U}{\partial a_2} \qquad & \frac{da_2}{d\theta} &= i\frac{\partial U}{\partial \bar{a}_2} \,. \end{aligned} \tag{A.16}$$

Let us next consider the longitudinal oscillations of the synchrotron motion which satisfy the two following equations, when θ is the independent variable

$$\begin{aligned} \frac{dW}{d\theta} &= -\frac{e\hat{V}}{2\pi\,\omega_{rf}}\,(\sin\varphi - \sin\varphi_s) \\ \frac{d\varphi}{d\theta} &= \frac{h^2\eta}{Rp_0}\,W \,. \end{aligned} \tag{A.17}$$

It must be noted at this point that the definition of W used in these equations, is $W = -\Delta E/\omega_{rf}$, where ΔE is the energy deviation from the synchronous value, and ω_{rf} is the RF angular frequency; and this definition differs from the one commonly quoted in the literature by a factor of $-1/2\pi h$, where h is the RF harmonic number. The other quantities have their usual meaning; φ is the RF phase, φ_s the synchronous RF phase, $\hat{V}$ the RF voltage, η the change of frequency with momentum deviation, R the average radius (of the synchronous particle), and p_0 the central momentum. Combining these two equations (A.17), we obtain a second-order equation as expected

$$\frac{d}{d\theta}\left[\frac{Rp_0}{h^2\eta}\,\frac{d\varphi}{d\theta}\right] + \frac{e\hat{V}}{2\pi\omega_{rf}}\,(\sin\varphi - \sin\varphi_s) = 0 \,, \tag{A.18}$$

which becomes, when considering time t and using $\theta = \omega_s t$ and $\omega_{rf} = h\omega_s$

$$\frac{d}{dt}\left[\frac{Rp_0}{h\eta\omega_s}\,\frac{d\varphi}{dt}\right] + \frac{e\hat{V}}{2\pi}\,(\sin\varphi - \sin\varphi_s) = 0 \,. \tag{A.19}$$

The parameters in the square brackets either being constant or varying slowly, it is possible to write the following equations, the second of which is an approximation of the first when $\Delta\varphi = \varphi - \varphi_s$ is small

$$\begin{aligned} & \ddot{\varphi} + \frac{h\eta\omega_s e\hat{V}}{2\pi Rp_0}\,(\sin\varphi - \sin\varphi_s) = 0 \\ \text{or} \qquad & \ddot{\varphi} + \Omega_s^2\,\Delta\varphi = 0 \,, \end{aligned} \tag{A.20}$$

where $\Omega_s^2/\cos\varphi_s$ is equal to the bunch of constants entering the first equation. The last equation is very similar to the equation (A.8) for betatron oscillations. It describes the synchrotron oscillations of non-synchronous particles in terms of the synchrotron angular frequency Ω_s or synchrotron tune Q_s (when one keeps θ instead of using time t) given by $R\Omega_s/\beta c$ and of an amplitude a [$\sim \sqrt{I_s}\,\exp(i\psi)$ in complex notation], similar to a_1 of Eqs. (A.9). Starting from this, perturbation theory can be applied to synchrotron motion, just as described above for betatron motion.

MODERN TOOLS FOR PARTICLE TRACKING

F. Willeke
DESY, Hamburg, Germany

Abstract
In order to investigate the stability of single particles in large hadron accelerators it is necessary to track particles for a long time. To avoid unreasonably large computing times, it has been proposed to use truncated Taylor maps. This lectures describes recently developed tools to generate such truncated maps and to make the maps symplectic, a property necessary for long term tracking.

1 INTRODUCTION

The material presented in this chapter is part of a lecture on particle tracking given at the Advanced Accelerator Physics Courses of the CERN Accelerator School in Uppsala, September 1989, and in Rhodos, September 1983. It is a complement to the article published in the proceedings of the Uppsala School [1].

Particle tracking is a widely used tool, in fact still the only reliable tool for investigating the stability of particle beams exposed to nonlinear fields in circular accelerators. For large accelerators in particular, the designer wishing to use this tool to make decisions on tolerable magnetic field imperfections is confronted with two conflicting requirements. The weak incoherent instabilities of particle beams in circular accelerators caused by small nonlinear field distortions take a long time to develop. In order to decide on stability one therefore has to track the particles for a long time. This requires a large amount of computing time. On the other hand, the beam behavior might be influenced by effects like exact distribution of the nonlinear field errors, small oscillations of the fields (as caused by power supply ripples) or collective effects. This requires a precise modelling of the accelerator and the beam physics. A very detailed description is necessary which also leads to long computing times.

We are looking therefore for methods which resolve the conflict between precision of the model and the required long tracking times. Recently it was proposed to use truncated polynominal maps in phase-space coordinates for particle tracking. These maps should contain up to the necessary order all the relevant physical effects. The idea to use a truncated map instead of element-by-element tracking of accelerator particles is not new. But due to rather recent developments, this approach has now become feasible and has been applied successfully in tracking for large accelerators [2] [3] [4]. This lecture describes the models and procedures necessary to turn truncated maps into a useful tool for the investigation of particle dynamics in accelerators.

If one integrates the equation of motion of particles which move around the accelerator the result is a one turn map which describes the transformation of initial into final phase space coordinates. The usual technique is to evaluate the coordinate transformation element by element by approximating the nonlinear forces by an instantaneous kick. The map for one element is then given by a polynominal in the coordinates at the entrance of the element. Typically, this polynominal is of degree up to 10. The resulting implicit one-turn map for a large accelerator with thousands of single elements contains very high powers ($\geq 10^{1000}$) of the coordinates. The dynamics of the particle however is dominated by the lower-order terms. Even the lowest-order nonlinear terms of the map contribute after a few hundreds of turns to very high-order mapping which is beyond the resolution of digital numbers. Therefore one might say that if it is justified to consider tracking as a relevant tool to investigate stability of the motion of accelerator particles, then the essential physics is likely to be described by relatively low orders of the total map. What relatively low order means however depends on the problem which is investigated. This is the basic assumption in the use of a truncated map. Such a map may be explicitly given as polynominals in the initial phase space coordinates. The iteration of such a map will be referred to as 'map tracking'. The technical resources allow the generation and iteration of explicit power series of order up to about ten. There is a loss of information associated with the use of truncated maps which is the price one has to pay for a fast tracking which is independent of the complexity of the physical model under investigation. The experience obtained so far suggests that map tracking indeed provides the required compromise, at least for large accelerators with weak nonlinearities.

There are several problems which have to be solved before map tracking is turned into a useful

tool:

- We have to generate a power series map
- Since a truncated power series is not an exact solution of the Hamiltonian equations of motion for an accelerator particle, the map is not symplectic which means it may contain small pseudo dissipative terms which are not in the physical model. These terms have to be cancelled. In other words, the map has to be 'symplectified'.
- After symplectification, the map has passed through a metamorphosis by which, as we shall see, it has apparently lost its main advantage. It is no longer easy to evaluate. Therefore one has to develop special representations of the symplectified map which restore the original advantage at the expense of further approximations.

Each of the tasks listed above is quite difficult and cumbersome or at least extremely tedious. However two tools have been made available in recent years which have allowed considerable progress and which finally provided the technical solutions. These tools include the Lie algebraic description of beam dynamics as developed by A.Dragt and co-workers and the automatic differentiation introduced by M. Berz to accelerator physics by the name of 'Differential Algebra'. On both topics there is a rich bibliography available. Some of the articles suggested for reading are included in the references [5] [6] [7] [8] [9] [10].

The scope of this lecture is as follows: First there is an introduction to automatic differentiation and generation of power series (Taylor series, polynominal) maps which is followed by a discussion of the issue of symplecticity. The next topic will be an introduction to the Lie algebraic language with the most important rules and procedures used to make polynominal maps symplectic. The last step is then to turn the Lie transformations into a form which allows them to be used effectively for particle tracking.

2 GENERATION OF TRUNCATED MAPS USING DIFFERENTIAL ALGEBRA

The purpose of using truncated maps for particle tracking is to save computer time. The main idea is that the dynamics of particles is determined by the lower-order terms of the complete map which describes a turn around the accelerator. There is of course no proof of this statement. Whether, or to what degree, this is true has to be tested case by case. However since the weak instabilities which one wants to understand build up over many thousands of terms, even a map of low order produces very high order contributions which might influence stability. Moreover it is well known that the lowest order nonlinear system with quadratic forces has already all the features of a nonlinear system such as amplitude dependence of tunes, excitation of all nonlinear resonances and chaotic behavior.

Map tracking does not come for free. A sixth-order map in six-phase space variables has already $\simeq$ 900 monominals to evaluate. In order to obtain good results (these are results which compare well with element-by-element tracking) one might have to go up to order 10 which contains $\simeq$ 8000 terms. In this case map tracking saves only a factor of 3-10 in computing time for large accelerators such as LHC. (For the HERA accelerator the evaluation of a 10th-order map requires almost as many operations on a computer as element-by-element tracking with multipolar fields up to order 10.) For smaller accelerators, map tracking can be interesting if additional features such as a detailed description of end and fringe fields is included.

A breakthrough in the use of truncated maps for tracking was the introduction of automatic differentiation called 'Differential Algebra' by M. Berz into accelerator physics. In the following, the basic ideas of this procedure will be presented.

Differential algebra is a technique for systematically propagating the derivatives of a function $f(x_i)$ through mathematical transformations on f by simply applying the familiar sum, product and chain rule of differentiation. If two functions f_1, f_2 in variables x_i are to be combined by a mathematical operation and one knows the derivatives of f_1, f_2 with respect to the x_i, then one also knows the derivatives of the result of the combination of f_1 and f_2. The derivatives of any complicated function which may be obtained by successive mapping can be calculated by extending any function f to a vector $\mathbf{f}$ which contains the value of the function as the first element and the values of the derivatives with respect to all the variables up to the desired order in the subsequent elements. Michelotti called these vectors

prolongations [10], sometimes accelerator physicists call them 'DA-vectors'.

$$f(x_i) \to \mathbf{f}(x_i) = \{f, ..\partial f/\partial x_i .., .\partial^2 f/\partial x_i \partial x_j .., ...\} = \{f, ..f_{x_i} .., ..f_{x_i x_j} .., ...\} \tag{1}$$

According to the rules of differentiation, the sum of two such vectors is defined as

$$\mathbf{f}(x_i) + \mathbf{g}(x_i) = \begin{pmatrix} f+g \\ f_{x_i} + g_{x_i} \\ . \\ f_{x_i x_j} + g_{x_i x_j} \\ . \\ . \end{pmatrix} \tag{2}$$

their product is defined as

$$\mathbf{f}(x_i) \cdot \mathbf{g}(x_i) = \begin{pmatrix} f \cdot g \\ f_{x_i} \cdot g + g_{x_i} \cdot f \\ . \\ f_{x_i x_j} \cdot g + f_{x_i} \cdot g_{x_j} + f_{x_j} \cdot g_{x_i} + g_{x_i x_j} \cdot f \\ . \\ \prod_i \sum_{m_i + k_i = n_i} \prod_i \begin{pmatrix} n_i \\ m_i \end{pmatrix} (\partial^K f / \prod_i \partial x_i^{k_i})(\partial^M g / \prod_i \partial x^{m_i}) \end{pmatrix} \tag{3}$$

($K = \sum_i k_i$; $M = \sum_i m_i$). Multiplication of the vector with a constant c is just the product of each component with that constant

$$f_i \to c \cdot f_i. \tag{4}$$

One can easily extend this to the case where the function f has to be taken as the argument of an analytic function $h(f)$ by using the chain rule of differentiation

$$\mathbf{f} \to \mathbf{g} = \begin{pmatrix} h(f) \\ h'(f) f_{x_i} \\ . \\ h''(f) f_{x_i} f_{x_j} + h' f_{x_i x_j} \\ . \\ . \end{pmatrix}. \tag{5}$$

This reduces the calculation of derivatives of complicated functions with respect to their variables to a book keeping problem.

Any computation of the value of a function $f(x_i)$ by computer is iterative and always starts with the very simple function

$$f(x_i) = x_i. \tag{6}$$

The full prolongation of this function is always known, for the first derivatives one finds

$$f_{x_j} = \delta_{ij} \tag{7}$$

and all other derivatives are zero thus

$$\mathbf{f} = \{x_i, 0, .., 0, 1, 0, .., 0\}. \tag{8}$$

In order to demonstrate the procedure let us calculate the distance of a point from the origin together with the derivatives with respect to the cartesian coordinates x, y at $x = 4$ and $y = 3$ up to order two $\mathbf{r} = (\mathbf{x}^2 + \mathbf{y}^2)^{1/2}$. The calculation starts by the assignment of two vectors

$$\mathbf{x} = \begin{pmatrix} 4 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 3 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{9}$$

In the next step we add the products of the vectors to themselves

$$\mathbf{h} = \mathbf{x}^2 + \mathbf{y}^2 = \begin{pmatrix} x_0^2 \\ 2x_0 \\ 0 \\ 2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} y_0^2 \\ 0 \\ 2y_0 \\ 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 16+9 \\ 8 \\ 6 \\ 2 \\ 0 \\ 2 \end{pmatrix} \tag{10}$$

and finally we have to take the square root

$$\mathbf{r} = \begin{pmatrix} (h)^{1/2} & \\ 1/2(h)^{-1/2} \cdot h_x & \\ 1/2(h)^{-1/2} \cdot h_y & \\ -1/4(h)^{-3/2} \cdot h_x^2 & +1/2(h)^{-1/2} \cdot h_{xx} \\ -1/4(h)^{-3/2} \cdot h_x h_y & +1/2(h)^{-1/2} \cdot h_{xy} \\ -1/4(h)^{-3/2} \cdot h_y^2 & +1/2(h)^{-1/2} \cdot h_{yy} \end{pmatrix} = \begin{pmatrix} 5 \\ 0.8 \\ 0.6 \\ 0.072 \\ -0.096 \\ 0.128 \end{pmatrix} \tag{11}$$

so that one obtains the Taylor expansion of r in x and y as

$$\begin{aligned} r \quad = \quad & 5 + 0.8 \cdot (x-4) + 0.6 \cdot (y-3) \\ & +0.036 \cdot (x-4)^2 - 0.096 \cdot (x-4) \cdot (y-3) + 0.064 \cdot (y-3)^2. \end{aligned} \tag{12}$$

In the same way the procedure is used to generate a one-turn map in initial phase-space coordinates for map tracking. All phase-space coordinates z_i become vectors $\mathbf{z_i}$. The first element of $\mathbf{z_i}$ contains the current value of the coordinate z_i, and the subsequent elements contain the derivatives with respect to the initial values of z_i. The vectors $\mathbf{z_i}$ are initialized by setting the first element to the initial value of z_i, the element which contains the first derivative with respect to the $i-th$ initial coordinate is set to one and all other elements are zero. All the mathematical operations which involve phase-space variables have then to be replaced by one of the vector operations discussed above. For example the application of a sextupole kick during the tracking is written in scalar form as

$$p_x^{final} = p_x + 1/2ml(x^2 - y^2)$$

and in DA vector form as

$$\mathbf{p_x}^{final} = \mathbf{p_x} + \frac{1}{2}ml(\mathbf{x}^2 - \mathbf{y}^2).$$

Several ways have been chosen to implement automatic differentiation on a computer. Berz has written a precompiler for FORTRAN codes. Most existing FORTRAN source codes are suited as an input for the precompiler which only has to be supplied with the information of which are the independent variables and which are to be treated as DA-vectors. The modified codes perform all calculations as before (only much more slowly). In addition to the motion of the particle, the derivations of specified quantities with respect to specified variables are calculated up to the specified order. This differential-algebra package is available from the author.

Michelotti used the advantages of the C^{++} language which makes it possible to define one's own mathematical operations between elements or objects. There are several papers where this approach is described in some detail [10, 11]. A program package in C-language is in preparation [12].

Having modified a particle tracking code such as SIXTRACK [14] or TEAPOT [13] in such a way, it may be used to generate a one-turn Taylor or polynominal map which describes the motion of a particle once around the accelerator by

$$\vec{z}_{final} = \sum_{j_1+..+j_n \leq N} \vec{a}_{j_1 \ldots j_n} \prod_i z_i^{j_i}. \tag{13}$$

In the previous section we have seen how differential algebra supplies us with a tool to generate one-turn polynominal maps of very high order. However, the capacity of modern computers is exhausted

if the order of the map is beyond about ten. Beyond this order, the evaluation of the map takes as much time as element-by-element tracking.

There is a serious drawback of polynominal maps. They are non-physical since they are the result of a truncation. The truncated map is not symplectic in general. The non-physical effects are of the order $N+1$ where N is the highest order of the map. Artificial growth of the amplitudes might be produced. Any attempt to explicitly symplectify the map is extremely difficult because one is dealing with very high orders and with a huge number of terms. In the next section the issue of symplecticity will be made clear and discussed.

3 SYMPLECTIC MAPS

The fact that we consider the motion of particles under the influence of forces which can be derived from a Hamiltonian function introduces a strong constraint on the motion of the particles. The most important consequence for accelerators is that the number of particles inside a cell of phase space is preserved. This cell is given by the phase-space coordinates of particles at the surface of that cell. There is a formal property of the maps describing the propagation of particles which is responsible for this constraint. This property is called symplecticity. In order to track the motion of particles effectively, approximations are usually necessary. If the symplecticity is violated by these approximations the consequence is an artificial growth or damping of the size of a cell in phase space. This corresponds to growth or damping of the beam emittances or beam oscillation amplitudes. A non-symplectic model thus might cause artificial instabilities which in practice may be very difficult to distinguish from genuine weak instabilities. These arise from the complex distortions of phase space under the influence of nonlinear forces. In a nonlinear system, a phase-space cell might be deformed into infinitely fine filaments which may occupy a large region of phase space. These regions may be visited by particles during their motion. Although the total volume inside the filamented structure has not changed, the volume of a surface enclosing the filaments is much larger than the volume of the filament. This might appear as a growth of beam emittance. If these regions extend outside the vacuum vessel, this eventually leads to the loss of particles. This is what we might call genuine beam instability. Not all the forces acting upon a particle in a real accelerator can be conveniently described by a Hamiltonian model. Consider for example scattering processes with the remaining gas molecules in the vacuum vessel or intrabeam scattering. Since these may interfere with nonlinear effects, which might be tested by a simulation, it is important that the starting point is a symplectic model to which other physical effects might be added in a controlled fashion. In the following, a formal criterion for symplectic behavior will be derived.

The symplectic condition is derived from the equation of motion under the influence of a Hamiltonian force

$$\begin{array}{rcl} \frac{dx_i}{dt} & = & \frac{\partial H}{\partial p_i}(x_i,p_i) \\ \frac{dp_i}{dt} & = & -\frac{\partial H}{\partial x_i}(x_i,p_i). \end{array} \tag{14}$$

The $2n$ phase space coordinates of a single particle may be represented by the phase space vector $\vec{z} = \{x_1, p_1, x_2, p_2, ...x_n, p_n\}$ and with the symplectic form $\underline{S}$

$$\underline{S} = \begin{pmatrix} 0 & 1 & . & 0 & 0 \\ -1 & 0 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & 0 & 1 \\ 0 & 0 & . & -1 & 0 \end{pmatrix} \tag{15}$$

we can write the equations of motion in matrix form

$$\frac{d\vec{z}}{dt} = \underline{S}\vec{\nabla}_z H. \tag{16}$$

Consider the Jacobian matrix $\underline{J}$ which describes how the phase space coordinates $\vec{z}(t)$ vary with an infinitesimal change in the initial conditions $\vec{z}^0$

$$J_{ik} = \frac{\partial z_i(t)}{\partial z_k^0}. \tag{17}$$

The Jacobian satisfies the symplecticity condition

$$\underline{J}^T \cdot \underline{S} \cdot \underline{J} = \underline{S}. \tag{18}$$

This can be seen by inserting the Hamiltonian equations of motion:

$$\frac{dJ_{ik}}{dt} = \frac{\partial}{\partial z_k^0}\frac{dz_i}{dt} = \frac{\partial}{\partial z_k^0}\sum_j S_{ij}\frac{\partial H}{\partial z_j} = \sum_n \frac{\partial z_n}{\partial z_k^0}\sum_j S_{ij}\frac{\partial^2 H}{\partial z_n \partial z_j} \tag{19}$$

thus

$$\dot{\underline{J}} = \underline{S} \cdot \underline{H} \cdot \underline{J}, \quad H_{ij} = \frac{\partial^2 H}{\partial z_i \partial z_j}. \tag{20}$$

Taking into account that $\underline{S} \cdot \underline{S} = -1$ and $\underline{S}^T = -\underline{S}$ one finds

$$\begin{array}{rcl} \frac{d}{dt}(\underline{J}^T \cdot \underline{S} \cdot \underline{J}) & = & \underline{J}^T \cdot \underline{H}^T \cdot \underline{S}^T \cdot \underline{S} \cdot \underline{J} - \underline{J}^T \cdot \underline{S} \cdot \underline{S}^T \cdot \underline{H} \cdot \underline{J} \\ & = & \underline{J}^T \cdot \underline{H} \cdot \underline{J} - \underline{J}^T \cdot \underline{H} \cdot \underline{J} \\ & = & 0 \\ \rightarrow \underline{J}^T \cdot \underline{S} \cdot \underline{J} & = & const. \end{array} \tag{21}$$

Since $\underline{J}(t=0) = 1$, this constant matrix must be $\underline{S}$.

The motion of particle coordinates in the vicinity $\delta\vec{z}(t)$ of a reference trajectory $\vec{z}(t)$ is described by

$$\vec{z} + \delta\vec{z} = \vec{z} + \underline{J}\delta\vec{z}_0. \tag{22}$$

Consider two orthogonal phase space vectors $\delta\vec{z_1}$ and $\delta\vec{z_2}$which define an (infinitesimally small) elliptical area in phase space. The area of this ellipse is $\pi\delta W = \pi\delta\vec{z_1}^T(t) \cdot \underline{S} \cdot \delta\vec{z_2}(t)$. Since $\underline{J}$ is symplectic, the expression

$$\begin{aligned} \delta W = \delta\vec{z_1}^T \cdot \underline{S} \cdot \delta\vec{z}_2 &= \delta\vec{z}_{10}^T \cdot \underline{J}^T \cdot \underline{S} \cdot \underline{J} \cdot \delta\vec{z}_{20} \\ &= \delta\vec{z}_{10}^T \cdot \underline{J}^T \cdot \underline{S} \cdot \underline{J} \cdot \delta\vec{z}_{20} \\ &= \delta\vec{z}_{10}^T \cdot \underline{S} \cdot \delta\vec{z}_{20} \end{aligned} \tag{23}$$

is an invariant. This invariant is called the Lagrange invariant. This means that the sum of the projections of a phase-space cell on the two-dimensional planes $x_i - p_i$ in 2n-dimensional phase space are preserved during the time evolution of this cell. This implies the volume preservation of this cell during its journey through phase space, which is also expressed by Liouville's theorem. The phase space volume element $\int dx_1 dp_1 ... dx_n dp_n$ at the time $t = 0$ transforms into $\int dx_1 dp_1 ... dx_n dp_n \times det\underline{J(t)}$ at the time t. Since $\underline{J}$ is symplectic, its determinant is equal to one

$$det(\underline{J}^T \underline{S} \underline{J}) = det(\underline{J}^T) det(\underline{S}) det(\underline{J}) = (det(\underline{J}))^2 = det(\underline{S}) = 1, \tag{24}$$

(using $det(\underline{J}(t=0)) = +1$). It is the symplecticity of the Jacobian $\underline{J}$ of the map which guarantees this important property of a Hamiltonian system.

4 SYMPLECTIFICATION OF POLYNOMIAL MAPS BY LIE ALGEBRA METHODS

As we discussed in the previous section, symplecticity of the accelerator map is important in studying weak instabilities of particle orbits. Therefore procedures which preserve the symplectic nature of the particle motion in the process of modelling the accelerator are very desirable. The methods of applying the physical forces in the form of 'kicks', sudden changes of the particle coordinate, provides such a method and is the basis of most tracking codes. For very large accelerator lattices, the tracking of particles from element to element takes a lot of computing time. It is therefore desirable to describe a whole section of the accelerator by an explicit map, the Lie algebra methods provide a useful tool. In the following it will be shown how to derive these methods starting from Hamilton's equations of motion

and the most important properties of the method will be reviewed. In doing so, A. Dragt's review article [7] will be followed.

The time evolution of a function in the coordinates x_i, p_i of a particle is described by the Poisson Bracket:

$$\frac{df}{dt} = \sum_i \frac{\partial f}{\partial x_i}\dot{x}_i + \sum_i \frac{\partial f}{\partial p_i}\dot{p}_i \tag{25}$$

$$\dot{x}_i = \frac{\partial H(x_i, p_i)}{\partial p_i}; \qquad \dot{p}_i = -\frac{\partial H(x_i, p_i)}{\partial x_i} \tag{26}$$

$$\frac{df}{dt} = \sum_i \frac{\partial f}{\partial x_i}\frac{\partial H}{\partial p_i} - \sum_i \frac{\partial f}{\partial p_i}\frac{\partial H}{\partial x_i} = [f, H] = \hat{H}f. \tag{27}$$

The action of a Poisson Bracket $[f, H]$ is written symbolically by a Poisson Bracket operator or Lie operator $\hat{H}f$. If f and H have no explicit time dependence, its evolution in time may be written as

$$\begin{aligned} f(t + \Delta t) &= \textstyle\sum_{n=0}^{\infty} \frac{1}{n!} d^n f(t)/dt^n \Delta t^n \\ &= \textstyle\sum_{n=0}^{\infty} \frac{1}{n!} (\hat{H} \cdot \Delta t)^n f \\ &= exp(\hat{H} \cdot \Delta t) f \\ &= exp(\hat{L}) f. \end{aligned} \tag{28}$$

The expression $exp(\hat{L})$ is called a Lie transformation. It describes the evolution of a function of particle coordinates in time. This function can in particular be the coordinates themselves. A Lie transformation is generated by any analytic function $L(x_i, p_i)$ of the phase-space coordinates. In particular however we will consider homogeneous polynominals such as for example

$$L_k = \sum_{n+m=k} h_{nm} x^n p^m \tag{29}$$

which describes the motion in one degree of freedom. Any analytical function in the phase space coordinates may serve as a Hamiltonian, and since the corresponding Lie transformation is a solution of this Hamiltonian system, the Lie transformation is a symplectic transformation and preserves the density in phase space.

As a demonstration of the Lie operator technique, we consider the harmonic oscillator. The equations of motion can be integrated using Lie transformations in the following way:

$$H = \frac{1}{2}(p^2 + \omega^2 x^2) \rightarrow L = \frac{1}{2} t \cdot (p^2 + \omega^2 x^2). \tag{30}$$

If we apply the Lie operator once to the particle coordinate x we get

$$\hat{L}x = \frac{\partial L}{\partial p} \cdot \frac{\partial x}{\partial x} - \frac{\partial L}{\partial x} \cdot \frac{\partial p}{\partial x} = \frac{\partial L}{\partial p} = p \cdot t. \tag{31}$$

If we apply it twice we obtain

$$\hat{L}^2 x = t \cdot \hat{L}p = t \cdot (-\frac{\partial L}{\partial x}) = -(\omega t)^2 \cdot x \tag{32}$$

and if we apply it n-times we obtain

$$\begin{aligned} \hat{L}^n x &= (-1)^{n/2} \quad (\omega t)^n \cdot x \quad (n\ even) \\ \hat{L}^n x &= (-1)^{(n-1)/2} \quad (\omega t)^n \cdot p/\omega \quad (n\ odd) \end{aligned} \tag{33}$$

so that we finally get

$$\begin{aligned} x(t) = \quad & exp(\hat{L})x_0 \\ & = \sum_{n=0,2}^{\infty} \frac{(-1)^{n/2}}{n!}(\omega t)^n \cdot x_0 + \sum_{n=1,3}^{\infty} \frac{(-1)^{(n-1)/2}}{n!}(\omega t)^n \cdot p_0/\omega \\ & = x_0 \cdot cos(\omega t) + p_0/\omega \cdot sin(\omega t). \end{aligned} \tag{34}$$

Lie transformations may be used to describe the propagation of particle orbits in an accelerator. Beam-line elements such as drift spaces, quadrupole or sextupole magnets can be described by Lie transformations using the generators

$$\begin{aligned} Drift \quad & : \quad L = \tfrac{1}{2}(p_x^2 + p_y^2) \cdot l \\ Quadrupole \quad & : \quad L = \tfrac{1}{2}(p_x^2 + p_y^2 + k(x^2 + y^2)) \cdot l \\ Sextupole \quad & : \quad L = \tfrac{1}{2}(p_x^2 + p_y^2 + \tfrac{m}{3}(x^3 - 3xy^2)) \cdot l \end{aligned} \tag{35}$$

(here only the lowest-order terms generated by these elements are taken into account). The motion of a particle through a lattice composed of various elements is then described by successive application of Lie operators each describing the motion through a beam line element analogous to the multiplication of matrices in linear beam optics.

$$\begin{aligned} x_n \quad & = exp(\hat{L}_1)(exp(\hat{L}_2)(.....exp(\hat{L}_n)x_0))..)) \\ & = exp(\hat{L}_1)exp(\hat{L}_2) \cdot .... \cdot exp(\hat{L}_n)x_0 \end{aligned} \tag{36}$$

Just as in linear optics, where transfer matrices for single elements are combined into a revolution matrix, one can combine Lie transformations into a combined transformation which describes the motion along an accelerator lattice. Unfortunately, the concatenation of Lie transformation turns out to be very tedious. It can be performed only as an approximation which is correct only up to a chosen order. The advantage of the method however is that the concatenation procedure can be carried out conveniently on a computer since it is based on repetition of simple basic steps, the derivatives of polynominals. The result will always be in the form of a Lie transformation. So it will be a symplectic transformation.

The most important tool for the concatenation of Lie transformations is the Campbell-Baker-Hausdorff formula

$$\begin{aligned} exp(\hat{L}_1)exp(\hat{L}_2) \quad & = \sum_n \sum_{n'} \tfrac{1}{n!n'!} \hat{L}_1^n \hat{L}_2^{n'} \\ & = \sum_n \sum_m \tfrac{1}{n!} \begin{pmatrix} n \\ m \end{pmatrix} \hat{L}_1^{n-m} \hat{L}_2^m \\ & = \sum_n \tfrac{1}{n!}(\hat{L}_1 + \hat{L}_2)^n + \tfrac{1}{2}[\hat{L}_1, \hat{L}_2] + ... \\ & = exp(\hat{L}_1 + \hat{L}_2 + \ higher\ order\ terms). \end{aligned} \tag{37}$$

The expression $[\hat{L}_1, \hat{L}_2]$ stands for

$$[\hat{L}_1, \hat{L}_2] = \hat{L}_1\hat{L}_2 - \hat{L}_2\hat{L}_1 = [L_1, \hat{L}_2] \tag{38}$$

The higher-order terms are formed by an infinite series of commutator operations on $\hat{L}_1$ and $\hat{L}_2$

$$+\tfrac{1}{2}[\hat{L}_1, \hat{L}_2] + \tfrac{1}{12}[\hat{L}_1, [\hat{L}_1, \hat{L}_2]] + \tfrac{1}{12}[\hat{L}_2, [\hat{L}_1, \hat{L}_2]] + ... \tag{39}$$

The coefficients of the commutator series are known up to any order. In the case that the Lie transformations are generated by polynominals, the commutator series results in another polynominal the coefficients of which can be evaluated by computer in a straight forward manner.

For each Lie transformation $exp(\hat{L})$ its inverse transformation $exp(-\hat{L})$ exists.

$$\begin{aligned} exp(-\hat{L})exp(\hat{L}) &= \sum_{n,n'} \frac{(-1)^{n'}}{n!n'!} \hat{L}^{n+n'} \\ &= \sum_m \hat{L}^m \cdot \sum_{n=0}^m \frac{(-1)^n}{n!(m-n)!} = 1 \end{aligned} \tag{40}$$

Important for the manipulation of maps is the exchange rule

$$\begin{aligned} exp(\hat{L}_1)exp(\hat{L}_2) &= exp(\hat{L}_2)exp(-\hat{L}_2) \cdot exp(\hat{L}_1)exp(\hat{L}_2) \\ &= exp(\hat{L}_2) \sum_{n=0}^{\infty} \frac{1}{n!} exp(-\hat{L}_2)\hat{L}_1^n exp(\hat{L}_2) \\ &= exp(\hat{L}_2) \sum_{n=0}^{\infty} \frac{1}{n!} (exp(-\hat{L}_2)\hat{L}_1 exp(\hat{L}_2))^n \\ &= exp(\hat{L}_2)exp(\hat{L}_3) \end{aligned} \tag{41}$$

where $\hat{L}_3$ is generated by the transformed function L_1

$$L_3 = exp(-\hat{L}_2)L_1. \tag{42}$$

This rule can be used to factorise an accelerator map into a linear and a nonlinear part. Imagine a sequence of 'linear' and 'nonlinear' elements in an accelerator lattice which are represented by linear maps $exp(\hat{R}_i)$ and nonlinear maps $exp(\hat{L}_i)$

$$\mathcal{M} = exp(\hat{R}_1)exp(\hat{F}_1).....exp(\hat{R}_{n-1})exp(\hat{F}_{n-1})exp(\hat{R}_n)exp(\hat{F}_n) \tag{43}$$

Applying the exchange rule to change the sequence of the first linear and second nonlinear transformation one obtains

$$\mathcal{M} = exp(\hat{R}_1)exp(\hat{F}_1).....exp(\hat{R}_{n-1})exp(\hat{R}_n)exp(\hat{N}_{n-1})exp(\hat{L}_n) \tag{44}$$

with $N_{n-1} = exp(-\hat{R}_n)L_{n-1}$, which means that the phase space coordinates $\vec{z}$ in the argument of L_{n-1} have to be replaced by the result of a linear transformation $\underline{R}_n^{-1}\vec{z}$ where the matrix $\underline{R}_1$ corresponds to $exp(\hat{R}_1)$. This procedure can be continued until one obtains

$$\mathcal{M} = exp(\hat{R})exp(\hat{N}_n).....exp(\hat{N}_2)exp(\hat{N}_1) \tag{45}$$

The transformation $exp(\hat{R})$ corresponds to the linear part of the transfer map of the lattice and the transformation $\prod_i exp(\hat{N}_i)$ corresponds to the nonlinear part.

Let us now turn to the symplectification of polynominal maps. A polynominal map $\mathcal{M}_N$ in the phase space coordinates z_i (N is the maximum order of the map) can be described by a Lie transformation of the form $\prod_{n=2}^{N+1} exp(\hat{L}_n)$ which agrees up to order N with $\mathcal{M}_N$. This has been shown by Dragt and Finn [15]. While $\mathcal{M}_N$ is in general not symplectic, the associated Lie transformation however is a symplectic map. For orders $m \leq N$ the polynominal map and the result of the Lie transformation agree, the Lie transformation contains higher order terms which make the transformation symplectic.

This can be seen in the following way. A map of an accelerator element or a whole lattice section may be given by a polynominal as a result of a Taylor expansion in the initial phase space coordinates z_i

$$\vec{z}_{final} = \mathcal{M}^N(z_i) = \sum_{j_1+..j_i+..j_n \leq N} \vec{a}_{j_1 \cdots j_n} \prod_i z_i^{j_i} \tag{46}$$

There is an important property of a Lie transformation which is generated by a homogeneous polynominal L_k of order k. If the corresponding transformation is applied to an initial point in phase space, the result is a map in these coordinates which extends in general to infinitely high order but the lowest-order terms are of order $k-1$.

Using this property one can find for any polynominal map, the corresponding Lie transformation order by order. Consider a polynominal map, a truncation of a symplectic map

$$\mathcal{M}^N = \sum_{k=1}^{N} \mathcal{M}_k \tag{47}$$

where the $\mathcal{M}_k$ are homogeneous polynominals of degree k in the phase space coordinates x_i, p_i, and consider also a Lie transformation $exp(\hat{L}_{N+1}) \cdot ... \cdot exp(\hat{L}_3)$ which is generated by homogeneous polynominals L_{k+1} of degree $k+1$ in x_i, p_i. Starting to iterate the lowest-order Lie transformation $exp(\hat{L}_{k+1}) = \sum_n \frac{1}{n!} \hat{L}^n_{k+1}$, the lowest-order terms of the corresponding map read

$$\begin{array}{l} x + \partial L_{k+1}/\partial p_i + higher\ order \\ p - \partial L_{k+1}/\partial x_i + higher\ order. \end{array} \tag{48}$$

The coefficients of L_{k+1} can then be found by comparison with the map $\mathcal{M}_k$.

Before one can proceed to the next order, the higher-order effects of $exp(\hat{L}_{k+1})$ have to be taken into account by modifying $\mathcal{M}^N$ to

$$\tilde{\mathcal{M}}^N = (exp(-\hat{L}_{k+1})\mathcal{M}^N)_N. \tag{49}$$

(The bracket with the index N means the expression is to be truncated beyond order N.) Now one can determine the coefficients of the next-higher-order generator L_{k+2} by comparison with the lowest order of the modified map $\tilde{M}^N$ namely $\tilde{\mathcal{M}}_{k+1}$.

This procedure involves carrying out the exponential operator series

$$exp(-\hat{L}_{k+1})...exp(\hat{L}_{k+n+1})$$

on the $n-th$ iteration until all terms of up to order $k+n+1$ have been generated. This can be conveniently performed by a computer.

As an illustration consider the symplectification of the following nonlinear map which includes terms up to third order

$$\begin{array}{rrl} \mathcal{M}^3(x,p): & x \rightarrow & x + x^2 + 2xp + p^2 + x^3 + x^2p + 2xp^2 + p^3 \\ & p \rightarrow & p + x^2 - 2xp - p^2 + x^3 + 5x^2p + 3xp^2 - 2/3p^3. \end{array} \tag{50}$$

We apply the first term of a Lie transformation generated by a homogeneous polynominal of degree 3

$$L_3 = h_{30}x^3 + h_{21}x^2p + h_{12}xp^2 + h_{03}p^3 \tag{51}$$

$$\begin{array}{rrl} x \rightarrow & x + \partial L_3/\partial p & = x + h_{21}x^2 + 2h_{12}xp + 3h_{03}p^2 \\ p \rightarrow & p - \partial L_3/\partial x & = p - 3h_{30}x^2 - 2h_{21}xp - h_{12}p^2. \end{array} \tag{52}$$

We determine the h_{ij} by comparison with the quadratic part of the map $\mathcal{M}_2$.

$$\begin{array}{ll} h_{30} = -1/3 & h_{21} = 1 \\ h_{12} = 1 & h_{03} = 1/3 \end{array} \tag{53}$$

so that

$$L_3 = -1/3 \cdot x^3 + x^2p + xp^2 + 1/3 \cdot p^3. \tag{54}$$

In the next step, we have to take into account in the original map the higher-order terms generated by $exp(\hat{L}_3)$.

$$\begin{array}{lll} \frac{1}{2}\hat{L}_3^2 x & = \frac{1}{2}(\frac{\partial L_3}{\partial p}\frac{\partial^2 L_3}{\partial p \partial x} - \frac{\partial L_3}{\partial x}\frac{\partial^2 L_3}{\partial p^2}) & = 2x^3 + 2x^2p \\ \frac{1}{2}\hat{L}_3^2 p & = -\frac{1}{2}(\frac{\partial L_3}{\partial p}\frac{\partial^2 L_3}{\partial x^2} - \frac{\partial L_3}{\partial x}\frac{\partial^2 L_3}{\partial x \partial p}) & = 2x^2p + 2xp^2. \end{array} \tag{55}$$

This modifies the cubic part of the map to:

$$\begin{array}{rrl} \tilde{\mathcal{M}}_3(x,p): & x \rightarrow & x - x^3 - x^2p + 2p^2x + p^3 \\ & p \rightarrow & p + x^3 + 3x^2p + p^2x - 2/3p^3 \end{array} \tag{56}$$

This modified map is represented by the lowest-order contribution of a map $exp(\hat{L}_4)$ with the generator

$$L_4 = -1/4 \cdot x^4 - x^3p - 1/2x^2p^2 + 2/3xp^3 + 1/4p^4 \tag{57}$$

Thus our original nonsymplectic polynominal map is represented by the symplectic Lie transformation $exp(\hat{L}_4)exp(\hat{L}_3)$.

We have learned so far how the tool of Lie transformations can be used to generate a symplectic map from a nonsymplectic polynomial map. If one wants to use Lie transformations however for tracking, one realizes that now we have to deal with infinite series of transformations which do not terminate. Truncating the map at a given order again results in a nonsymplectic map. In order to make the symplectification scheme useful, one has to find a way to make the Lie transformation executable without losing the symplectic property. This will be discussed in the next section.

5 EVALUATION OF SYMPLECTIC MAPS

Up to this point we know how to generate a truncated polynomial map by differential algebra techniques, and we know how to make this map symplectic by finding the associated Lie transformation. The next step is to turn the Lie transformation into a form which allows fast particle tracking to be performed.

Lie transformations are represented in general by infinite series. The evaluation of Lie transformations, i.e. actually performing the transformation, is a difficult task. Considerable effort has gone into the development of procedures to improve the performance of Lie transformations. For fast tracking the following procedure has been developed:

- First decompose the Lie generator into a sum of certain components S_n. The closed solution of the Lie transformation generated by each single component S_n is known. Such transformations are called a 'solvable' map.

- In the second step, the Lie transformation which is generated by the sum of solvable generators $exp(\sum_n \hat{S}_n)$ has to be factorized $\prod_n exp(\hat{S'}_n)$ in such a way that $exp(\hat{S'})$ is solvable and the map can then be carried out factor by factor.

It is easy to find a decomposition of a Lie generator $L_k = \sum_{n+m=k} h_{nm} x^n p^m$ into solvable parts. Each isolated monomial $h_{nm} x^n p^m$ corresponds to a solvable system.

$$\begin{aligned} & H = h_{nm} x^n p^m \\ & \dot{x}/(mx) = -\dot{p}/(np) \\ & x^n \cdot p^m = const = x_0^n p_0^m \\ & x(t) = x_0 \big(1 + (m-n) \cdot h_{nm} \cdot p_0^{(m-1)} x_0^{n(m-1)/(m-n)} \cdot t\big)^{m/(m-n)} \end{aligned} \tag{58}$$

or for $m = n$

$$x(t) = x_0 \cdot exp(n h_{nn} (x_0 p_0)^{n-1} \cdot t) \tag{59}$$

Therefore one might consider factorising the map in a straightforward manner by using the Campbell-Baker-Hausdorff formula

$$\begin{aligned} exp(\textstyle\sum_i \hat{S}_i) &= exp(\hat{S}_1) exp(\textstyle\sum_{i\geq 2} \hat{S}_i - \frac{1}{2}[\hat{S}_1, \hat{S}_i] + ...) \\ &\equiv exp(\hat{S}_1) exp(\textstyle\sum_{i\geq 2} \hat{G}_i) \\ &= exp(\hat{S}_1) exp(\hat{G}_2) exp(\textstyle\sum_{i\geq 3} \hat{G}_i - \frac{1}{2}[\hat{G}_2, \hat{G}_i] + ...) \\ &\equiv exp(\hat{S}_1) exp(\hat{G}_2) exp(\textstyle\sum_{i\geq 3} F_i) \\ &= exp(\hat{S}_1) exp(\hat{G}_2) exp(\hat{F}_3) exp(\textstyle\sum_{i\geq 4} \hat{F}_i) - \frac{1}{2}[\hat{F}_3, \hat{F}_i] + ...) \\ &. \\ &. \\ &. \end{aligned} \tag{60}$$

In principle one arrives that way at successive Lie transforms which are generated by monomials $S_1, G_2, F_3, ...$ respectively. The problem here is the huge number of terms one has to evaluate. A 10-th order map in 6 variables for example contains 12376 terms. For each of these terms a complicated expression of the type of equation (58) (generalized to three degrees of freedom) has to be evaluated. This would not be economic and map tracking would lose its advantage.

In order to make the concept of solvable maps feasible, one has to find another decomposition into solvable maps. There are two approaches which have been discussed, tested and made available recently. These are

- Kick factorization and
- Symmetric factorization.

We will first discuss the kick factorization.

5.1 Kick Factorization

Sudden changes in the particle momenta are called kicks. Kicks are used to describe the impact of nonlinear forces in particle tracking (thin-lens approximation). There are special Lie transformations which generate just a kick. All generators of the form $K = \sum_n a_n x^n$ (or more generally $\sum_{\vec{n}} a_{\vec{n}} \prod_i x_i^{n_i}$) produce such transformations. These transformation series terminate after the first step:

$$\begin{aligned} \hat{K}f(x) &= \partial f/\partial x \cdot \partial K/\partial p - \partial f/\partial p \cdot \partial K/\partial x = 0 \\ \hat{K}p &= -\textstyle\sum_n n a_n x^{n-1} = g(x) \\ \hat{K}^2 p &= 0 \end{aligned} \tag{61}$$

A thin-lens kick from a sextupole magnet for example is described by a generator K

$$K = \frac{1}{6} m \cdot l \cdot (x^3 - 3xy^2) \rightarrow exp(\hat{K}) \begin{pmatrix} x \\ p_x \\ y \\ p_y \end{pmatrix} = \begin{pmatrix} x \\ p_x - ml/2(x^2 - y^2) \\ y \\ p_y + mlxy \end{pmatrix} \tag{62}$$

There are more general 'kicks'. A Lie transformation produced by a generator of the form

$$K = \sum_n A_n \cdot (c \cdot x + s \cdot p)^n \tag{63}$$

(where c and s are constants e.g. $s = sin\phi$, $c = cos\phi$) or, for more degrees of freedom

$$K = \sum_{\vec{n}} A_{\vec{n}} \cdot \prod_i (c_i \cdot x_i + s_i \cdot p_i)^{n_i} \tag{64}$$

terminates as well after the first iteration.

$$\begin{aligned} \hat{K}x = &\quad \textstyle\sum_n n \cdot s \cdot A_n (c \cdot x + s \cdot p)^{n-1} \\ \hat{K}p = &\quad \textstyle\sum_n c \cdot n \cdot A_n (c \cdot x + s \cdot p)^{n-1} \\ \hat{K}^2 x = &\quad \textstyle\sum_{n,n'} nn'(n-1) A_n A_{n'} (c \cdot x + s \cdot p)^{n+n'-3} (s \cdot s \cdot c - s \cdot c \cdot s) = 0 \\ \hat{K}^2 p = &\quad \textstyle\sum_{n,n'} nn'(n-1) A_n A_{n'} (c \cdot x + s \cdot p)^{n+n'-3} (c \cdot c \cdot s - c \cdot s \cdot c) = 0 \end{aligned} \tag{65}$$

Such transformations may be called 'generalized kicks'. J. Irwin has worked out a scheme to evaluate a Lie transformation which involves generalized kicks [17] which is called 'Kick factorization'.

The idea of the kick factorization is to express a given Lie transformation by a sequence of kicks.

$$exp(\hat{L}_1)....exp(\hat{L}_n) \rightarrow exp(\hat{K}_1)....exp(\hat{K}_m). \tag{66}$$

For a given order k one finds always a set of N kick generators $K_{lk} = A_{kl}(c_l \cdot x + s_l \cdot p)^k$ which satisfy

$$\sum_l^N A_{kl}(c_l \cdot x + s_l \cdot p)^k = \sum_{n+m=k} h_{nm} x^n y^m \tag{67}$$

for properly chosen c_l, s_l. Then the set of equations

$$h_{nm} = \sum_l \left(\begin{array}{c} k \\ m \end{array} \right) c_l^n s_l^m \cdot A_{kl} \tag{68}$$

can be solved . The h_{nm} and A_{kl} form vectors $\vec{h}$ and $\vec{A}$ and the coefficients of the A_{kl} form a matrix $\underline{B}$. Then the solution for the kick coefficients A_{kl} is

$$\vec{A} = (\underline{B}^T \cdot \underline{B})^{-1} \underline{B}^T \vec{h}. \tag{69}$$

A Lie transformation which is generated by a sum of generalized kick generators $\sum_l K_{lk}$ unfortunately does not terminate. One starts the procedure with the lowest order, say k. Having found the kick generators for the lowest order k which results in a Lie transformation of the form $exp(\sum_{l=1}^N \hat{K}_{kl})$ we must factorise the result. We may use the Campbell-Baker-Haussdorff formula introduced in the previous paragraph for this purpose (the inverse of equation (37))

$$\begin{aligned} &exp(\textstyle\sum_{l=1}^N \hat{K}_{lk}) = exp(\hat{K}_{1k}) exp(\hat{K}_{-1k}) \\ &K_{-1k} = \textstyle\sum_{l=2,3..} K_{lk} - 1/2[K_{1k}, \sum_{l=2,3..} K_{lk}] + 1/6[K_{1k}, [K_{1k}, \sum_{l=2,3..} K_{lk}]] + ... \end{aligned} \tag{70}$$

Proceeding like that, one arrives at

$$exp(\hat{K}_{1k}) exp(\hat{K}_{2k}) exp(\hat{K}_{3k}) .... exp(\hat{K}_{Nk}) \cdot exp(\frac{1}{2} \sum_m^N \sum_{l \le m} [\hat{K}_{lk}, \hat{K}_{mk}] + ...) \tag{71}$$

The last factor contains the higher orders which are a contribution to the higher order parts of the map and they have to be taken into account if one is proceeding to the next order kick factorization. Eventually we have represented the whole original map in the form of generalized kicks.

Kicks of different orders n, m but with the same coefficients s_l, c_l generated by $K_{lm} = A_{ml}(c_l x + s_l p)^m$ and $K_{lm} = A_{nl}(c_l x + s_l p)^n$ commute

$$[\hat{K}_{ln}, \hat{K}_{lm}] = 0. \tag{72}$$

Therefore the sequence of kicks of different orders can be simplified to a sum of kicks

$$\prod_n exp(\hat{K}_{ln}) = exp(\sum_n \hat{K}_{ln}) \equiv exp(\hat{K}_l) \tag{73}$$

The kick factorization does not change the original polynomial map up to its order. For the lowest order, k, this is quite obvious. The sequence of the lowest-order kicks produces a map of order up to $(k-1)^2$. The part of the lowest order $k-1$ is just the sum of the kicks. This sum is identical with the first iteration of the corresponding Lie transformation $\partial L_k / \partial p$ and this is identical with the lowest order of the original map $\mathcal{M}$

$$\begin{aligned} \textstyle\prod_l exp(\hat{K}_{kl}) x \quad &= \textstyle\sum_l A_{kl} \cdot k \cdot s_l \cdot (c_l x + s_l p)^{k-1} + higher\ order \\ &= \textstyle\sum_{n+m=k} m h_{nm} x^n p^{m-1} + higher\ order \\ &= \textstyle\sum_{n+m=k-1} a_{nm} x^n p^m + higher\ order \end{aligned} \tag{74}$$

The kick factorization reproduces the original polynomial map but it adds higher order terms. The additional terms depend on the choice of the factors c_l, s_l. It is unfortunately very difficult to choose these coefficients in order to minimize the unwanted higher-order terms, which are the byproduct of the symplectification procedure. Therefore we now consider a procedure which provides better control of the higher-order terms.

5.2 Symmetric Factorization

Symmetric factorization is a symplectic approximation of a Lie transformation. It is a useful algorithm in cases where one can decompose a Lie generator into a relatively small number of solvable parts. Yan and Shi [18] have demonstrated that one can cut down the number of solvable parts considerably using the fact that generators which are the product of terms $g(x_i, p_i) \times f(x_j, p_j)...$ are solvable when f or g are either quadratic in the variables or $f,\ g$ are only functions in only one degree of freedom, or $f,\ g$ depend only on action variables $J_i = x_i^2 + p_i^2$ or if $f,\ g$ are kicks. Using these properties, one can reduce the number of solvable constituents of a map considerably. The 56 monomials of a cubic polynomial in six variables can for example be written as a sum of only 8 solvable constituents. A listing of the decomposition of the generators up to order six and the corresponding solutions are contained in reference [18].

Having reduced the number of solvable parts, the following symmetric factorization procedure is useful. Consider a map $exp(\hat{S} + \hat{G})$ where $\hat{S}$ and $\hat{G}$ constitute solvable maps respectively. We can approximate this map by

$$exp(\hat{S} + \hat{G})) \simeq exp(\frac{1}{2}\hat{S})exp(\hat{G})exp(\frac{1}{2}\hat{S}) \tag{75}$$

If one concatenates the r.h.s. of equation (77) using equation (40) the quadratic contributions in S and G given by the commutator terms $[\frac{1}{2}\hat{S}, \hat{G}]$ and $[\hat{G}, \frac{1}{2}\hat{S}]$ cancel and only third- and higher-order terms in $\hat{S}$) and $\hat{G}$ appear. The factors $\frac{1}{2}$ of $\hat{S}$ are universal and do not depend on the choice of S or G. One can find a symmetric factorization also for higher orders. For the next order in the factorization procedure one finds

$$\begin{array}{l} exp(\hat{S} + \hat{G})) \\ \simeq \\ exp(\alpha_1\hat{S})exp(\beta_1\hat{G})exp(\alpha_2\hat{S})exp(\beta_2\hat{G})exp(\alpha_2\hat{S})exp(\beta_1\hat{G})exp(\alpha_1\hat{S}) \end{array} \tag{76}$$

In this case, the concatenation of the r.h.s. produces, besides the expression on the l.h.s, only terms of order seven and higher in S and G. The coefficients α, β are universal as well and do not depend on S or G but only on the order of the factorization. In order to illustrate the kind of approximation let us consider the tracking through a long sextupole. Note that this problem can also be solved exactly but this example gives a hint as to the physical meaning of the approximation we are discussing.

$$H = \frac{1}{2}(p^2 + \frac{1}{3}mx^3)$$

is the Hamiltonian of the system.

$$L = \frac{1}{2} \cdot l \cdot (p^2 + \frac{1}{3}mx^3)$$

is the associated Lie generator, which is split into two easily solvable parts

$$L_1 = \frac{1}{2}lp^2; \quad L_2 = \frac{1}{6}mlx^3.$$

First-order symmetric factorization results in

$$exp(\hat{L}) \simeq exp(\frac{1}{4}l\hat{p^2})exp(\frac{1}{6}ml\hat{x^3})exp(\frac{1}{4}l\hat{p^2}) \tag{77}$$

We see that we have approximated the long sextupole by a drift space of length $l/2$ a sextupole kick $\frac{1}{2}mx^2$ and another drift of $l/2$. In this case, the first order symmetric factorization corresponds to the thin lens approximation. If we would proceed to the next order, this would correspond to a numerical integration according to Simpson's rule.

The procedures discussed here have been applied with great success for particle tracking for the SSC [3].

6 CONCLUSION

The methods of Differential Algebra and Lie Algebra helped to make fast particle tracking with truncated polynomial maps feasible. So far, good results have been obtained when the methods have been tested in particle tracking for the SSC, LHC and HERA [3] [4].

However map tracking is far from being completely explored or exploited. In particular, it could be useful to apply maps in conjunction with additional effects which affect the stability of particle motion in nonlinear fields. One example is the influence of tune modulation, which requires a large number of turns to be tracked. Another interesting application would be to include the particle spin in the tracking to investigate depolarisation effects in polarized beams [19]. The tools to attack these interesting and relevant issues are available now.

7 ACKNOWLEDGEMENTS

The author is indebted to Drs. D. Barber and G. Hoffstätter for enlightening discussions and for carefully reading the manuscript.

References

[1] F. Willeke, Analysis of Particle Tracking, Proceedings of the CAS Third Advanced Accelerator Physics Course, Uppsala 1989, CERN 90-04 (1990)

[2] Y. Yan, Success in One-Turn Maps for Dynamic Aperture Studies, Proceedings of the Workshop on Stability of Particle Motion in Storage Rings, Upton, N.Y. (1992), p 177

[3] Y.T. Yan, P.J. Channell, M. Li, M.J. Syphers, Long Term Tracking with Symplectic Implicit One Turn Maps, SSCL-PREPRINT-452 (1993)

[4] R. Kleiss, F. Schmidt, Y. Yan and F. Zimmermann, On the Feasibility of Tracking with Differential Algebra Maps in Long Term Stability Studies for Large Hadron Colliders, DESY-HERA 92-01, CERN SL/9202 SSCL-564 (1992)

[5] A. Dragt, Lectures on Nonlinear Orbital Dynamics, AIP Conference Proceedings, vol 87 (1982)

[6] A. Dragt, E. Forest, Computation of Nonlinear Behavior of Hamilton Systems Using Lie Algebra Methods, J. Math. Phys. 24(12), (1983), p 2734

[7] A. Dragt, F. Neri, G. Rangarajan, D.R. Douglas, L.M. Healy, R.D. Ryne, Lie Algebraic Treatment of Linear and Nonlinear Beam Dynamics, Ann. Rev. Nucl. Particle Sci. (1988) 38, p 455-495

[8] M. Berz, Differential Algebra Description of Beam Dynamics to Very High Orders, SSC-152 (1988)

[9] M. Berz, A Survey of Differential Algebra and its Use for the Extraction of Maps to Arbitrary Order, SSC-166 (1988)

[10] L. Michelotti, Differential Algebra without Differentials, Proceedings of the US Particle Accelerator Conference, Chicago (1989)

[11] L. Michelotti, MXYZPTLK: A Practical, User-Friendly C++ Implementation of Differential Algebra: User Guide, Fermilab FN-535 (1990)

[12] L. Michelotti, A Differential Algebra Package optimized for Beam Dynamics Applications, in preparation, FNAL

[13] L. Schachinger, R. Talman, Teapot: A Thin-Element Accelerator Program for Optics and Tracking, Particle Accelerators, Vol 22 pp 35-56 (1987)

[14] F. Schmidt, Tracking Results with SIXTRACK for the 1988 Dynamical Aperture Experiment, CERN SPS/88-49 (AMS) (1988)

[15] A.J. Dragt and J.M. Finn, Lie Series and Invariant Functions for Analytic Symplectic Maps, J. Math. Phys. 17,2215 (1976)

[16] A.J. Dragt, L.M. Healy, F. Neri, R.D. Ryne, D.R. Douglas, E. Forest, MARILIE 3.0, A Program for Nonlinear Analysis of Accelerator and Beamline Lattices, IEEE Trans Nucl. Sci, Vol NS-32 No 5, (1985)

[17] J. Irwin, A Multi-Kick Factorization Algorithm for Nonlinear Maps, SSC-228 (1989)

[18] Ji-cong Shi and Y.T. Yan, Symplectic Integrable Polynomial Factorization for Symplectic One-Turn Map Tracking, SSCL-PREPRINT-452 (1993)

[19] M. Berz, private communication

LONGITUDINAL BEAM DYNAMICS

W. Pirkl
CERN, Geneva, Switzerland

Abstract
An overview of longitudinal beam dynamics is given based on a Hamiltonian treatment of longitudinal motion. Emphasis is put on the conditions in circular accelerators; the differences with respect to linear accelerators are briefly mentioned.

1. ENERGY GAIN OF PARTICLES

1.1 Acceleration by electric fields and transit time factor

Acceleration of particles by electric fields is the predominant method to achieve the required energy gain.

Continuous DC electric fields are used in "electrostatic" (pre)accelerators. Insulation becomes a major problem since it is necessary to provide an electrode capable of holding a voltage equal to the total kinetic energy of the particle. The maximum kinetic particle energy is therefore limited to about 20 Mega-eV, even if tricks are applied such as charge inversion in the "Tandem" configuration.

Conversely the use of RF fields permits the adding of an arbitrary number of accelerating gaps to act on the beam, without summing the total voltage on an electrode. This is possible due to the "skin effect" (separating the inside volume of RF resonators from the outside world) together with the "cut-off frequency" of transmission in the vacuum chamber (separating the individual RF cavities). The electric field is no longer continuous but becomes sinusoidal with alternating half-periods of acceleration and deceleration. As a consequence the beam becomes structured in individual bunches, also the energy gain is reduced by the so-called transit time factor. Large particle accelerators with end energies up to the TeV range rely exclusively on this method of energy transfer to the beam.

Consider a typical accelerating gap and related electric field strength at one particular moment in time (Fig. 1). Far away from the axis the cavity geometry corresponds to a flat disk capacitor with uniform field strength $E_z(z)$ as depicted by the rectangular curve on top of the right side. The field pattern becomes bell-shaped as the axis is approached where the field penetrates partially into the vacuum chamber (bottom). These patterns are valid for a static field or for a snapshot within the RF cycle.

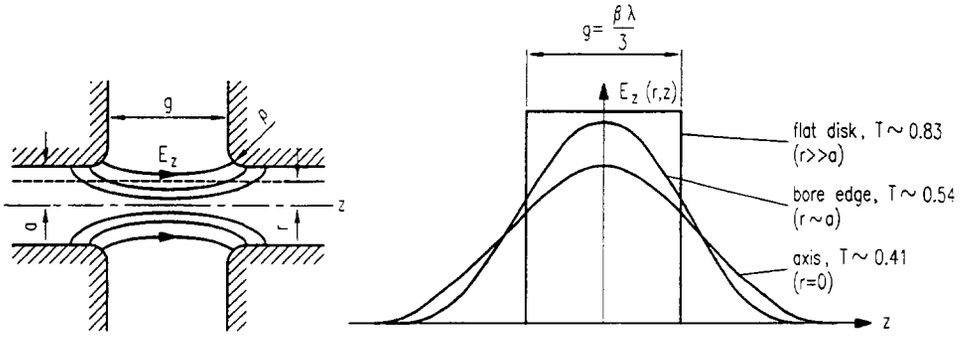
The energy gain of a particle of charge q traversing the gap is in general

$$\Delta E(r,t) = q \cdot \int E_z(r,z,t)dz$$

with

$$E_z(r,z,t) = E_z(r,z) \cdot \sin(\omega_{RF} + \phi)$$

i.e. the field pattern $E_z(r,z)$ discussed above is as a whole multiplied by a sinusoidal time dependence. Zero time corresponds by definition to the instant where the particle is situated in the mid-gap plane.



a) General geometry

b) Field pattern and transit time factor T for $g = \beta\lambda/3$

Fig. 1 Electric field in accelerating gap (adapted from [1])

If the particle spends an infinitely small fraction of the RF cycle time within the field region, the time factor $\sin(\omega_{RF} + \phi)$ is everywhere uniform and can, as a constant factor, be taken out and put in front of the integral. Under this condition it makes no difference whether the field pattern is rectangular or bell shaped as long as the area under the curve (i.e. the gap or cavity voltage $V_{cav} = \int E_z\, dz$) is the same. The energy gain can then be maximized by choosing the moment of maximum field for the passage of the particle, i.e. for $\phi = 90°$ (or $\pi/2$, respectively).

For a beam passage time that is not infinitely short with respect to the RF cycle the beam "sees" an instantaneous field whose local pattern is multiplied by a varying time factor (Fig. 2). The energy gain is then a function of the particle velocity; its value expressed in eV is certainly smaller than the achievable maximum (i.e. the gap voltage) since the full RF amplitude can only be seen at one particular moment and not during the whole passage through the gap.

This effect is usually described by the transit time factor which is defined as

$$T = \frac{\text{energy gain of particle with velocity } v = \beta c}{\text{maximum energy gain (particle with velocity } v \to \infty)}$$

It can be shown that T does not depend on the instant at which a particle crosses this gap. In other words the relative loss in acceleration voltage is the same for any RF phase ϕ.

For the simple case of uniform electric field across the gap and under the assumption that the particle velocity is constant the transit time factor is easily derived analytically as

$$T = \frac{\sin\left(\frac{\omega}{v}\frac{g}{2}\right)}{\left(\frac{\omega}{v}\frac{g}{2}\right)} = \frac{\sin\left(\pi\frac{g}{\beta\lambda}\right)}{\left(\pi\frac{g}{\beta\lambda}\right)} \quad \text{with} \quad \frac{\omega}{2v} = \frac{2\pi f}{2\beta c} = \frac{\pi}{\beta\lambda}$$

The shorter the gap length g with respect to the beam wavelength $\beta\lambda$ (i.e. the distance the beam travels during one RF period) the higher the transit time factor. Note that the exposure to the RF field is shortest at the bore edge. On-axis particles encounter longer effective gap length, their transit time factor and energy gain is therefore smaller than for peripheral particles. This can lead to beam dilution and synchrotron-betatron coupling with subsequent instabilities if the beam occupies an excessive fraction of the gap diameter.

Approximations are in widespread use to express the transit time factor for real RF field patterns: the influence of the beam pipe and of the curvature radius at the bore edge are taken into account by a larger effective gap length, the radial field dependence by the Bessel function I_0. These methods are largely superseded by numerical integration of the fields on and off axis in computer programs like SUPERFISH or MAFIA that allow to calculate all relevant cavity and gap parameters with high precision.

It should be pointed out that the overall transit time factor of multigap structures depends critically on proper synchronization of the beam passage between individual gaps. On top of the transit time factor of the individual gaps there may be an additional reduction factor that takes synchronization errors into account. Figure 2 gives an example of optimum synchronization between two gaps whose spacing of $\beta\lambda/2$ corresponds exactly to the distance the particle runs during 1/2 RF period.

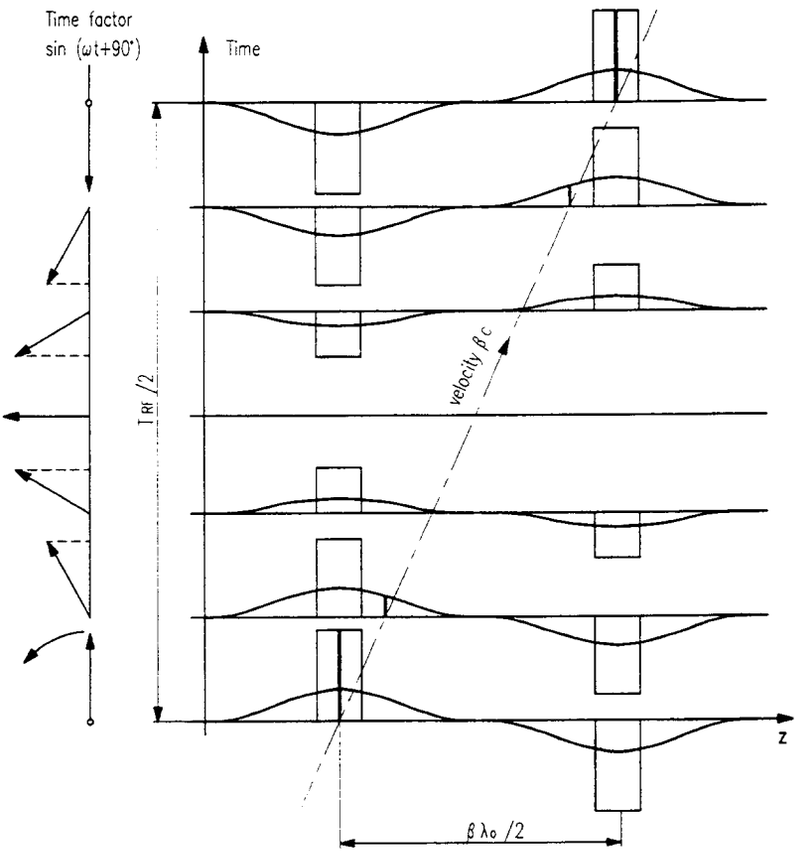


Fig. 2 Space-time diagram for a particle in a $\beta\lambda/2$ structure, for very short gaps (rectangular field pattern) and RFQ (sinusoidal field pattern).

The energy gain for a cavity with peak RF voltage V_{cav} and RF phase ϕ is given by the so-called "Panofsky" equation

$$\Delta E = q \cdot V_{cav} \cdot T \cdot \sin(\phi)$$

The transit time factor is normally of little concern in circular accelerators. It may become critical in Linacs, especially in bunches for low-velocity ion beams, where the inherent transit time factor can be as low as a few percent if no corrective action is taken.

1.2 Acceleration by magnetic fields: the "Betatron" voltage

Circular machines use a magnetic "bending" field to keep the particles on orbit. If the machine is an accelerator (as opposed to a storage ring at constant energy) there is a programmed increase in particle energy that has to be driven by a certain amount of RF voltage V_{acc} presented to the particles during one turn. On the other hand the increasing particle momentum leads to increasing beam rigidity that has to be compensated by a rising magnetic field B. It can be shown that the rate of rise of B, commonly designated as Bdot, is related to the voltage V_{acc} by the relation

$$V_{acc} = B\text{dot} \cdot 2 \cdot \pi \cdot R \cdot \rho$$

where R represents the average radius of the orbit and ρ the magnetic bending radius ($\rho <= R$ because of a magnet filling factor (magnet length)/(lattice length) of the machine).

Bdot will in turn induce along the particle orbit a "betatron" voltage V_b that has a tendency to accelerate the beam:

$$V_b = B\text{dot} \cdot 2 \cdot \pi \cdot R \cdot b$$

where b is the average equivalent magnet thickness. The fictive area $2 \cdot \pi \cdot R \cdot b$ would enclose, for the nominal magnetic induction at nominal orbit, the same total flux as generated by the real magnet geometry. A precise derivation has to take orbit variations for different energies into account as presented in [2].

It follows that

$$V_b / V_{acc} \cong b / \rho$$

The length b is approximately one half of the magnet dimension in transverse direction, further reduced by the magnet filling factor, i.e. of the order of less than one metre. On the other hand ρ may be as large as hundreds of meters or even kilometres. In conclusion the betatron voltage can be safely neglected in comparison with the accelerating voltage (and even more in comparison with the peak RF voltage per turn) for all but very small machines.

Historically the acceleration by varying magnetic fields was actually used in the (fairly small) "Betatron" machines from where the name was adopted.

1.3 Overall accelerating voltage per turn

For circular machines all parameters are usually referred to a complete turn. In the presence of N cavities delivering V_{cav} each the total RF voltage per turn is therefore

$$V_t = V_{cav} \cdot N \cdot T + V_b$$

The second term can usually be neglected.

On the other hand there are effects like beam loading and space charge that tend to reduce the apparent available RF voltage. It is supposed that these influences are also kept negligibly small.

Unless otherwise noted, it is assumed in what follows that the RF voltage can be represented by a single sinusoid of amplitude V_t. This is a realistic assumption for the majority of practical situations and allows a transparent presentation.

2. DIFFERENTIAL EQUATIONS OF BEAM MOTION

2.1 Variables

The beam motion is in this context represented by a single particle whose dynamics are, in good approximation, described by a second-order non-linear differential equation. As usual this equation can be broken down into two equations of order one for further treatment.

There is a range of possibilities for the selection of the two variables. To apply Hamiltonian treatment the two variables have to be canonically related which restricts but not determines the choice; a thorough discussion and derivation can be found in Ref. [3].

What follows is based on the set

$$\left(\frac{\Delta E}{h\omega_0}\right), \quad (\phi - \phi_s)$$

i.e. on the one hand the particle energy deviation ΔE divided by the RF harmonic number h and angular revolution frequency ω_0 , on the other hand the particle phase ϕ. Both quantities are referred to the synchronous particle, i.e. the reference particle of nominal energy and position for which the basic machine parameters are designed and orbit equilibrium conditions satisfied.

2.2 Time derivative of the particle energy

The phase of the particle is defined by

$$\phi = \int \omega_{RF} \; dt - h \cdot \Theta \approx h \cdot \omega_0 \cdot t - h \cdot \Theta$$

where Θ stands for the azimuthal position of the particle with respect to some geometric datum point, h for the RF harmonic number. It is well known that the RF system usually works at a multiple h of the revolution frequency, e.g. to allow more than one single bunch to be simultaneously on orbit, or to implement the best compromise between beam dynamics and RF hardware.

Note that the phases are referred to the RF frequency of h times the angular revolution frequency ω_0 of the particles rather than to ω_0 itself. A particle with nominal revolution time not subject to synchrotron oscillations keeps the same phase ϕ everywhere on the orbit. Inversely the quantity $\int \omega_0 \, dt$ can be thought of as a rotating phasor whose passage at some azimuth in the machine marks the positive zero crossing of the RF sinusoid at that spot and allows to calibrate the particle's time of passage in terms of a phase at ω_{RF}.

The energy gain accumulated during one turn by a particle of charge q and phase ϕ is

$$\Delta E_t = V_t \cdot q \cdot \sin\phi$$

Under the assumption that this process happens smoothly and continuously over one turn one can write

$$\frac{dE}{dt} = \Delta E_t \cdot f_0 = q \cdot U_t \cdot \sin\phi \cdot f_0$$

Referring to $h\,\omega_0$ and forming the difference to the synchronous particle of phase ϕ_S leads to

$$\frac{d}{dt}\left(\frac{E-E_s}{h\omega_0}\right)=\frac{q\ V_t}{2\pi}(\sin\phi-\sin\phi_s)$$

and finally

$$\frac{d}{dt}\left(\frac{\Delta E}{h\omega_0}\right)=\frac{q\ V_t}{2\pi}(\sin\phi-\sin\phi_s) \tag{2.1}$$

Remarks:

* It is important to include the possible variation of ω_0 in the time derivative (i.e. include it in the left-hand bracket), rather than treat it as a constant (and place it outside that bracket). Different orbits lead on the one hand to different energy gains due to the different betatron voltage, on the other hand to different revolution times. It is demonstrated in [1] that these two effects precisely cancel if the ratio of $\Delta E/\omega_0$ is considered. This specific result is a direct confirmation of the general arguments concerning the choice of canonically related variables discussed in [2].

* For ion accelerators where the charge-to-mass ratio is in general different from the proton there are two possibilities to express masses, energies and charges:

- Either they are referred to the full ion, the "bullet" consisting of a certain number of unit masses combined with some other number of unit charges (this leads to high figures for the particle energy),
- or everything is referred to the mass and rest energy of a single nucleon, with proportionally reduced charge that becomes smaller than unity (all figures then stated as "per nucleon"). The latter approach is generally preferred, although some confusion may occasionally occur when operating computer programs based on the other system.

2.3 Time derivative of the particle phase

The aim is to evaluate the phase change of a particle as a function of energy. Consider first the change in particle travel time τ as a function of the orbit length C and velocity v

$$\frac{d\tau}{\tau}=\frac{dC}{C}-\frac{dv}{v}$$

In an accelerator the change of orbit length C as a function of particle momentum p is an important characteristic of the magnetic lattice known as "momentum compaction factor" α_p.

$$\frac{dC}{C}=\alpha_p\frac{dp}{p}$$

Relativistic kinematics yield relations between the derivatives of p (particle normalized momentum), γ (particle nomalized energy) and β (particle normalized velocity)

$$\frac{dp}{p}=\frac{1}{\beta^2}\frac{d\gamma}{\gamma}\quad\text{and}\quad\frac{d\beta}{\beta}=\frac{1}{\beta^2\gamma^2}\cdot\frac{d\gamma}{\gamma}$$

Substituting them into the above formula leads to

$$\frac{d\tau}{\tau} = \alpha_p \frac{dp}{p} - \frac{d\beta}{\beta} = \frac{1}{\beta^2}\underbrace{\left(\alpha_p - \frac{1}{\gamma^2}\right)}_{-\eta}\frac{d\gamma}{\gamma}$$

$$\frac{d\tau}{\tau} = -\frac{\eta}{\beta^2}\frac{d\gamma}{\gamma}$$

The term η is sometimes also referred to as "phase slip factor".

On the other hand the relation between particle phase and travel time referred to one turn reads

$$\frac{d\phi}{\phi} = \frac{d\tau}{\tau}, \quad \text{or} \quad d\phi = 2\pi h \frac{d\tau}{\tau}$$

Combining yields

$$d\phi = -2\pi h \frac{\eta}{\beta^2}\frac{d\gamma}{\gamma}$$

Assuming that the process is smooth and uniform during one turn leads to

$$\frac{d\phi}{dt} = d\phi \cdot f_0 = -\omega_0 h \frac{\eta}{\beta^2}\frac{d\gamma}{\gamma} = -\omega_0 h \frac{\eta}{\beta^2}\frac{\Delta E}{E}$$

Referring the deviation to the synchronous particle of phase ϕ_S and energy E_S

$$\frac{d}{dt}(\phi - \phi_s) = -\omega_0 h \frac{\eta}{\beta^2 E}(E - E_s)$$

Expanding the right side by $h\omega_0$ for normalization leads to the final result

$$\frac{d}{dt}(\phi - \phi_s) = -\omega_0^2 h^2 \frac{\eta}{\beta^2 E}\left(\frac{\Delta E}{h\omega_0}\right) \tag{2.2}$$

Remarks:

- According to the above definition, the phase slip factor η starts at a value of $+(1 - \alpha p)$ for low energies ($\gamma \sim 1$) and tends to a value of $-\alpha p$ for very high energies ($\gamma >> 1$). The energy at which η necessarily crosses zero is called "transition energy γt" and the simple relation holds

  $$\eta = 0 = \frac{1}{\gamma_{tr}^2} - \alpha_p \quad \text{or} \quad \alpha_p = \frac{1}{\gamma_{tr}} \quad \text{hence} \quad \eta = \frac{1}{\gamma^2} - \frac{1}{\gamma_{tr}^2}$$

  (This does not imply that one and the same γ_t is valid over the full energy range; α_p and γ_{tr} may well vary over the machine cycle).

- η is frequently defined with the opposite sign, i.e. as negative below transition and positive above. That is most often the case in the context of large machines whose operating range is exclusively above transition.

2.4 Hamiltonian

Equations (2.1) and (2.2) describe the motion of a single particle in the phase space; they are repeated below for convenience.

$$\frac{d}{dt}\left(\frac{\Delta E}{h\omega_0}\right) = \frac{q\ V_t}{2\pi}(\sin\phi - \sin\phi_s)$$

$$\frac{d}{dt}(\phi - \phi_s) = -\omega_0^2 h^2 \frac{\eta}{\beta^2 E}\left(\frac{\Delta E}{h\omega_0}\right)$$

Because of the various simplifications made the accuracy of these equations is limited but normally sufficient for at least short-term simulations and predictions.

The two time derivatives may of course be combined into one single differential equation of second order but this formal operation does not facilitate the numerical integration nor significantly enhance the physical insight. On the other hand forming the Hamiltonian allows to invoke the principles of preservation of phase space and of energy which opens the door for very illustrative interpretations.

The function H assigned to equations (2.1) and (2.2) reads

$$H = \frac{1}{2}h^2\omega_0^2\frac{\eta}{\beta^2 E}\left(\frac{\Delta E}{h\omega_0}\right)^2 + \frac{gV_t}{2\pi h}\left(\underbrace{-\cos\phi - \phi\sin\phi_s}_{A} + \underbrace{\cos\phi_s + \phi_s\sin\phi_s}_{B}\right) \qquad (2.3)$$

Part A is the direct integration of the bracket $(\sin\varphi - \sin\varphi_s)a\varphi$ of Eq. (2.1), part B the same expression evaluated for $\phi = \phi_S$ i.e. a constant. It brings the Hamiltonian to zero for a particle with energy deviation $\Delta E = 0$ at phase ϕ_s. This is the equilibrium position, both time derivatives vanish at this point, the synchronous particle remains therefore at rest.

The canonically conjugate variables x,y of a Hamiltonian H have to satisfy the relations

$$\frac{\partial H}{\partial x} = -\frac{dy}{dt} \qquad \frac{\partial H}{\partial y} = \frac{dx}{dt}$$

This condition is met for the above set as can be seen by differentiation

$$\frac{\partial H}{\partial\left(\frac{\Delta E}{h\omega_0}\right)} = h^2\omega_0^2\frac{\eta}{\beta^2 E}\left(\frac{\Delta E}{h\omega_0}\right) = -\frac{d}{dt}(\phi)$$

$$\frac{\partial H}{\partial\phi} = \frac{g\ V_t}{2\pi}(\sin\phi - \sin\phi_s) = \frac{d}{dt}\left(\frac{\Delta E}{h\omega_0}\right)$$

In addition, the dimension of an area element in phase space defined by the two variables is [energy/(1/time)]·[1] = [energy·time] i.e. an "action" so that all formal requirements concerning a Hamiltonian are satisfied. Strictly speaking this is in essence an *a posteriori* justification for the legitimate choice of these two variables.

The expression for the Hamiltonian can be simplified

$$H = \frac{1}{2}\frac{\eta}{\beta^2 E}(\Delta E)^2 + \frac{q\ V_t}{2\pi h}\left(\underbrace{\sin\phi_s(\phi_s - \phi) + \cos\phi_s - \cos\phi}_{U(\phi)}\right) \tag{2.4}$$

The two terms of this Hamiltonian are of the same structure as those of a simple harmonic oscillator like a pendulum:

- the first term is proportional to the square of one canonical variable and can be assimilated to the kinetic energy $K = mv^2/2$ of the mechanical equivalent,
- the second term is a function of the second canonical variable and can be assimilated to the potential energy. It can be easily shown that $U(\phi)$ represents, apart from a scale factor, the energy needed to displace the particle of charge q from its equilibrium position ϕ_S to position ϕ against a restoring force $q \cdot U_t \cdot (\sin\varphi - \sin\varphi_s)$.

The Hamiltonian in this form is another expression for the total energy of the particle; in a conservative system this total energy has to remain constant. For a known initial position in phase space the Hamiltonian (or total energy) is a constant of motion that has to remain unchanged along a trajectory even though the individual amounts of kinetic and potential energy may be interchanged.

Although the factor $h\,\omega_0$ has now disappeared from the expression (2.4) it must be kept in mind that the canonically related energy variable is still $\Delta E/(h\,\omega_0)$ rather than the bare ΔE appearing in the formula.

Figure 3 gives a qualitative view of the relation between RF voltage (top), potential energy (middle) and phase space (bottom). The RF voltage is supposed to be a pure sinusoid, with a synchronous particle at phase ϕ_S. The corresponding RF voltage V_{acc} covers the energy per turn that has to be invested into the particle to follow the acceleration rate, and to replenish losses e.g. by synchrotron radiation. The synchronous particle is thus energetically at equilibrium.

As mentioned above the energy needed to displace the particle from the equilibrium position corresponds to a potential energy. This figure is in turn proportional to the area under the curve displaying the voltage difference V_{RF} - V_{acc} , its numerical value is plotted in the middle curve.

Suppose that a particle is displaced from equilibrium $(\phi = \phi_s;\ \Delta E = 0;\ H = 0)$ to the position $(\phi = \phi^*;\ \Delta E = 0;\ H = H^*)$. Released from there the particle follows a trajectory in phase space for which $H = H^*$.

$$H^* = \text{zero} + \frac{q\ V_t}{2\pi h} \cdot U(\phi^*)$$

At any phase ϕ the potential energy can be directly evaluated via $U(\phi)$; the difference with respect to H^* must correspond to the kinetic part, therefore

$$H = H^* = \frac{q\ V_t}{2\pi h} \cdot U(\phi^*) = \frac{1}{2}\frac{\eta}{\beta^2 E}(\Delta E)^2 + \frac{q\ V_t}{2\pi h} \cdot U(\phi)$$

$$\Delta E = \sqrt{\frac{\beta^2 E\ q\ V_t}{\eta \pi h}}\sqrt{U(\phi^*) - U(\phi)} \tag{2.5}$$

This is the general equation of a trajectory in phase space for a particle starting with initial phase deviation ϕ^*.

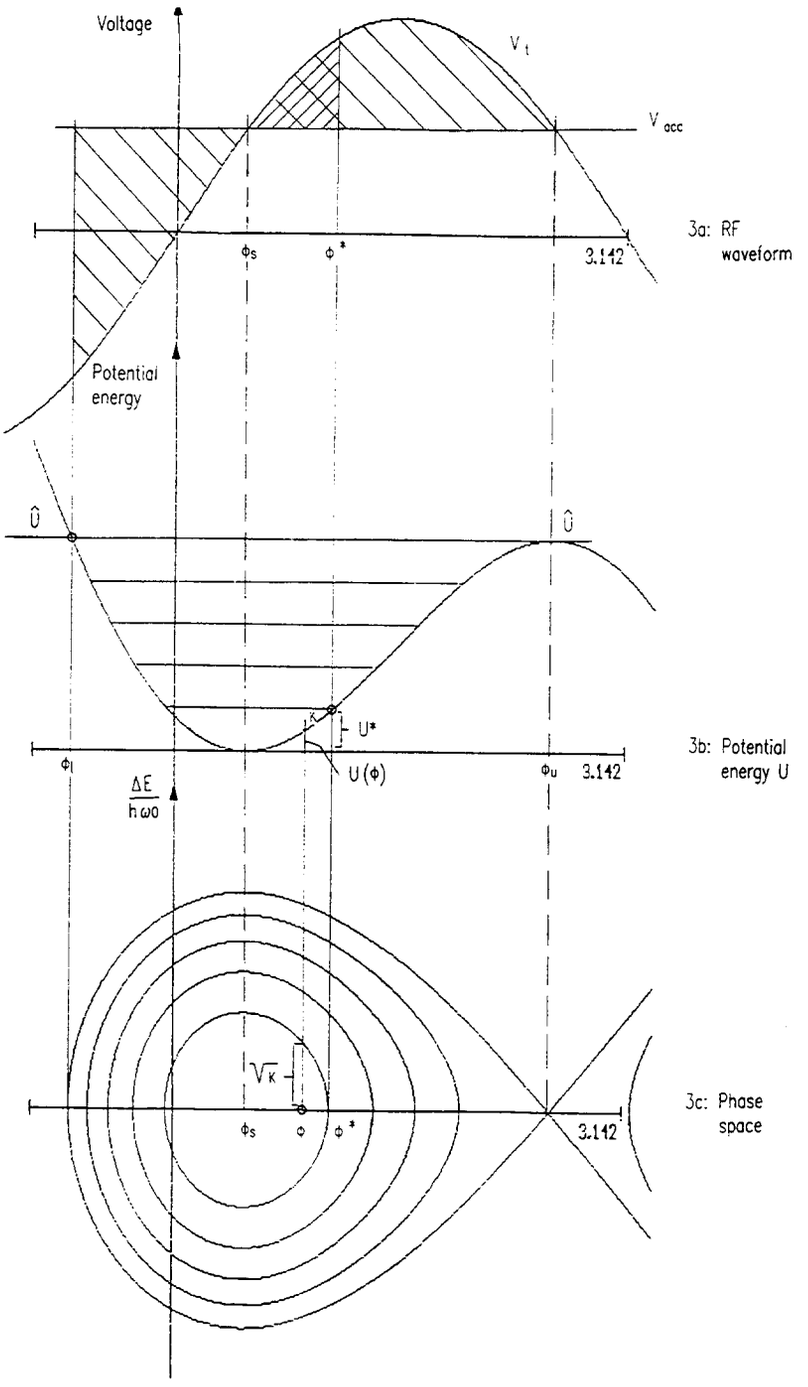


Fig. 3 RF voltage (a), potential function U (b), phase space (c)

A 3-D view of the Hamiltonian is given in Fig. 4 below. The bold trace (intersection of the Hamiltonian with plane $\Delta E = 0$) represents the potential energy of Fig. 3b. A trajectory in phase space corresponds to $H = \text{const}$. In other words, the trajectories are the contour lines of that surface (or rather the projections thereof to the horizontal $\phi - \Delta E$ plane).

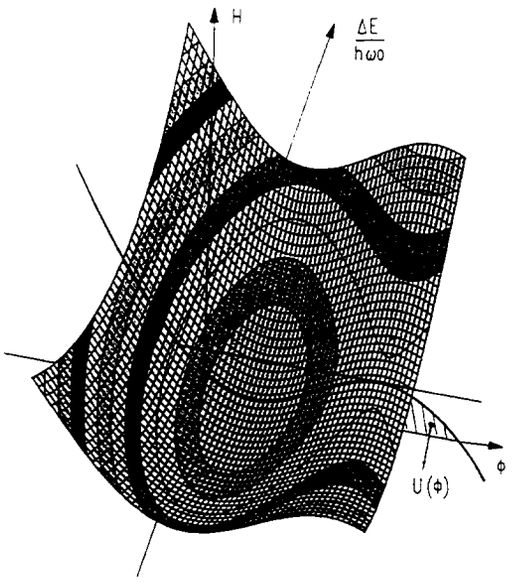


Fig. 4 3D-view of the Hamiltonian

There are two fixed or equilibrium points, namely at $(\phi = \phi_s,\ \Delta E = 0)$ and $(\phi = \pi - \phi_s,\ \Delta E = 0)$. At both points the time derivative of the two canonical variables is zero. This can be seen by evaluation of the differential equations (2.1) and (2.2), or by noting that the tangent plane to the Hamiltonian in 3D representation is horizontal. The former point is a true minimum, the contour lines around it are closed, it is a stable or "elliptical" fixed point. The latter, however, is a saddle point; the contour lines around it are not closed, it is therefore an unstable or "hyperbolic" fixed point.

3. STABILITY LIMITS: THE "(FULL) BUCKET"

One may ask how far a particle can be displaced from the stable phase ϕ_S and still remain linked to the synchronous particle. Looking at Fig. 3a it can be seen that the restoring force proportional to V_{RF} - V_{acc} changes sign at $\phi_u = \pi - \phi_s$ and that the potential function U is at its maximum value there. As explained above, this point is at unstable equilibrium; it represents one limit of the stable region that is commonly referred to as the "bucket".

If a particle is released from within the stable region in the immediate vicinity of this point it will first convert all energy from potential into kinetic on its way to ϕ_S, then recover all energy from kinetic back to potential on the way from ϕ_S to the other limiting point ϕ_1 where it comes again to rest. In other words the point of the lower limiting phase is characterized by the same potential energy as the known point of upper limiting phase. The stable region is therefore limited by the phases

$$\phi_u = \pi - \phi_s \qquad \text{(direct evaluation)}$$
$$U(\phi_1) = U(\phi_u) = \hat{U} \qquad \text{(implicit equation)}$$

The difference $\phi_u - \phi_1$ is called the "bucket width".

It has been mentioned that the potential energy for any phase can be represented by the area under the curve V_{RF} - V_{acc}. At the lower limiting phase this area is therefore the same as

the area at the upper limiting phase (see Fig. 3a). This allows a rapid qualitative evaluation of the position of ϕ_1.

The maximum energy deviation ΔE at phase $\phi = \phi_S$ is known as "bucket (half) height" and can be evaluated from Eq. (2.5). Since the potential function U is zero at the equilibrium phase it follows

$$\Delta\hat{E} = \sqrt{\frac{\beta^2 E \; q \; V_t}{\eta \pi h}} \cdot \sqrt{\hat{U}} = \sqrt{\frac{\beta^2 \gamma}{\eta} \cdot \frac{q \; V_t \cdot E_0}{\pi h}} \cdot Y(\phi_s) \tag{3.1}$$

with

$$\begin{aligned} Y(\phi_s) &= \sqrt{\sin\phi_s(2\phi_s - \pi) + 2\cos\phi_s} \; ; \\ Y(0) &= \sqrt{2} \end{aligned} \tag{3.2}$$

Bucket width and bucket half height are the extreme points in phase and in energy deviation that characterize the "separatrix", the closed curve that define the stable region or "bucket" where off-phase or off-energy particles rotate around the synchronous particle. Particles outside may vibrate between adjacent buckets for some time before they are definitely lost by contact with the vacuum chamber.

The equation of the separatrix between these extreme points is a special case of Eq. (2.5) where $U(\phi^*) = U(\phi_u) = \hat{U}$, or

$$\Delta E_x = \sqrt{\frac{\beta^2 \gamma}{\eta} \cdot \frac{q \; V_t E_0}{\pi h}} \sqrt{Y^2 - U(\phi)}$$

The index x refers to the separatrix.

The area A enclosed by the separatrix, known as "bucket area" for obvious reasons, has a special significance since it represents the longitudinal acceptance of a machine. As an area enclosed by a trajectory and expressed in canonical units $\left(\phi, \Delta E / (h \cdot \omega_0)\right)$ it should in principle be invariant (the apparent paradox that it changes nevertheless during acceleration will be discussed in section 4.6). Using the symmetry around the ϕ axis and converting into canonical units one has

$$A = 2 \int_{\phi_L}^{\phi_u} \frac{\Delta E_x}{h\omega_0} d\phi = A_y \cdot \alpha(\phi_s) \tag{3.3}$$

The expanded analytical expression for (3.3) is very complicated since the lower integration limit is an elaborate function of the stable phase angle ϕ_S. The area of a general bucket is therefore related to the area of the stationary bucket A_y which can be obtained analytically as explained below; the mediating scale factor $\alpha(\phi_S)$ is evaluated by numerical integration. The symbols Y and α are de facto reserved names in early litterature, together with $1' = \sin\phi_s$ which was generally used as the independent quantity. The functions $Y(\Gamma)$ and α (Γ) are widely tabulated; a transcript of $Y(\phi_S)$ and $\alpha(\phi_S)$ is given in annex 1.

The different systems of units used to express the bucket parameters deserve special mention. The canonically related parameters $(\Delta E/h\,\omega_0)$ for the energy deviation and ϕ for the longitudinal position are mandatory for theoretical considerations but not well suited for

practical applications. Machine related parameters like radial position or RF frequency shift are sometimes used instead of the energy variable. With the increasing importance of multiple ring machines which necessitate beam transfers between rings of radically different properties there is growing consensus to use machine independent parameters to characterize beams (the "bunches") and acceptances (the "buckets"). These units are time t for the abscissa and ΔE for the ordinate.

Note that the bucket area expressed in (t, ΔE) units is numerically and dimensionally the same as that expressed in the canonical units $\left(\phi, \Delta E / \left(h \cdot \omega_0\right)\right)$. The reason is that the conversion factor $(h \cdot \omega_0)$ is used for the abscissa and $1/(h \cdot \omega_0)$ for the ordinate such that the multiplication results in the same area expressed in [eV·sec].

3.1 The stationary bucket

A special case is a bucket whose stable phase angle ϕ_S is zero, i.e. a bucket whose acceleration voltage is zero. Its upper limiting phase is $\phi_u = 180° - 0° = +180°$ (or π) and by symmetry around the stable phase it follows immediately that the lower limiting phase $\phi_1 = -180°$ (or $-\pi$). Its separatrix is also symmetric about ϕ_s and has the shape of a cosine $\left(\text{since } \sqrt{Y^2 = U(\phi)} = \sqrt{0.5(1 - \cos(x))} = \cos(2 * x)\right)$. This fact allows to evaluate the stationary bucket area by inspection as

$$A_y = 4\left(\frac{\Delta\hat{E}_y}{h\omega_0} \cdot \phi_u\right) \cdot \frac{2}{\pi} = 8\frac{\Delta\hat{E}_y}{h\omega_0} = \frac{8\sqrt{2}}{h\omega_0}\sqrt{\frac{\beta^2\gamma}{\eta}\frac{q\ V_t E_0}{\pi h}} \tag{3.4}$$

with bucket half height $\Delta\hat{E}_y$

$$\Delta\hat{E}_y = \sqrt{\frac{\beta^2\gamma}{\eta}\frac{q\ V_t E_0}{\pi h}}\sqrt{2} \tag{3.5}$$

where the index y refers to the stationary case.

The stationary bucket has, for given machine parameters and RF voltage, the largest height and largest area. It is therefore often used in graphs and tables as a reference point for other stable phase angles. The fact that the bucket width extends over 360° or 2π permits (at least in theory and for ideally adiabatic processes) to capture continuous beams with 100% efficiency. This is the reason why the injection of a Linac into a synchrotron is normally performed at "flat bottom", i.e. a magnetic bending field that is constant in time. Then Bdot = $U_{acc} = 0$ allowing implementation of $\phi_S = 0$.

3.2 The partial bucket

All the above limiting quantities refer to the "full" bucket, i.e. to the full extent of the stable area. In practice the stable area is very rarely fully used in order to avoid particle losses. It is usually of more interest to know how far a beam of a given size (or area or longitudinal emittance) is away from these limits in order to evaluate the loss rate e.g. in a storage ring. It should be mentioned in this context that the "emittance" of a multi-particle beam is a statistical quantity and that the percentage of particles contained in a "limiting" contour depends strongly on their spatial distribution. The bucket limits in phase and in energy have usually to be three to four times the RMS dimensions of the beam to comprise some 98% of the total intensity. These

facts have to be kept in mind when speaking of the limits of a beam or of a bucket that are characterized by the orbit of a single particle.

The orbits of a particle around the stable phase and within the separatrix are closed curves for which the corresponding Hamiltonian is a constant of motion. Equation (2.5) describes directly the limiting trajectory whose upper limiting phase is ϕ^*. The lower limiting phase, maximum energy deviation and partial bucket area can be obtained from this formula in analogy to the procedure described for the full bucket.

Of particular interest is the region immediately around the point of stable equilibrium where the potential function for a small phase deviation Ψ can be approximated by

$$U(\Psi) = \Psi \sin\phi_s + \cos\phi_s - \cos(\phi_s + \Psi) \approx \frac{\cos\phi_s}{2}\Psi^2 + \ldots \tag{3.6}$$

There is obviously no linear term since the potential function has an extremum at ϕ_S. The corresponding peak energy deviation $\Delta\hat{E}_\Psi$ is then

$$\Delta\hat{E}_\Psi = \sqrt{\frac{\beta^2\gamma}{\eta}\frac{q\ V_t E_0}{\pi h}} \cdot \sqrt{\frac{\cos\phi_s}{2}}\ \Psi$$

or

$$\frac{\Delta\hat{E}_\Psi}{\Psi} = \sqrt{\frac{\beta^2\gamma}{\eta}\frac{qV_t E_0 \cdot \cos\phi_s}{2\pi h}} \tag{3.7}$$

This ratio is constant in the vicinity of the stable equilibrium point. It can be easily seen that in this region the energy spread of the partial bucket relative to the full bucket is

$$\frac{\Delta\hat{E}_\Psi}{\Delta\hat{E}} = \frac{\Psi}{Y}\sqrt{\frac{\cos\phi_s}{2}}$$

or in particular for the stationary case

$$\frac{\Delta\hat{E}_\Psi}{\Delta\hat{E}} = \frac{\Psi}{2}$$

In other words: in a partial stationary bucket the phase deviation 0.1 rad (or 5.7°) corresponds to an energy deviation of 0.05 (or 5%) of the full bucket height.

The dimensions of full and partial buckets are best evaluated numerically. Even a standard programmable pocket calculator, preferably with built-in root-finding and integration routines, can be coded to deliver fast and accurate results without the trouble and sources of error involved with the use of normalized graphs and tables.

4. DYNAMICS OF THE PARTICLE TRAJECTORY

4.1 Evolution of particle position with time

So far the trajectories in phase space have been described without consideration of time, i.e. without mention how far a particle will move from different positions in a given time interval. To obtain this information the time derivatives of the canonical variables have to be integrated by numerical methods. It is known beforehand that

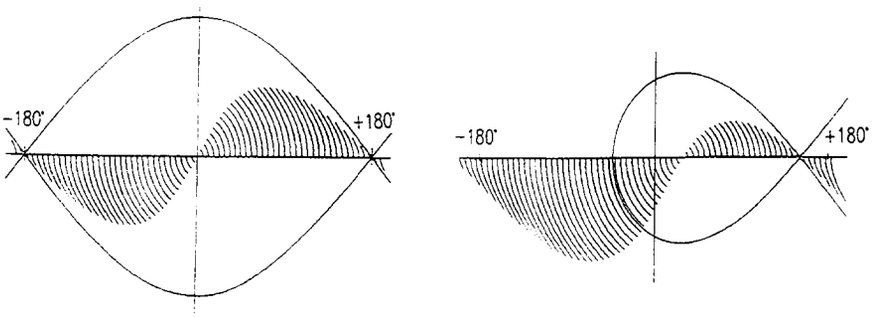
- the *direction* of motion follows a contour line of the Hamiltonian, i.e. proceeds orthogonally to its gradient,

- the *velocity* of motion $|ds/dt|$ in a phase space of coordinates p, g is proportional to the magnitude of the Hamiltonian's gradient at any given point since

$$\left|\frac{ds}{dt}\right| = \sqrt{\left(\frac{dp}{dt}\right)^2 + \left(\frac{dg}{dt}\right)^2} = \sqrt{\left(\frac{\partial H}{\partial g}\right)^2 + \left(-\frac{\partial H}{\partial p}\right)^2} = |\mathrm{grad}(H)|$$

Some examples of particle dynamics are given below. The plots have been obtained through numerical Runge-Kutta integration.

Figure 5 depicts the trajectory of particles during 1/8 of a "synchrotron" period after they started at the same time from different phases. The evolution of the particles is shown in a stationary and an accelerating bucket, for the same RF voltage per turn. In both cases the particles close to the stable phase rotate at sensibly the same angular frequency but slow gradually down when approaching the unstable fixed point. This phenomenon will be examined in more detail in section 4.2 below.

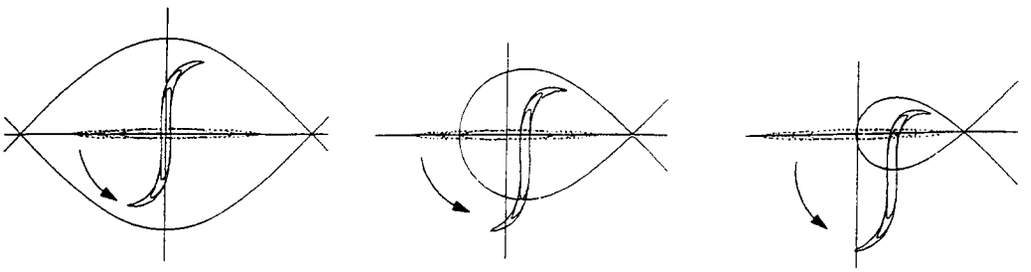


a) Stationary bucket b) Accelerating bucket

Fig. 5 Evolution of particle position

It can also be seen that the motion outside the separatrix is still orderly and not a priory a forbidden area for momentary beam gymnastics, provided that the bunch can be repositioned inside the separatrix before it migrates into neighbouring buckets.

In analogy to the transverse phase space the longitudinal properties of a beam are often idealized as an ellipse. Equations (2.1) together with (2.2) transform such ellipses into ellipses as long as the coefficients are linear, i.e. as long as the synchrotron frequency of all their constituents is uniform; otherwise a distortion of the beam boundaries occurs as illustrated in Fig. 6a for three different emittances. The initial beam is shown in dotted lines, the final beam in full lines. The process shown is the so-called "(non-adiabatic) bunch rotation" during 1/4 of a synchrotron period. It exchanges bunch length and energy spread and is often used in either direction. Figures 6b and 6c illustrate the surprising fact that bunch rotation can be carried out not only on stationary but also on accelerating buckets, with virtually the same distortion. This can be used to advantage in linear accelerating structures.



a) Stationary bucket b) Accelerating bucket $\phi_s = 0.4$ c) Accelerating bucket $\phi_s = 0.6$

Fig. 6 Bunch rotation

Another standard beam gymnastics procedure is shown in Fig. 7, namely bunch lengthening by evolution around the unstable fixed point. Note that the distortions occur in different directions which opens the possibility of some compensation if the two processes are carried out consecutively.

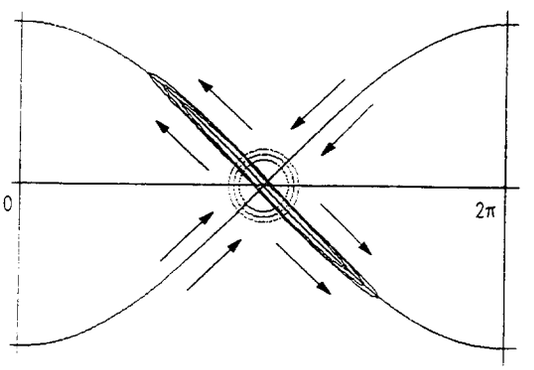


Fig. 7 Bunch elongation near unstable fixed point

4.2 Evaluation of the synchrotron frequency using the Hamiltonian

The frequency of revolution in phase space is known as the "synchrotron" frequency f_{syn}. For the evaluation of the time t needed for one full turn one starts with the equation obtained from simplification of (2.2)

$$\frac{d\phi}{dt} = -\omega_0 \cdot h \cdot \frac{\eta}{\beta^2 E} \Delta E$$

or

$$dt = -\frac{\beta^2 E}{\omega_0 h \eta} \cdot \frac{1}{\Delta E} \cdot d\phi$$

with

$$\Delta E = \sqrt{\frac{\beta^2 \gamma}{\eta} \frac{q \ V_t E_0}{\pi h}} \sqrt{Y^2 - U(\phi)}$$

The phase is chosen as the independent variable and the corresponding energy deviation obtained from the Hamiltonian, whose initial value Y^2 at the upper limit ϕ_u can be evaluated immediately through the potential energy function. The time needed for one full turn is then

$$\tau = 2\left(-\frac{\beta^2 E}{\omega_0 h\eta}\cdot\frac{1}{\Delta E}\right)\int_{\phi_e}^{\phi_u} d\phi = -2\frac{\beta}{\omega_0}\sqrt{\frac{\gamma}{\eta}\frac{E_0\cdot\pi}{hqV_t}}\cdot\int_{\phi_e}^{\phi_u}\frac{d\phi}{\sqrt{Y^2-U(\phi)}} \tag{4.1}$$

For the vicinity of the stable fixed point the integral reduces to

$$\int_{\phi_e}^{\phi_u}\frac{d\phi}{\sqrt{Y^2-U(\phi)}} = \int_{\phi_e}^{\phi_u}\frac{d\phi}{\sqrt{\frac{\cos\phi_s}{2}(\Psi^2-\phi^2)}} = \frac{1}{\sqrt{\frac{\cos\phi_s}{2}}}\int_{-1}^{1}\frac{dx}{\sqrt{1-x^2}} = -\pi\cdot\frac{1}{\sqrt{\frac{\cos\phi_s}{2}}}$$

Hence

$$\tau_0 = \frac{2\pi\beta}{\omega_0}\cdot\sqrt{\frac{\gamma}{\eta}\cdot\frac{2\pi E_0}{hqV_t\cos\phi_s}}\ ,$$

or

$$\frac{1}{\tau_0} = f_{\text{syn\_0}} = \frac{\omega_0}{2\pi\beta}\cdot\sqrt{\frac{\eta}{\gamma}\cdot\frac{hqV_t\cos\phi_s}{2\pi E_0}} \tag{4.2}$$

This synchrotron frequency $f_{\text{syn\_0}}$ is constant in the immediate vicinity of the stable fixed point. In regions where the approximation (3.6) does not hold the synchrotron frequency is in general lower. This can be implied from the fact that the ratio (energy deviation) /(phase deviation) decreases with distance from the stable phase point with the consequence that the revolution time according to equation (4.1) increases. The analytical expression becomes too complicated for direct interpretation and numerical evaluation is used instead.

Figure 8 gives the synchrotron frequency as a function of the phase deviation for a stationary bucket. The result of the approximation by the first terms of a truncated series given in [1] is shown for comparison. Note that the synchrotron frequency tends to zero if the particle approaches the separatrix.

To further investigate this effect, the vector field in the immediate vicinity of the two fixed points is shown in Fig. 9. The phase derivative (upper formula) is obtained directly from Eq. (2.2); the energy derivative (lower formula) by linearization of Eq. (2.1) or by differentiation of (3.6). The fact that the sign of the latter is different for the two fixed points leads to radically different particle orbits:

- around the *stable* fixed point (left) the orbits are closed locally. Both the orbit lengths and the orbit velocities decrease in proportion when approaching the fixed point, such that the revolution time (hence the synchrotron frequency) remains constant.

- around the *unstable* fixed point (right) the orbits are not closed locally but around the separatrix, with virtually constant lengths. When approaching the unstable fixed point the orbit velocities tend to zero, and in absence of the compensation mechanism described above the revolution times tend to infinite.

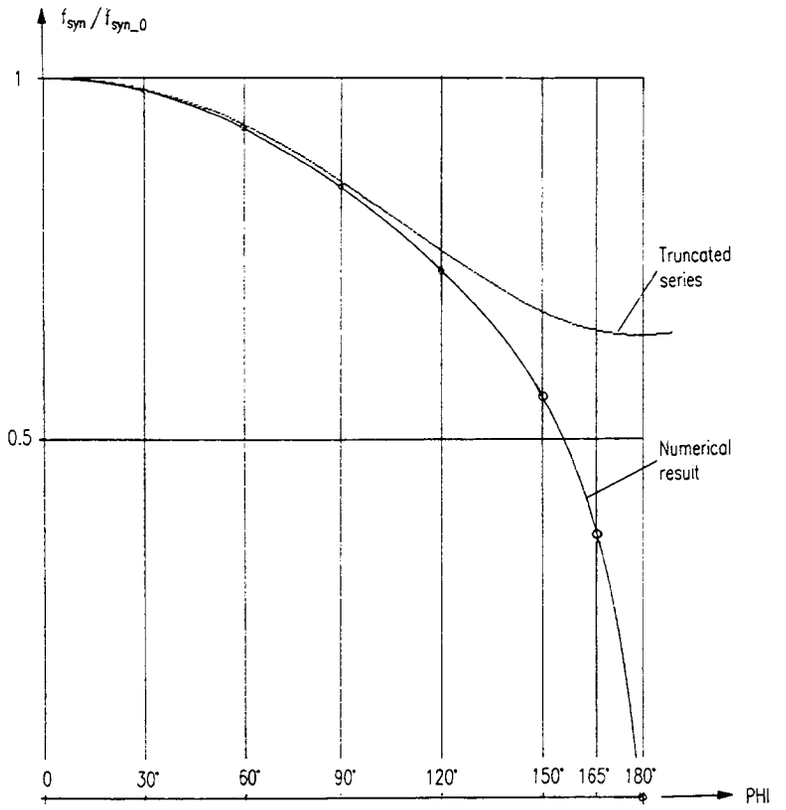


Fig. 8 Synchrotron frequency vs. phase of trajectory (stationary bucket)

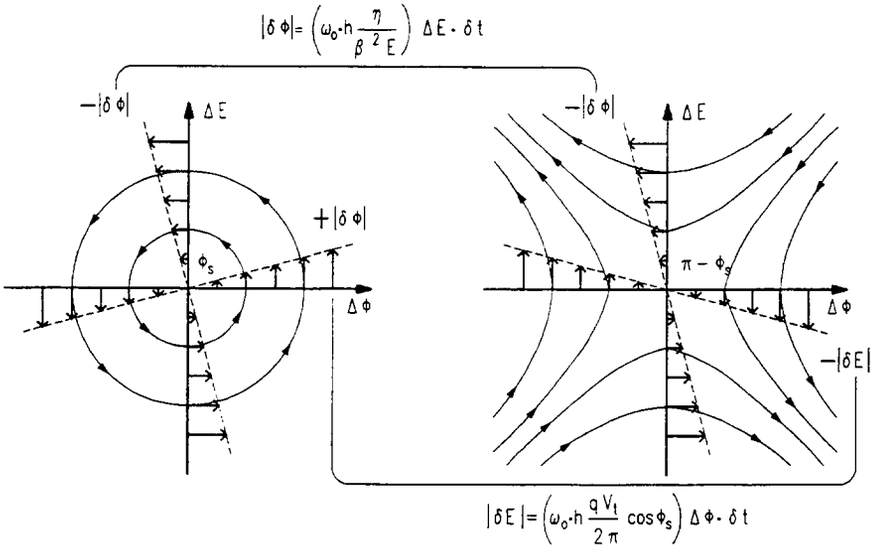


Fig. 9 Time derivatives along axes in the vicinity of a) stable fixed point, b) unstable fixed point

A word of caution concerning the numerical evaluation of the synchrotron frequency. The function to be integrated comprises poles at the two limiting phases ϕ_u and ϕ_l since the corresponding ΔE is zero, hence $1/\sqrt{\Delta E}$ infinite; numerical integration routines invariably stumble over these singularities. The solution is to integrate numerically only to the immediate

neighbourhood of the poles and add the result of analytical integration from there on, using the fact that $dx / \sqrt{x} = \sqrt{x} / 2$ which remains finite.

4.3 Beam dynamics above transition

Up to now all considerations supposed tacitly a beam below transition, i.e. in the region where the factor $\eta > 0$. From a formal point of view the transition crossing implies the change of the sign of η and hence of the right hand side of Eq. (2.2). The major consequence is the reversal of rotation sense in phase space, from counter-clockwise to clockwise, together the exchange of stable and unstable fixed point positions. The dynamics above transition are thus mirror images of the conditions below transition (see Fig. 10 below). All numerical results obtained for a given $|\eta|$ remain valid.

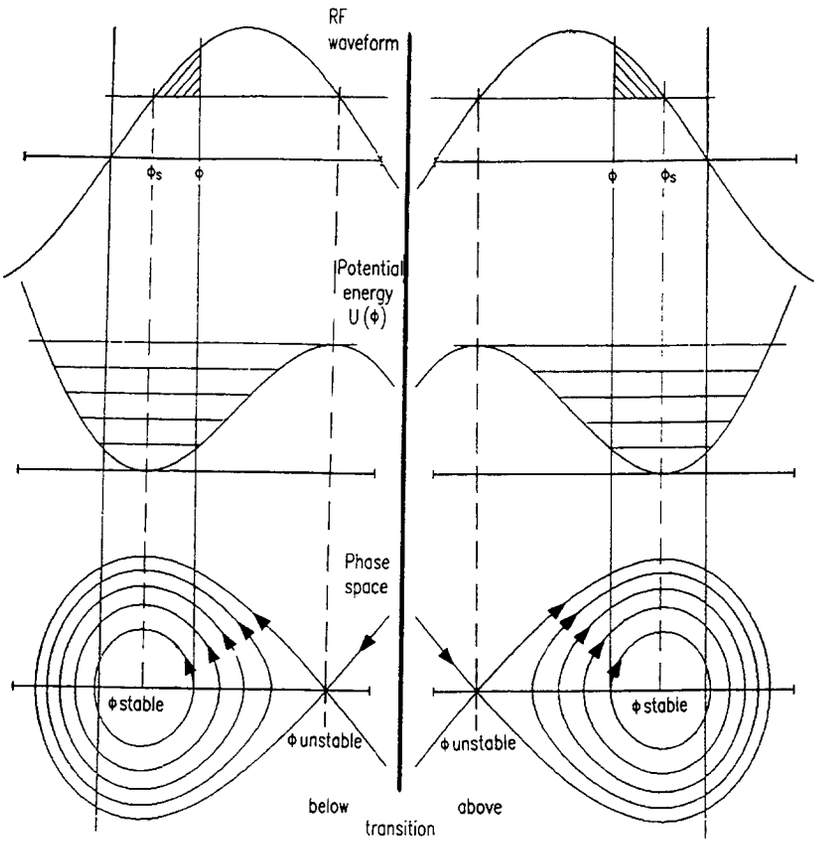


Fig. 10 Beam dynamics below and above transition

The particle motion comes to a complete stop at transition due to the reversal of the rotation sense, and is gradually increased or decreased around that energy. The transition crossing necessitates an RF phase jump by (180 - 2 ϕ_S) degrees. Special operating procedures like multiple phase jumps or an abrupt change of the machine tune (the so-called "gamma-transition jump") are often used to minimize beam losses and blowup during the process. At transition the synchrotron frequency and beam length tend towards zero and the energy spread towards infinite. A more detailed description can be found in [4].

4.4 Multifrequency RF voltage

In accelerators or storage rings there is in general one single RF frequency used at a time to accelerate or to hold the beam. For special purposes a second frequency may be added to shorten or lengthen the beam or to modify the structure of the energy spread.

Figure 11 shows as example the influence of a third-harmonic "Landau" cavity whose aim is to increase the energy spread at the core of the bunch to provide "Landau" damping. Starting with no harmonic content (left), the percentage of the third harmonic is increased to 33.3% (mid) for a maximally flat RF waveform at the centre and to 50% (right) for reversal of the RF slope at the stable phase of 0°.

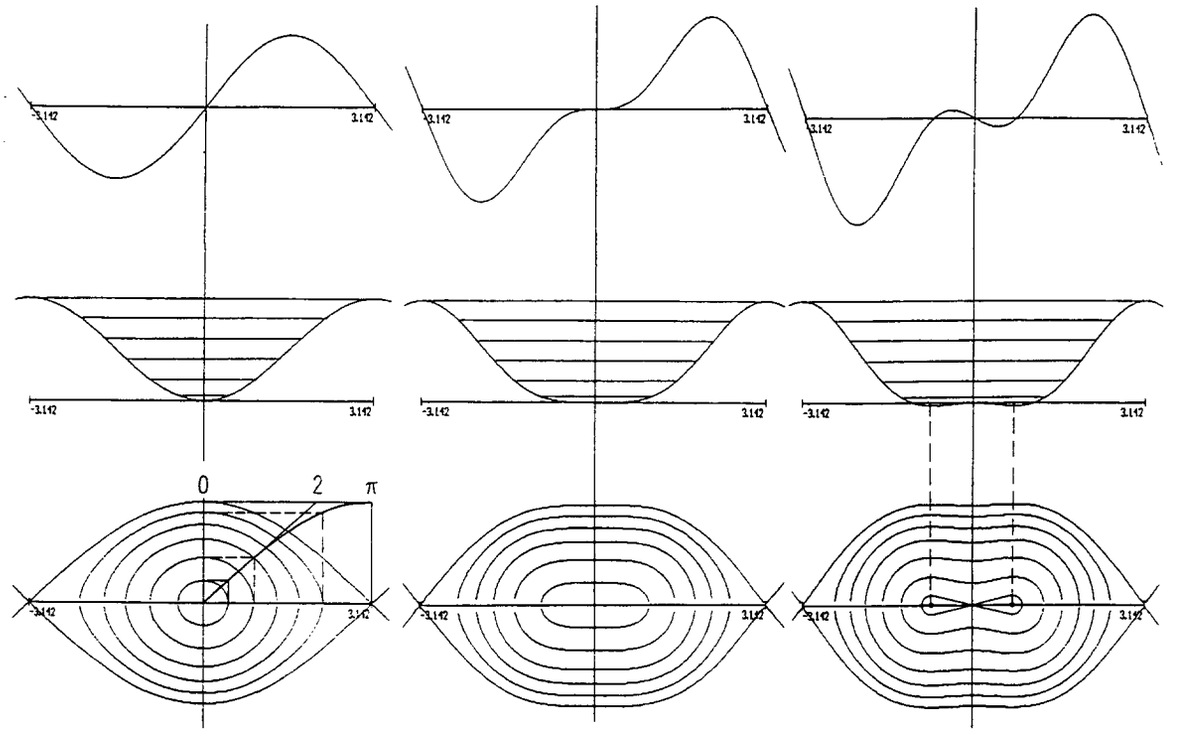


Fig. 11 Buckets for multi-frequency RF voltage (third harmonic).
a) 0%, b) 33%, c) 50% harmonic

For the case of the maximally flat RF waveform, the synchrotron frequency at the bunch centre becomes zero. It then increases to a maximum before decreasing again to zero at the boundary. The aim is to eliminate the area of uniform synchrotron frequency at the centre of the bunch, increase the local energy spread and therefore to enlarge the beam stability limit.

If the harmonic content is further increased the slope of the RF waveform is reversed and the hitherto stable point becomes a locally unstable singularity. Two sub-buckets are created in which particles circulate in stable trajectories.

Full analytical treatment of the parameters becomes very involved and numerical methods are used instead. The general approach remains the same as before, namely the establishment of a Hamiltonian via the potential function U that has to take the other components of the RF waveform into account. The power of single-particle analysis is however limited and multi-particle tracking codes and finally experiments are an essential part of the design process.

Note that the RF frequencies involved need not be harmonically related to each other. It is sufficient that both are a multiple of the revolution frequency.

4.5 Adiabatic damping of phase oscillations

A conservative system for which a Hamiltonian can be formulated has a very important property, namely the preservation of phase space which is also known as Liouville's theorem. It states that the volume of an element in phase space remains constant during subsequent adiabatic transformations.

The volume element in phase space comprises in general 3 * 2 = 6 dimensions. It is however supposed here that longitudinal and transverse dynamics are independent from each other and that a longitudinal phase space of only two dimensions can be considered separately from the rest (an assumption not valid for electron machines). With this premise in mind Liouville's theorem states that the adjacent surface elements constituting a beam may change in form but not in area during their journey along a trajectory. They behave like an incompressible fluid.

An even more remarkable corollary states that also the area of a *closed trajectory* remains constant provided all changes occur only adiabatically, i.e. infinitely slowly in comparison with the time needed for one complete turn. This is known as the theorem of Boltzmann-Ehrenfest. The Hamiltonian itself may change by orders of magnitude but provided these changes are adiabatic the area enclosed by a trajectory is preserved.

Consider a small partial bucket with phase deviation ψ in immediate vicinity of the stable fixed point. The corresponding energy deviation follows from Eq. (3.7)

$$\frac{\Delta\hat{E}_\Psi}{\Psi} = \sqrt{\frac{\beta^2\gamma}{\eta}\cdot\frac{qV_tE_0\cdot\cos\phi_s}{2\pi h}}$$

or

$$\Delta\hat{E}_\Psi = \Psi\sqrt{\frac{\eta}{\beta^2\gamma}\frac{2\pi h}{qV_tE_0\cos\phi_s}}$$

The partial bucket boundary is an ellipse. Its area A in canonical coordinates is

$$A = \pi\cdot\Psi\cdot\frac{\Delta\hat{E}_\Psi}{h\omega_0} = \frac{\pi}{h\omega_0}\Psi^2\sqrt{\frac{\eta}{\beta^2\gamma}\frac{2\pi h}{qV_tE_0\cos\phi_s}}$$

or rearranged

$$\psi = \sqrt{A}\cdot\left|\left(\frac{2h^3}{\pi qV_tE_0\cos\phi_s}\right)\left(\frac{\omega_0^2}{\beta^2}\right)\frac{\eta}{\gamma}\right|^{1/4} \tag{4.3}$$

On the other hand

$$\omega_0 = \frac{2\pi\beta c_0}{2\pi R} = \frac{\beta c_0}{R}$$

or

$$\frac{\omega_0}{\beta} = \frac{c_0}{R} = \text{const} \tag{4.4}$$

The only term in Eq. (4.3) that changes during acceleration is therefore the factor η/γ. Since the area enclosed by the trajectory remains constant for adiabatic changes of the machine parameters it follows that

$$\psi \sim \left|\frac{\eta}{\gamma}\right|^{1/4}$$

Starting from some initial value at injection the term η/γ decreases steadily and becomes zero at transition. The magnitude of the phase excursion of a particle (or the width of the bunch it delimits) decreases accordingly hence the name "adiabatic damping of phase oscillations" for this phenomenon.

This designation is however a misnomer since the term η/γ increases again up to a certain maximum after transition and so do the phase oscillations. A detailed discussion of this and of the impact on the other beam parameters is given in [1].

The preservation of longitudinal emittance is not guaranteed for "mismatched" beams whose boundaries do not correspond to a closed trajectory of the bucket. Consider the distorted beam boundary after bunch rotation shown in Fig. 7a. Since this boundary is obviously not a closed trajectory of the bucket, the basic requirement for the application of the Boltzmann-Ehrenfest theorem is not met. The bunch will not remain compact but will be "blown up" by "filamentation" after a few dozen turns of synchrotron oscillations.

Note that Liouville's theorem no longer applies here: the exposure of a given beam to a non-matched bucket is equivalent to an abrupt, non-adiabatic change of the former bucket parameters that shaped the bunch to the new conditions. The beam will not remain compact but will be infested by voids between batches of particles. This process has been compared to the – generally irreversible – transformation of a fixed volume of incompressible liquid into foam.

4.6 Non-preservation of full bucket area; effect of the "golf club"

The area of a stationary bucket as given by Eq. (3.4) is not constant for a given particle and fixed RF voltage, but changes during acceleration in proportion to the factor $|\eta / \gamma|^{-1/2}$ taking (4.4) into account. This is also true for an accelerating bucket whose area is linked to the area of the stationary bucket by the constant scale factor $\alpha(\phi_s)$.

The bucket area should in principle remain constant according to the Boltzmann-Ehrenfest theorem since its boundary, the separatrix, is a closed trajectory of the bucket. The fact that this area changes nevertheless during acceleration can be explained as a consequence of the violation of adiabaticity: since one turn around the separatrix lasts infinitely long there is no process that can possibly be slow in comparison.

The factor $|\eta / \gamma|^{-1/2}$ increases steadily during the acceleration process between injection and transition. The bucket area increases in proportion and the question arises from where the additional surface elements could come. The speculation is that even if adiabaticity is clearly violated for one complete turn around the separatrix, it might still be satisfied for most of the phase space since the slowing down of the synchrotron frequency is caused by the dynamics around an isolated singularity, namely the unstable fixed point. In that case the phase space preservation according to Liouville's theorem would still be valid everywhere else and some areas outside the separatrix could be identified that contribute to the build-up of bucket area.

Multi-particle tracking studies confirm this hypothesis. The area that crosses the "static" separatrix and penetrates into the formerly closed bucket area is situated in an elongated band above the unstable fixed point, giving the "dynamic" separatrix the characteristic shape of a golf

club. It is intuitively evident that a bucket that is moving towards higher energies may be able to capture additional particles whose energy is higher than nominal, i.e. rather those situated above the separatrix than those situated below.

A more detailed description of these effects is given in [1], considering also the zone of shrinking bucket size above transition.

5. LINACS

5.1 Differential equations

The concepts for circular accelerators exposed in the previous paragraphs remain in principle valid for the longitudinal beam dynamics in linear accelerators (linacs). The major differences are briefly

- the reference length is now the distance L between adjacent accelerating cells rather than the length of a non-existing closed orbit,
- the quantity $q\ V_t$ refers to the energy gain per cell taking the transit time factor in due account (this factor may become very critical at low energies),
- there is no physical mechanism that varies the average orbit length of a particle between cells as a function of its instantaneous energy, so formally the momentum compaction factor α_p is equal to zero and the phase slip factor $\eta \rightarrow 1/\gamma^2$.
- the frequency of acceleration f is fixed so the individual cell lengths L have to be modified in correspondence with the varying relative velocity β,
- under these circumstances it is more convenient to take the distance s along the linac axis as the independent variable rather than time,
- a convenient pair of canonically related variables is the normalized particle energy γ together with the phase ϕ,
- in a Linac environment it is usual to count phases from the top of the accelerating waveform rather than from positive zero crossing (see also annex 1). The convention used in synchrotrons is nevertheless maintained here.

The equivalent of equations (2.1) and (2.2) is

$$\frac{d}{ds}(\gamma) = \frac{qV_t}{E_0 \cdot L}(\sin\phi - \sin\phi_s) \tag{5.1}$$

$$\frac{d}{ds}(\phi) = -\frac{2\pi f}{c_0}\frac{(\Delta\gamma)}{\beta^3\gamma^3} \tag{5.2}$$

together with the Hamiltonian

$$H = \frac{\pi f}{c_0\beta^3\gamma^3}(\Delta\gamma)^2 + \frac{qV_t}{E_0 L}\cdot U(\phi) \tag{5.3}$$

Bucket length and height, bucket area and a spatial synchrotron "frequency" can be derived from these equations as in the case of circular accelerators.

The effects of non-adiabaticity are particularly pronounced in the first cells of a linear accelerator where the ratio (energy gain per cell / total energy) is usually quite high.

5.2 Non-relativistic vs relativistic linacs

A rough classification of linacs is based on the normalized velocity β of the particles they accelerate:

- varying $\beta < 1$: "non-relativistic" linacs, for protons or very low-energy electrons. The particle velocity varies appreciably in the machine, the cell lengths are variable. As a result the proper phasing between accelerating gaps relies on a certain energy profile along the axis which in turn depends on the cumulated energy gains at the upstream parts of the machine. There is in general a characteristic threshold of the accelerating tank voltage below which no acceleration is possible. At the threshold voltage the particles ride on the crest of the accelerating waveform, the bucket size is just emerging from zero.

- $\beta \equiv 1$: "(ultra)relativistic" linacs, mainly for electrons. The particle velocity is very close to the velocity of light and virtually constant, the cell lengths are therefore fixed. The proper phasing between cells is not dependent on an energy profile. As a consequence the beam will be further accelerated by almost any cell voltage, the output energy can be readily adjusted by the RF level or by addition/withdrawal of additional cells or tanks.

Full details are found in [2] and [5].

6. CONCLUDING REMARKS

This text is essentially the record of a 3-hour lecture on longitudinal beam dynamics at an advanced but not specialist level. The intention was to emphasise the underlying principles even if some mathematical tools like transformation of the Hamiltonian with reduced variables had to be sacrificed.

The content is a subset of the material presented by Dr. G. Dôme at an earlier meeting of the CERN Accelerator School [2]. The specialist will find in that article a wealth of detailed information which, together with the very complete list of references, represents the state-of-the-art on the subject.

REFERENCES

[1] P.M. Lapostolle, Proton Linear Accelerators, Los Alamos National Laboratory, Report LA-11601-MS.

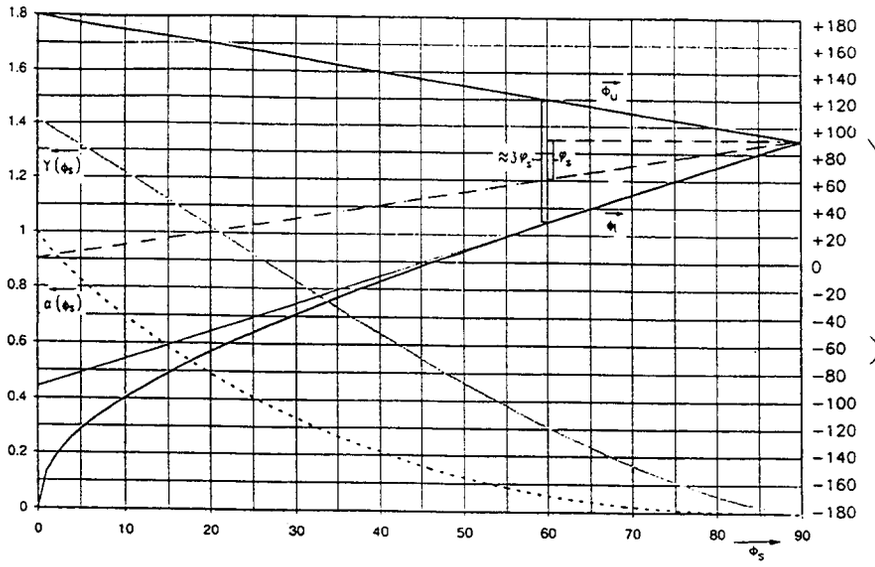
[2] G. Dôme, Theory of RF acceleration, Proceedings of CAS Oxford, CERN 87-03, p. 110-149.

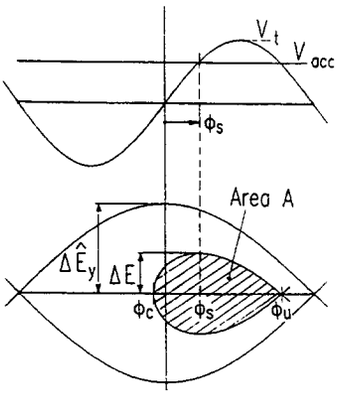
[3] P. Bryant and K. Johnsen, Lagrangian/Hamiltonian derivation of the phase equation, Proceedings of CAS Leewenhorst, CERN 92-01, p. 51.

[4] K. Johnsen, Transition, Proceedings of CAS Paris, CERN 85-19, p. 178

[5] P.M. Lapostolle, A.L. Septier (editors), Linear Accelerators, North Holland Publishing Company, Amsterdam, 1970.

ANNEX 1 BUCKET PARAMETERS





$$\Delta E = Y(\phi_s) \cdot \frac{1}{\sqrt{2}} \cdot \Delta\hat{E}_y \qquad \text{with } \Delta\hat{E}_y = \sqrt{2 \cdot \frac{\beta^2 \gamma}{\eta} \cdot \frac{q V_t E_o}{\pi h}}$$

$$A = \alpha(\phi_s) \cdot \frac{8}{h \cdot \omega_o} \cdot \Delta\hat{E}_y \qquad \text{(half height of stationary bucket)}$$

ϕ_s	ϕ_l	ϕ_u	$\alpha(\phi_s)$	$Y(\phi_s)$
0	-180	180	1	1.414
1	-154	179	0.954	1.395
2	-143.5	178	0.918	1.375
3	-135.5	177	0.885	1.356
4	-128.8	176	0.854	1.336
5	-122.9	175	0.825	1.317
6	-117.6	174	0.797	1.297
7	-112.6	173	0.770	1.278
8	-108.1	172	0.745	1.258
9	-103.7	171	0.720	1.238
10	-99.6	170	0.696	1.218
11	-95.7	169	0.673	1.199
12	-92	168	0.651	1.179
13	-88.4	167	0.629	1.159
14	-84.9	166	0.608	1.140
15	-81.5	165	0.587	1.120
16	-78.2	164	0.567	1.100
17	-75	163	0.548	1.081
18	-71.9	162	0.529	1.061
19	-68.9	161	0.510	1.041
20	-65.9	160	0.492	1.022
21	-63	159	0.474	1.002
22	-60.1	158	0.457	0.982
23	-57.3	157	0.440	0.963
24	-54.5	156	0.424	0.943
25	-51.8	155	0.408	0.924
26	-49.1	154	0.392	0.905
27	-46.4	153	0.377	0.885
28	-43.8	152	0.362	0.866
29	-41.2	151	0.347	0.847

ϕ_s	ϕ_l	ϕ_u	$\alpha(\phi_s)$	$Y(\phi_s)$
30	-38.7	150	0.333	0.828
31	-36.2	149	0.320	0.808
32	-33.7	148	0.306	0.789
33	-31.2	147	0.293	0.769
34	-28.8	146	0.281	0.752
35	-26.3	145	0.268	0.733
36	-23.9	144	0.256	0.714
37	-21.6	143	0.245	0.696
38	-19.2	142	0.233	0.677
39	-16.9	141	0.222	0.659
40	-14.6	140	0.215	0.640
41	-12.3	139	0.201	0.622
42	-10	138	0.191	0.604
43	-7.7	137	0.181	0.586
44	-5.4	136	0.172	0.569
45	-3.2	135	0.163	0.551
46	-1	134	0.154	0.533
47	1.2	133	0.145	0.516
48	3.5	132	0.137	0.499
49	5.6	131	0.129	0.482
50	7.8	130	0.121	0.465
51	10	129	0.114	0.448
52	12.2	128	0.107	0.431
53	14.3	127	0.100	0.415
54	16.4	126	0.093	0.399
55	18.6	125	0.087	0.383
56	20.7	124	0.081	0.367
57	22.8	123	0.075	0.351
58	24.9	122	0.070	0.335
59	27	121	0.064	0.320

ϕ_s	ϕ_l	ϕ_u	$\alpha(\phi_s)$	$Y(\phi_s)$
60	29.1	120	0.059	0.305
61	31.2	119	0.054	0.290
62	33.3	118	0.050	0.276
63	35.4	117	0.046	0.261
64	37.4	116	0.041	0.247
65	39.5	115	0.038	0.233
66	41.6	114	0.034	0.219
67	43.6	113	0.031	0.206
68	45.7	112	0.027	0.193
69	47.7	111	0.024	0.180
70	49.7	110	0.022	0.167
71	51.8	109	0.019	0.155
72	53.8	108	0.017	0.143
73	55.8	107	0.014	0.131
74	57.9	106	0.012	0.120
75	59.9	105	0.011	0.109
76	61.9	104	0.009	0.098
77	63.9	103	0.007	0.088
78	65.9	102	0.006	0.078
79	68	101	0.005	0.069
80	70	100	0.004	0.059
81	72	99	0.003	0.051
82	74	98	0.002	0.043
83	76	97	0.002	0.035
84	78	96	0.001	0.028
85	80	95	0.001	0.021
86	82	94	0.000	0.015
87	84	93	0.000	0.010
88	86	92	0.000	0.005
89	88	91	0.000	0.002
90	90	90	0	0.000

Remark: Two different definitions coexist for the stable phase angle phi:

-in circular accelerators, phi is counted from the zero crossing (symbol Φ_s),

-in linacs, phi is counted from top (symbol φ_s, where $\varphi_s = 90 - \Phi_s$).

The bucket width $\Phi_u - \Phi_l$ is very close to $3 * \varphi_s$ (law of " $3\,\varphi_s$ ", see figure above).

KINETIC THEORY

A. Hofmann
CERN, Geneva, Switzerland

Abstract
The treatment of the kinetic theory presented here concentrates on the physical principles involved and on practical applications. It does not pretend to give rigorous mathematical derivations. From the conservation of the phase-space area - as stated by Liouville's theorem - the Vlasov equation is derived. In the case of stationary conditions this equation states that the phase-space distribution must be a function of the Hamiltonian only. This fact is a very powerful tool for finding stationary distributions in certain conditions. As an application the longitudinal distributions of the particles in a non-linear RF wave form and in the presence of a potential well created by a reactive impedance are derived. For the non-stationary case we often use a perturbation method and write the Vlasov equation for a small deviation from the stationary distribution. In the second part dissipative and random forces are investigated. Using an intuitive approach the effects of such forces on the phase-space area are taken into account and the Fokker-Planck equation is derived. As an application the energy distribution of the particles is evaluated for the dissipative and random effects due to the emission of synchrotron radiation. Often only global parameters such as the rms energy spread are of interest. They can be obtained more directly from Campbell's theorem.

1 INTRODUCTION

The kinetic theory is treated here in an intuitive way with emphasis on physical principles and practical applications rather than rigorous derivations. Detailed mathematical treatments of the equations involved can be found in several publications on this subject [1, 2]. We start here with Liouville's theorem which states the invariance of the phase-space density in the immediate neighborhood of a probe particle we follow. Expressing this fact as seen by a stationary observer a simple but not rigorous derivation of the Vlasov equation can be obtained. This equation describes the flow of particles in a way similar to that of a moving liquid. It is therefore only valid when a very large number of particles is involved. The case of stationary distributions is of special interest. It can be shown from the Vlasov equation that in this case the distribution depends only on the Hamiltonian which describes the motion of the individual particles. This can be used to obtain stationary distributions for special cases.

2 THE VLASOV EQUATION

We consider now particles moving in phase-space with coordinates q and p. Usually we associate q with a space coordinate and p with a momentum or a velocity but the situation can be more general. These coordinates q and p should be canonically conjugate which means that they are derived from a Hamiltonian $H(q,p)$ by the canonical equations

$$\dot{q} = \frac{\partial H(q,p)}{\partial p} \ , \quad \dot{p} = -\frac{\partial H(q,p)}{\partial q}. \tag{1}$$

We will not explain these fundamental equations any further here but refer to the standard literature and in particular to the CAS courses [2, 3]. The theorem of Liouville is visualized

in Fig. 1 where the phase-space trajectories of three particles are drawn. Their positions and the triangular surface element they determine are shown for two different times t. The form of this surface element changes but its area stays constant. The phase-space trajectories can of course not cross each other since (except at a singularity) this would mean that two particles have the same position and velocity at one moment but different values later.

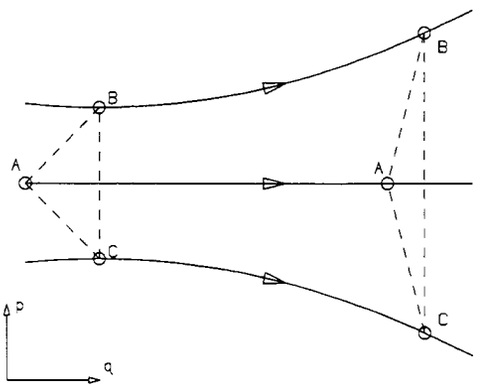


Figure 1: Phase-space trajectories and Liouville's theorem

According to Liouville's theorem the particles in phase space move like an incompressible liquid. This conservation of the phase-space density seems to contradict common experience with accelerators where one often talks about dilution of phase-space due to non-linearities, called filamentation. This paradox is illustrated in Fig. 2. On the left hand side we show a phase space occupied by particles which has the simple form of a rectangle. The instruments observing the distribution of the particles have, in practice, a limited resolution indicated by the grid in the figure. For the left diagram the instruments would localize the beam within seven such resolution elements. After the beam has gone through some non-linear elements (e.g. a mismatched bunch rotating in the non-linear RF wave form) the form of the phase-space occupied by the beam is distorted. Although its actual area has not changed the limited resolution of the instruments indicates now a distribution of particles over many more resolution grid elements. The situation is similar to a liquid which is transformed into foam. On a microscopic level the density is still the same but the global density is much less. For practical purposes, such as for determining the luminosity, the global phase-space density is of course also relevant.

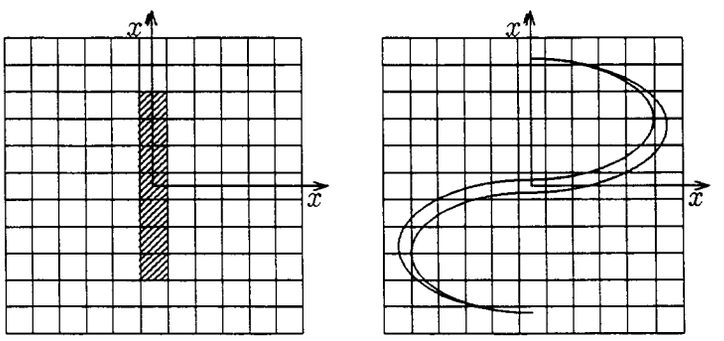


Figure 2: Resolution and phase-space area conservation [6]

The constancy of the phase density $\psi(q,p)$ is expressed by the equation

$$\frac{d\psi(q,p)}{dt} = 0 \tag{2}$$

where the absolute differentiation indicates that one follows the particle while measuring the density of its immediate neighborhood. We would like to know the development of this density as seen by a stationary observer (like a beam monitor) which does not follow the particle. It depends now not only directly on the time t but also indirectly through the coordinates q and p of the moving particles which change with time. Therefore, we have to express the absolute differential with respect to t by the partial differentiations with respect to time as well as with respect to the coordinates q and p multiplied with their time derivative

$$\frac{d\psi(q,p)}{dt} = \frac{\partial\psi(q,p)}{\partial t} + \dot{q}\frac{\partial\psi(q,p)}{\partial q} + \dot{p}\frac{\partial\psi(q,p)}{\partial p} = 0. \tag{3}$$

This expression is the Vlasov equation in its most simple form and is nothing else but an expression for Liouville's conservation of phase-space density seen by a stationary observer. Rigorous derivation of the Vlasov equation can be found in the literature e.g. [1, 2]. In this presentation we apply it to particular accelerator problems. In the first group of applications we search for a stationary distribution which fulfills certain boundary conditions. Such distributions do not depend explicitly on time. For example a stationary bunch looks the same each revolution as it is observed through an intensity monitor. For this stationary case the Vlasov equation becomes

$$\frac{\partial\psi(q,p)}{\partial t} = 0 \;\rightarrow\; \dot{q}\frac{\partial\psi(q,p)}{\partial q} + \dot{p}\frac{\partial\psi(q,p)}{\partial p} = 0. \tag{4}$$

Expressing $\dot{q}$ and $\dot{p}$ with the canonical equations we get

$$\frac{\partial\psi}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial\psi}{\partial p}\frac{\partial H}{\partial q} = [\psi, H] = 0. \tag{5}$$

The above expression is called a Poisson bracket. For a system with one variable pair (p,q) a stationary distribution is a function of the Hamiltonian only, $\psi = \psi(H)$. It is easy to show the reverse statement. Assuming $\psi(q,p)$ being a function of the Hamiltonian we get

$$\psi(q,p) = \psi(H) \;\rightarrow\; \frac{\partial\psi}{\partial q} = \frac{d\psi}{dH}\frac{\partial H}{\partial q}\;,\;\; \frac{\partial\psi}{\partial p} = \frac{d\psi}{dH}\frac{\partial H}{\partial p}, \tag{6}$$

or

$$\frac{\partial\psi/\partial q}{\partial H/\partial q} = \frac{\partial\psi/\partial p}{\partial H/\partial p} \;\rightarrow\; \frac{\partial\psi}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial\psi}{\partial p}\frac{\partial H}{\partial q} = 0, \tag{7}$$

which leads again to the stationary Vlasov equation. A stationary distribution is therefore a function of the Hamiltonian only and does not depend explicitly on the coordinates q and p but only implicitly through their relation with the Hamiltonian. This fact is a very powerful tool for finding stationary particle distributions [4, 5].

Before we go to the practical applications we have to make some remarks on the coordinates q and p we use. In Liouville's theorem the phase-space area is only conserved if expressed in canonically conjugate variables q, p. The same criterion applies to the validity of the Vlasov equation. Examples of such pairs are position and momentum x, p or time and energy t, E. These variables are often not very practical for accelerator applications. We could for example choose the longitudinal position s and the corresponding momentum p_s as coordinates to describe the longitudinal phase-space. However the path length s around the ring is different for particles with different momenta. As a result, individual

particles would enter a certain machine element, such as a quadrupole, at different values of the coordinate s. From this point of view other coordinate pairs in which the position of the machine element is given for all particles by the same value of one coordinate are more suitable. Such a variable is for example the angle θ of the motion around the ring. However the conjugate variable is quite complicated [7]

$$W = \int_{E_0}^{E_0+\Delta E} \frac{\Delta E}{\omega} d(\Delta E) \tag{8}$$

where ΔE is the deviation from the nominal energy E_0 and $\omega(\Delta E)$ is the revolution frequency which depends on this energy deviation. In our applications we only use terms up to first order in ΔE and use as an approximation θ and ΔE as variables. However we should always remember that this is a linear approximation in $\Delta E/E_0$ which can become very inaccurate in some cases such as the neighborhood of transition energy. As a further convenience we will us the RF phase angle ϕ instead of θ to which it is related through $\phi = h\theta$ with h being the harmonic number.

3 APPLICATIONS OF THE VLASOV EQUATION

3.1 Stationary distribution in a non-linear RF wave form

As an introduction we will first go through the derivation of the longitudinal beam dynamics and try to find the longitudinal particle distribution in the presence of the non-linearity represented by the RF wave form. We consider a storage ring with circumference C which depends on the deviation Δp from the nominal momentum p or ΔE from the nominal energy E.

$$C = C_0\left(1 + \alpha\frac{\Delta p}{p}\right) = C_0\left(1 + \frac{\alpha}{\beta^2}\frac{\Delta E}{E}\right), \tag{9}$$

where α is the momentum compaction factor. The revolution time $T = C/\beta c$ becomes

$$T = T_0\left(1 + \left(\alpha - \frac{1}{\gamma^2}\right)\frac{1}{\beta^2}\frac{\Delta E}{E}\right) \sim T_0(1 + \alpha\frac{\Delta E}{E}), \tag{10}$$

where an approximation for ultra-relativistic particles ($\gamma \gg 1$ and $\gamma^2 \gg 1/\alpha$) has been made and will be used from now on.

In each revolution a particle goes through an RF cavity which is driven with a frequency $\omega_{RF} = h\omega_0$ and a peak voltage $\hat{V}$ with ω_0 being the nominal revolution frequency $\omega_0 = 2\pi\beta c/C_0$ and the integer h being the harmonic number.

$$V(t) = \hat{V}\sin(h\omega_0 t). \tag{11}$$

We assume now that the particle looses in each turn a certain amount of energy U_s due to the emission of synchrotron radiation

$$U_s = U_{s0} + \frac{dU}{dE}\Delta E. \tag{12}$$

For the nominal energy E we have a stationary condition such that the particle receives at each passage through the cavity the same amount of energy U_{s0} as it looses due to synchrotron radiation. It has to traverse the cavity at the synchronous time t_s such that

$$V(t_s) = U_{s0}/e \;\rightarrow\; \sin(h\omega_0 t_s) = \frac{U_{s0}}{e\hat{V}}. \tag{13}$$

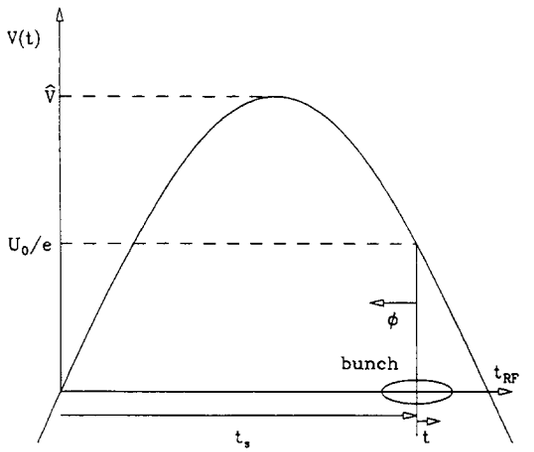


Figure 3: RF acceleration and phase focusing

Instead of the time t we use now the RF phase angle ϕ_{RF}

$$\phi_{RF} = \pi - h\omega_0 t \ , \ \ \phi_s = \pi - h\omega_0 t_s \tag{14}$$

and call the deviation from this synchronous phase angle ϕ_s

$$\phi = \phi_{RF} - \phi_s \tag{15}$$

as indicated in Fig.3. The energy gain and loss in one turn becomes

$$\delta(\Delta E) = e\hat{V}\sin\phi_{RF} - U_s = e\hat{V}(\cos\phi_s\sin\phi + \sin\phi_s\cos\phi) - \left(U_{s0} + \frac{dU}{dE}\Delta E\right). \tag{16}$$

Using the equilibrium for the synchronous particle

$$U_{s0} = e\hat{V}\sin\phi_s \tag{17}$$

we get

$$\delta(\Delta E) = e\hat{V}(\cos\phi_s\sin\phi - \sin\phi_s(1-\cos\phi) - \frac{dU}{dE}\Delta E. \tag{18}$$

In most practical cases the energy gain and loss $\delta(\Delta E)$ per turn is very small compared to the energy E itself. As an approximation we can replace the sudden energy gain while going through the cavity by a smooth acceleration.

$$\Delta\dot{E} = \frac{\omega_0}{2\pi}\left(e\hat{V}(\cos\phi_s\sin\phi - \sin\phi_s(1-\cos\phi)) - \frac{dU}{dE}\Delta E\right). \tag{19}$$

The dependence of the revolution time on the energy deviation ΔE leads to a change of the revolution time of the particle

$$\frac{\Delta T}{T} = \alpha\frac{\Delta E}{E}. \tag{20}$$

This leads in the smooth approximation to a time derivative of the phase

$$\dot{\phi} = -\omega_0\alpha h\frac{\Delta E}{E}. \tag{21}$$

Using this relation to replace ΔE in Eq. (19) gives

$$\ddot{\phi} + \frac{\omega_0}{2\pi}\frac{dU}{dE}\dot{\phi} + \frac{\omega_0^2 \alpha h e \hat{V}}{2\pi E}(\cos\phi_s \sin\phi - \sin\phi_s(1-\cos\phi)) = 0. \tag{22}$$

Introducing the synchrotron frequency ω_{s0}, the synchrotron tune Q_{s0} and the damping rate α_E

$$\omega_{s0}^2 = \omega_0^2 \frac{\alpha h e \hat{V}\cos\phi_s}{2\pi E}\ ,\ Q_{s0} = \frac{\omega_{s0}}{\omega_0} \text{ and } \alpha_E = \frac{\omega_{s0}}{2\pi}\frac{dU}{dE} \tag{23}$$

in Eq. (22) leads to a more compact expression

$$\ddot{\phi} + 2\alpha_E\dot{\phi} + \omega_{s0}^2\left(\frac{\cos\phi_s\sin\phi - \sin\phi_s(1-\cos\phi)}{\cos\phi_s}\right) = 0. \tag{24}$$

For small amplitudes $\phi \ll 1$ we can linearize the above equation

$$\ddot{\phi} + 2\alpha_E\dot{\phi} + \omega_{s0}^2\phi = 0 \tag{25}$$

which has the solution

$$\phi(t) = \hat{\phi}e^{-\alpha_E t}\cos(\omega_{s0}\sqrt{1-(\alpha_E/\omega_{s0})^2}t + \phi_0), \tag{26}$$

where the amplitude $\hat{\phi}$ and the phase constant ϕ_0 are given by the initial conditions. Usually $\alpha_E \ll \omega_{s0}$ and the solution is approximately

$$\phi(t) \sim \hat{\phi}e^{-\alpha_E t}\cos(\omega_{s0}t). \tag{27}$$

This is the equation of a damped oscillator with damping rate α_E and resonant frequency $\omega_{s0}\sqrt{1-(\alpha_E/\omega_{s0})^2} \sim \omega_{s0}$. The Vlasov equation applies to systems without dissipative or random forces. We can therefore not use it to describe particles undergoing damped synchrotron oscillations. For our next applications we will neglect the damping and take as approximation $\alpha_E \sim 0$. Strictly speaking, we should, according to Eq. (13), also set the synchronous phase angle ϕ_s to zero. We will however carry the term $\cos\phi_s$ along and pretend that the finite synchronous phase angle is here as a compensation for a conservative force.

We neglect now the damping and go back to our two first-order differential equations

$$\Delta\dot{E} = \frac{\omega_0}{2\pi}e\hat{V}(\cos\phi_s\sin\phi - \sin\phi_s(1-\cos\phi))\ ,\ \dot{\phi} = -\omega_0\alpha h\frac{\Delta E}{E} \tag{28}$$

or the second-order differential equation

$$\ddot{\phi} + \frac{\omega_{s0}^2}{\cos\phi_s}(\cos\phi_s\sin\phi - \sin\phi_s(1-\cos\phi) = 0. \tag{29}$$

Multiplying this last equation with $\dot{\phi}$ and integrating once gives

$$\frac{\dot{\phi}^2}{2} + \frac{\omega_{s0}^2}{\cos\phi_s}(\cos\phi_s(1-\cos\phi) - \sin\phi_s(\phi - \sin\phi) = H' = \text{const.} \tag{30}$$

This expression is a constant of motion and must therefore be directly related to the Hamiltonian. We find easily that the Hamiltonian is just

$$H = \frac{EH'}{\omega_0\alpha h} = \omega_0\alpha h E\left[\frac{1}{2}\left(\frac{\Delta E}{E}\right)^2 + \left(\frac{Q_s}{\alpha h}\right)^2\frac{(\cos\phi_s(1-\cos\phi) - \sin\phi_s(\phi-\sin\phi))}{\cos\phi_s}\right] \tag{31}$$

which satisfies the canonical equations

$$\frac{\partial H}{\partial \phi} = \Delta\dot{E} = \frac{\omega_0}{2\pi} e\hat{V}(\cos\phi_s \sin\phi - \sin\phi_s(1-\cos\phi)) = \Delta\dot{E} \ , \ -\frac{\partial H}{\partial \Delta E} = -\omega_0 \alpha h \frac{\Delta E}{E} = \dot{\phi}. \tag{32}$$

We show later in section 4 that the electron beam has a Gaussian energy distribution with rms value σ_E due to the interplay between damping and quantum excitation by the emitted synchrotron radiation

$$\psi(\phi, \Delta E) \propto \exp\left(-\frac{1}{2}\left(\frac{\Delta E/E}{\sigma_E/E}\right)^2\right). \tag{33}$$

On the other hand we also know that the stationary distribution should only be a function of the Hamiltonian. Comparing Eq. (31) with Eq. (33) we find that the phase-space distribution must be of the form

$$\psi(\phi, \Delta E) = C_2 \exp\left(-\frac{\left(\frac{\Delta E}{E}\right)^2 + 2\left(\frac{Q_s}{\alpha h}\right)^2 (\cos\phi_s(1-\cos\phi) - \sin\phi_s(\phi - \sin\phi))}{2\left(\frac{\sigma_E}{E}\right)^2}\right). \tag{34}$$

The constant factor can be determined from the normalizing condition

$$\int\int \psi(\phi, \delta E)(\Delta E) d\phi = 1 \tag{35}$$

which has to be solved numerically for the general case. In many practical cases the bunch is much shorter than the RF wavelength, i.e. $\phi \ll 1$, and we can approximate $1 - \cos\phi \sim \phi^2/2$ leading to

$$\psi(\phi, \Delta E) = \frac{1}{2\pi\sigma_\phi \sigma_E/E} \exp\left(-\frac{1}{2}\left(\left(\frac{\Delta E/E}{\sigma_E/E}\right)^2 + \left(\frac{\phi}{\sigma_\phi}\right)^2\right)\right), \tag{36}$$

where σ_ϕ is the rms bunch length measured in RF phase

$$\sigma_\phi = \frac{\alpha h}{Q_s}\frac{\sigma_E}{E}. \tag{37}$$

In this short-bunch approximation the bunch is Gaussian in energy and in phase.

3.2 Stationary distribution in a double RF system

The treatment used above can easily be extended to the case of a double RF system. This system has, besides the normal RF system with frequency $\omega_{RF} = h\omega_0$ and amplitude $\hat{V}$, another system with a frequency $n\omega_{RF}$ being a harmonic of the RF frequency and amplitude $k\hat{V}$, Fig. 4. The total voltage seen by the beam becomes

$$V(\phi) = V_1 + V_n = \hat{V}\left[\sin(\phi_s + \phi) + k\sin(n\phi_n + n\phi)\right]. \tag{38}$$

Such double RF systems are often used to control the bunch length and to produce a large spread in synchrotron frequency to get Landau damping. They are particularly interesting if operated in the bunch-lengthening mode. For this the RF wave form should fulfill two neccessary and one convenient conditions:

a) The energy loss U_s is compensated $V(0) = \hat{V}[\sin\phi_s + k\sin(n\phi_n)] = U_s/e$.

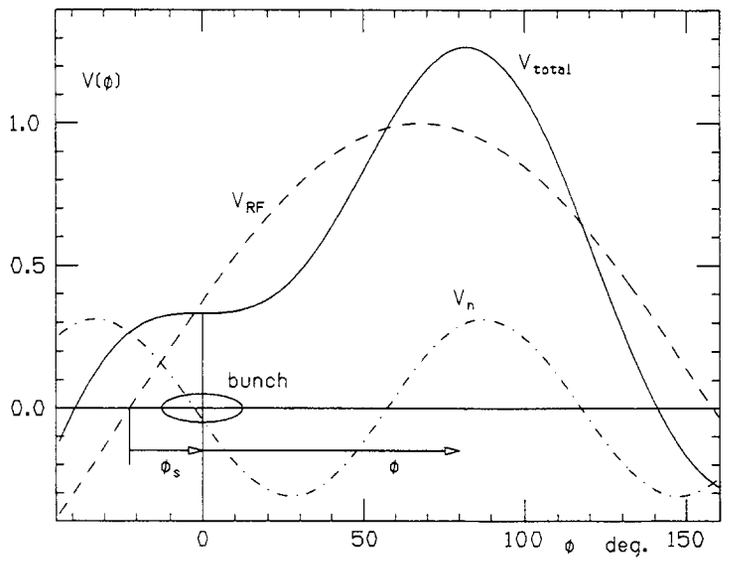


Figure 4: Double RF system

b) The slope vanishes at the bunch center $V'(0) = \hat{V}[\cos\phi_s + kn\cos(n\phi)] = 0$.

c) The curvature vanishes at the center $V''(0) = -\hat{V}[\sin\phi_s + kn^2\sin(n\phi_n)] = 0$.

Using these three conditions and approximating for the case of a bunch short compared to the higher harmonic wave-length we get for the voltage seen by the beam

$$V(\phi) \sim \hat{V}\left[\left(1-\frac{1}{n^2}\right)\sin\phi_s\left(\frac{n^2-1}{6}\right)\phi^3 + \cos\phi_s\right] = \frac{U_s}{e} + \hat{V}\frac{n^2-1}{6}\phi^3. \tag{39}$$

Using the same procedure as in the last subsection we find for the two first-order differential equations

$$\dot{\phi} = -\omega_0\alpha h\frac{\Delta E}{E}\ ,\ \ \Delta\dot{E} = \frac{\omega_0}{2\pi}e\hat{V}\cos\phi_s(\frac{n^2-1}{6})\phi^3 \tag{40}$$

and for the second-order differential equation

$$\ddot{\phi} + \frac{\omega_0^2\alpha h e\hat{V}\cos\phi_s}{2\pi E}\left(\frac{n^2-1}{6}\right)\phi^3 = \ddot{\phi} + \omega_{s0}^2\frac{n^2-1}{6}\phi^3 = 0. \tag{41}$$

where the synchrotron frequency ω_{s0} of a single RF system is used here as a convenient parameter. This last expression is a non-linear differential equation the solution of which can be expressed with Jacobian elliptic functions [8]. However we are here only interested to find the Hamiltonian. Multiplying it with $\dot{\phi}$ and integrating gives

$$\frac{\dot{\phi}^2}{2} + \omega_{s0}^2\frac{n^2-1}{24}\phi^4 = H' = \text{const.} \tag{42}$$

The Hamiltonian itself is

$$H = \frac{E}{\omega_0\alpha h}H' = \frac{\omega_0\alpha h E}{2}\left[\left(\frac{\Delta E}{E}\right)^2 + \left(\frac{Q_s}{\alpha h}\right)^2\frac{n^2-1}{12}\right]. \tag{43}$$

It can easily be checked that it is related to the coordinates ϕ and ΔE through the proper canonical equations. Using again the fact that the energy distribution of an electron beam is Gaussian

$$\psi(\phi, \Delta E) \propto \exp\left(-\frac{1}{2}\left(\frac{\Delta E/E}{\sigma_E/E}\right)^2\right) \tag{44}$$

and that the total distribution is a function of the Hamiltonian only we find the phase-space distribution

$$\psi(\phi, \Delta E) = C_2 \exp\left(-\frac{H}{\omega_0 \alpha h (\sigma_E/E)^2}\right) = C_2 \exp\left(-\frac{\left(\frac{\Delta E}{E}\right)^2 + \left(\frac{Q_s}{\alpha h}\right)^2 \frac{n^2-1}{12}\phi^4}{2\left(\frac{\sigma_E}{E}\right)^2}\right). \tag{45}$$

To get the instantaneous current of this bunch distribution we integrate over ΔE

$$I(\phi) = 2\pi h I_0 \int \psi(\phi, \Delta E) d(\Delta E) = C_3 \exp\left(-\frac{\left(\frac{Q_s}{\alpha h}\right)^2 \frac{n^2-1}{12}\phi^4}{2(\sigma_E/E)^2}\right). \tag{46}$$

This current has no longer a Gaussian distribution. The rms bunch length expressed in RF phase angle is

$$\sigma_\phi = \frac{2\sqrt{\pi}}{\Gamma(1/4)} \sqrt{\sqrt{\frac{3}{n^2-1}} \frac{\alpha h}{Q_s} \frac{\sigma_E}{E}}. \tag{47}$$

Normalizing leads to the final expression for the instantaneous current

$$I(\phi) = \frac{4\pi\sqrt{2\pi} I_0}{\sqrt{\sqrt{2}}\Gamma^2(\frac{1}{4})\sigma_\phi} \exp\left(-\frac{2\pi^2}{\Gamma^2(1/4)}\left(\frac{\phi}{\sigma_\phi}\right)^4\right). \tag{48}$$

From the phase-space and current distribution other properties such as the distribution of synchrotron frequencies, peak current, space-charge effects, etc can be calculated [8, 9].

3.3 Potential-well bunch lengthening

We calculate now the longitudinal particle distribution for a beam with a Gaussian energy distribution in the presence of an inductive wall impedance [10]. We consider a bunch in a single RF system which we linearize in the neighborhood of the RF phase angle

$$V_{RF} = \hat{V} \sin \phi_{RF} = \hat{V} \sin(\phi + \phi_s) \sim \hat{V}(\cos \phi_s \phi + \sin \phi_s). \tag{49}$$

We assume that the beam surroundings represent an inductive impedance. This is in fact a reasonable assumption for long bunches. Their spectrum lies mainly below the resonant frequencies of the typical parasitic modes where the impedance is inductive. Taking an inductance L which is approximately constant over the bunch spectrum we get an induced voltage of the form

$$V_i = L\frac{dI}{dt} = -\omega_0 h L \frac{dI}{d\phi}. \tag{50}$$

The impedance of the inductance is

$$Z_L(\omega) = j\omega L. \tag{51}$$

We introduce the absolute value of this impedance divided by the mode number $n = \omega/\omega_0$ which is directly related to the inductance L and the revolution frequency ω_0

$$\left|\frac{Z}{n}\right| = \frac{\omega L}{\omega/\omega_0} = \omega_0 L. \tag{52}$$

With this we find for the induced voltage

$$V_i = -h|Z/n|\frac{dI}{d\phi}. \tag{53}$$

The introduction of the impedance divided by the mode number seems to be just a convenient way to describe the inductance. It is however a good way to characterize the relevant impedance of a ring. It is typically between 10 and 20 Ohms for older machines with many aperture changes and between 1 and 5 Ohms for more modern rings with smooth vacuum chambers.

This induced voltage V_i is seen by the particles in the bunch and has to be subtracted from the RF voltage. We have for the total voltage

$$V(\phi) = V_{RF} - V_i = \hat{V}\left(\cos\phi_s\phi + \sin\phi_s + \frac{h|Z/n|}{\hat{V}\cos\phi_s}\frac{dI}{d\phi}\right). \tag{54}$$

The smoothed energy gain per unit time is

$$\Delta\dot{E} = \frac{\omega_0 e\hat{V}\cos\phi_s}{2\pi}\left(\phi + \frac{h|Z/n|}{\hat{V}\cos\phi_s}\frac{dI}{d\phi}\right) + \frac{\omega_0}{2\pi}(e\hat{V}\sin\phi_s - U_s). \tag{55}$$

For the equilibrium condition $e\hat{V}\sin\phi_s = U_s$ we get the two first-order differential equations

$$\Delta\dot{E} = \frac{\omega_0 e\hat{V}\cos\phi_s}{2\pi}\left(\phi + \frac{h|Z/n|}{\hat{V}\cos\phi_s}\frac{dI}{d\phi}\right) \quad , \quad \dot{\phi} = -\omega_0\alpha h\frac{\Delta E}{E} \tag{56}$$

and the second-order equation

$$\ddot{\phi} + \frac{\omega_0^2\alpha h e\hat{V}\cos\phi_s}{2\pi E}\left(\phi + \frac{h|Z/n|}{\hat{V}\cos\phi_s}\frac{dI}{d\phi}\right) = \ddot{\phi} + \omega_{s0}^2\left(\phi + \frac{h|Z/n|}{\hat{V}\cos\phi_s}\frac{dI}{d\phi}\right) \tag{57}$$

where we used with ω_{s0} the synchrotron frequency without an inductive impedance as a convenient parameter. Multiplying this second-order equation with $\dot{\phi}$ gives

$$\frac{\dot{\phi}^2}{2} + \omega_{s0}^2\frac{\phi^2}{2} + \frac{\omega_{s0}^2 h|Z/n|}{\hat{V}\cos\phi_s}(I(\phi) - I(0)) = H' = \text{const}, \tag{58}$$

from which we get the Hamiltonian

$$H = \frac{E}{\omega_0\alpha h}H' = \frac{\omega_s\alpha h E}{2}\left[\left(\frac{\Delta E}{E}\right)^2 + \left(\frac{Q_s}{\alpha h}\right)^2\left(\phi^2 + \frac{h|Z/n|}{\hat{V}\cos\phi_s}(I(\phi) - I(0))\right)\right]. \tag{59}$$

It is interesting to note that the above Hamiltonian depends on the current distribution $I(\phi)$ which is unknown at present. We take the case of electrons which have a Gaussian distribution in energy

$$\psi(\phi, \Delta E) \propto \exp\left(-\frac{1}{2}\left(\frac{\Delta E/E}{\sigma_E/E}\right)^2\right). \tag{60}$$

Furthermore, the distribution is assumed to be stationary and depends therefore only on the Hamiltonian

$$\begin{aligned}\psi(\phi, \Delta E) &= C_4\exp\left(-\frac{H}{\omega_0\alpha h(\sigma_E/E)^2}\right) \\ &= C_4\exp\left(-\frac{\left(\frac{\Delta E}{E}\right)^2 + \left(\frac{Q_s}{\alpha h}\right)^2\left(\phi^2 + \frac{h|Z/n|}{\hat{V}\cos\phi_s}(I(\phi) - I(0))\right)}{2(\sigma_E/E)^2}\right).\end{aligned} \tag{61}$$

Integrating the distribution over ΔE leads to an implicit expression for the current

$$I(\phi) = C_4 \frac{2\pi h I_0}{\sqrt{2\pi}\sigma_E} \exp\left(-\frac{\left(\frac{Q_s}{\alpha h}\right)^2 \left(\phi^2 + \frac{h|Z/n|}{\hat{V}\cos\phi_s}(I(\phi) - I(0))\right)}{2(\sigma_E/E)^2} \right). \tag{62}$$

This represents a transcendent equation for the current distribution $I(\phi)$. We can write it in a more compact form by introducing the current $I(0)$ at the bunch center in the presence of the impedance $|Z/n|$

$$I(0) = C_4 \frac{2\pi h I_0}{\sqrt{2\pi}\sigma_E}, \tag{63}$$

$$I(\phi) = I(0) \exp\left(\frac{\left(\frac{Q_s}{\alpha h}\right)^2 \left(\phi^2 + \frac{h|Z/n|}{\hat{V}\cos\phi_s}(I(\phi) - I(0))\right)}{2(\sigma_E/E)^2} \right). \tag{64}$$

For a vanishing impedance $|Z/n| \to 0$ this leads to the usual Gaussian expression

$$I(\phi)_{Z\to 0} = I_0(\phi) = I_0(0) \exp\left(-\frac{\left(\frac{Q_s}{\alpha h}\right)^2}{2(\sigma_E/E)^2}\phi^2 \right) = \frac{\sqrt{2\pi} h I_0}{\sigma s0} \exp\left(-\frac{1}{2}\left(\frac{\phi}{\sigma_{s0}}\right)^2 \right) \tag{65}$$

where $I_0(0) = \sqrt{2\pi} h I_0/\sigma_{so}$ is the peak current and

$$\sigma_{so} = \left(\frac{Q_s}{\alpha h}\right)^2 \frac{\sigma_E}{E} \tag{66}$$

is the bunch length measured in RF phase angle in the absence of the impedance. Using these quantities and introducing a further parameter ξ which measures the strength of the induced field relative to the guide field of the RF system

$$\xi = \frac{\sqrt{2\pi} h^2 I_0 |Z/n|}{\hat{V}\cos\phi_s \sigma_{s0}^3} \tag{67}$$

we can write the equation in a more compact form

$$I(\phi) e^{\xi I(\phi)/I_0(0)} = I(0) e^{\xi I(0)/I_0(0)} e^{-\phi^2/2\sigma_{s0}^2}. \tag{68}$$

This implicit and transcendent equation determines the self-consistent current distribution $I(\phi)$ for a Gaussian energy distribution in the presence of an inductive impedance. Since the equation is transcendent it has to be solved numerically. Once a solution is obtained the phase-space distribution $\psi(\phi, \Delta E)$, the induced voltage $V(\phi)$ or the synchrotron frequency distribution can also be found. Figure 5 shows such solutions for different values of the strength parameter ξ. Often it is convenient to have an analytic expression for the current distribution even if it is only approximate. We assume now that the impedance strength parameter ξ is small and develop with respect to this parameter

$$I(\phi) \sim \frac{\sqrt{2\pi} h I_0}{\sigma_{s0}} e^{-\frac{\phi^2}{2\sigma_{s0}^2}} \left[1 - \xi \frac{2e^{-\frac{\phi^2}{2\sigma_{s0}^2}} - \sqrt{2}}{2} + \xi^2 \frac{2 - \sqrt{3} - 2\sqrt{2} e^{-\frac{\phi^2}{2\sigma_{s0}^2}} + 3e^{-\frac{\phi^2}{\sigma_{s0}^2}}}{2} + \ldots \right] \tag{69}$$

which can be used for further approximate calculations such as synchrotron frequency distributions, etc.

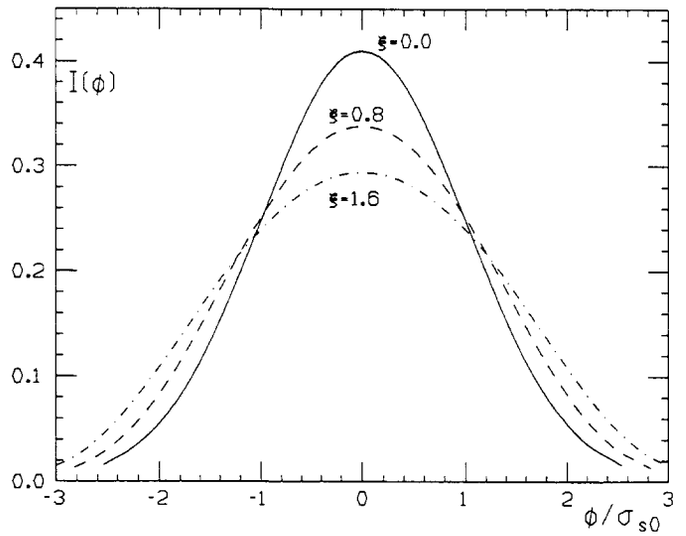


Figure 5: Longitudinal bunch distribution in the presence of an inductive impedance

3.4 Perturbation method

In many practical cases we do not look for a stationary solution but for small deviations from it. The most common examples of this are beam instabilities. The stationary distribution is usually known. Small oscillations around this distribution are considered and the forces they introduce through the wall impedance calculated. We then see whether these forces increase the small initial oscillation leading to an instability or whether they reduce the oscillation indicating stability. Starting with the Vlasov equation

$$\frac{\partial \psi(q,p)}{\partial t} + \dot{q}\frac{\partial \psi(q,p)}{\partial q} + \dot{p}\frac{\partial \psi(q,p)}{\partial p} = 0 \tag{70}$$

we split the distribution $\psi(q,p,t)$ into a stationary part $\psi_0(H)$ and an oscillating perturbation $\psi_1(q,p,t)$ which is assumed to be small

$$\psi(q,p,t) = \psi_0(H) + \psi_1(q,p,t). \tag{71}$$

Furthermore the forces $\dot{p}$ consist of a part due to the guide field like the RF system or the focusing magnets and a small part due to fields induced by the oscillating disturbance $\psi_1(q,p,t)$ in the impedance of the beam surroundings

$$\dot{p} = \dot{p}_0 + \dot{p}_1(\psi_1). \tag{72}$$

Introducing this into the Vlasov equation and making use of the fact that the disturbance is small we can linearize it and obtain an expression involving the disturbance ψ_1. From this equation we find whether the amplitude of the disturbance is growing (instability) or decaying (stability). Usually this last equation determines the frequency of the oscillating disturbance which can be real or complex. Expressing the oscillation with $\exp(j\omega t)$ the situation is stable if the obtained frequency has a positive imaginary part and unstable for a negative imaginary part. On the other hand the form of the disturbance has to be given as an input and is not obtained from the Vlasov equation. Therefore we have to consider a very general disturbance ψ_1 of the stationary distribution, develop it into a complete system of orthogonal modes and check each of them for stability. In practice things are easier since only a few modes are usually relevant. Examples for these procedures can be found in many publications in particular in the CAS lectures [11]

4 EFFECTS OF DAMPING AND RANDOM FORCES

4.1 The Fokker-Planck equation

In accelerators or storage rings the motion of interest is often periodic and the phase-space trajectories are closed curves. In the absence of dissipative forces the phase-space area of these trajectories as well as the number of particles inside, i.e. the phase-space density, is conserved

$$\frac{d\psi(q,p)}{dt} = \frac{\partial\psi(q,p)}{\partial t} + \dot{q}\frac{\partial\psi(q,p)}{\partial q} + \dot{p}\frac{\partial\psi(q,p)}{\partial p} = 0. \tag{73}$$

However, this is no longer guaranteed if dissipative forces, such as energy loss due to synchrotron radiation, or random forces, like the quantum excitation by the emitted photons, are present. We will use an intuitive approach to consider the effect of such forces on the phase-space density and refer to the standard literature such as CAS lectures [1, 12] for rigorous derivations.

We follow here the method used in Ref. [13] and start with the effect of the energy dissipation due to synchrotron radiation which has already been treated in section 3.1 for the phase oscillation. We found there that both canonical variables ϕ and ΔE follow a damped oscillation

$$\begin{aligned} \phi(t) &= \hat{\phi}e^{-\alpha_E t}\cos\left(\omega_{s0}\sqrt{1-(\alpha_E/\omega_{s0})^2}t + \Phi_0\right) \\ \Delta E(t) &= \Delta E_0 e^{-\alpha_E t}\sin\left(\omega_{s0}\sqrt{1-(\alpha_E/\omega_{s0})^2}t + \Phi_0\right). \end{aligned} \tag{74}$$

The two coordinates ϕ and ΔE spiral inwards to smaller values. Usually the damping is relatively small $\alpha_E \ll \omega_{s0}$ and we can still talk of a phase-space area A covered by this trajectory which decreases now at twice the rate of that of the individual coordinates. As a consequence the phase-space area decreases and the phase-space density increases

$$\frac{dA}{dt} = -2\alpha_E A\ ,\quad \frac{d\psi}{dt} = 2\alpha_E\psi. \tag{75}$$

While the energy dissipation due to synchrotron radiation and its replacement by the RF system leads to damping and therefore to an increase of the phase-space density the random forces due to quantum excitation lead to a diffusion and therefore to a reduction of the phase-space density. Assuming random forces which lead primarily to a change of the momentum coordinate p of the particles, as is the case for the quantum excitation, the resulting diffusion is given by the equation

$$\frac{d\psi}{dt} = D\frac{\partial^2\psi}{\partial E^2}. \tag{76}$$

We have not derived it here and refer to textbooks. However the above equation is rather transparent stating that the diffusion leads to a reduction of the density where its second derivative is negative, i.e. where the distribution is curved such that particles can diffuse into neighboring areas. On the other hand the density increases at locations where the distribution is curved such that diffusion from neighboring areas is more likely. We can now express the effect of the dissipative and the random forces on the phase-space density ψ

$$\frac{d\psi(q,p)}{dt} = \frac{\partial\psi(q,p)}{\partial t} + \dot{q}\frac{\partial\psi(q,p)}{\partial q} + \dot{p}\frac{\partial\psi(q,p)}{\partial p} = 2\alpha_E\psi(q,p) + D\frac{\partial^2\psi(q,p)}{\partial p^2}. \tag{77}$$

This is the Fokker-Planck equation in its most simple form which can be regarded as an extension of the Vlasov equation to dissipative and random forces. We write it now for our example of longitudinal motion in a bunch using the coordinates ϕ and ΔE

$$\frac{\partial\psi(\phi,\Delta E)}{\partial t}+\dot{\phi}\frac{\partial\psi(\phi,\Delta E)}{\partial\phi}+\Delta\dot{E}\frac{\partial\psi(\phi,\Delta E)}{\partial\Delta E}=2\alpha_E\psi(\phi,\Delta E)+D\frac{\partial^2\psi(\phi,\Delta E)}{\partial\Delta E^2}. \quad (78)$$

We would like to find a stationary solution of this equation i.e. the distribution for which the damping and the diffusion are in equilibrium and follow the method used in Ref. [13]. Taking the time derivative of the energy deviation from section 3.1 but for a very general RF wave form $V(\phi)$

$$\Delta\dot{E}=\frac{\omega_0}{2\pi}\left(eV(\phi)-\frac{dU}{dE}\Delta E\right)=\Delta\dot{E}_0-2\alpha_E\Delta E. \quad (79)$$

We introduce this into the Fokker-Planck equation and separate the damping and diffusion terms

$$\frac{\partial\psi}{\partial t}+\dot{\phi}\frac{\partial\psi}{\partial\phi}+\Delta\dot{E}_0\frac{\partial\psi}{\partial\Delta E}=2\alpha_E\Delta E\frac{\partial\psi}{\partial\Delta E}+2\alpha_E\psi+D\frac{\partial^2\psi}{\Delta E^2}. \quad (80)$$

The left hand side of the above equation has no damping and diffusion terms and is just the Vlasov equation as we used it in the earlier section but for a general RF wave form and must therefore vanish. For the right hand side we get

$$2\alpha_E\left(\psi+\Delta E\frac{\partial\psi}{\partial\Delta E}+\frac{D}{2\alpha_E}\frac{\partial^2\psi}{\partial E^2}\right)=2\alpha_E\frac{\partial}{\partial\Delta E}\left(\Delta E\psi+\frac{D}{2\alpha_E}\frac{\partial\psi}{\partial\Delta E}\right)=0. \quad (81)$$

From this it follows that the last expression in the brackets does not depend on ΔE

$$\frac{\partial\psi}{\partial\Delta E}+2\frac{\alpha_E}{D}\psi\Delta E=C_5=f(\phi). \quad (82)$$

It has the solution

$$\psi(\phi,\Delta E)=F(\phi)\exp\left(-\frac{1}{2}\left(\frac{\Delta E}{\sigma_E}\right)^2\right)\text{ with }\sigma_E^2=\frac{D}{2\alpha_E}. \quad (83)$$

The above derivation was carried out for a very general RF wave form and resulted in a Gaussian distribution in energy determined only by damping and diffusion, a result we used already before. The longitudinal distribution in ϕ, however, is not determined by damping and diffusion alone but depends on the RF wave form. These statements are only correct if the random force effects the energy directly, as in the case of the emission of a synchrotron radiation photon, and if neither the damping nor the quantum excitation depends on the energy deviation ΔE itself. For all the examples treated here these conditions are fulfilled in very good approximation. There are some exotic cases where this is not so, like the (usually very weak) quantum excitation in quadrupoles which depend on the displacement of the particle, or for some non-linear wigglers in which the damping becomes a function of the displacement [12, 14]

In the above derivation of the Gaussian distribution we did not calculate the rms energy spread σ_E but related it just to the diffusion constant D which should be determined from the properties of the synchrotron radiation. We will do that in the next subsection using a different method.

4.2 Campbell's theorem - energy spread due to synchrotron radiation emission

In the previous section it was shown that the quantum fluctuation and the damping provided by the synchrotron radiation leads to a Gaussian energy distribution of an electron beam in a storage ring. We will now use a direct method which does not give the form of the distribution but rather its variance or its rms value. This will be done by using Campbell's theorem [1, 15]. We will not prove the theorem here but refer to the references mentioned above. We follow the method used by M. Sands [16] who first derived the energy spread in electron storage rings.

We assume that we have a linear system and know its response to a δ-pulse excitation of amplitude a at the time t_0

$$x(t) = ag(t - t_0) \ , \ \ g(\tau) = 0 \text{ for } \tau < 0. \tag{84}$$

If such excitations occur randomly in time with average frequency f but with the same amplitude a, the resulting average of the square of the excursion (variance) is

$$\langle x^2 \rangle = a^2 f \int_{-\infty}^{\infty} g^2(t - t_0) dt. \tag{85}$$

This can be generalized for the case where the amplitude also has a random distribution of the form $\dot{n}(a)$

$$\langle x^2 \rangle = \int_0^{\infty} a^2 \dot{n}(a) da \int_{-\infty}^{\infty} g^2(t - t_0) dt. \tag{86}$$

This is Campbell's theorem which we now wish to apply to the case of quantum excitation by synchrotron radiation.

We have first to recapitulate some properties of the synchrotron radiation. The total power emitted by a particle of charge e and energy $m_0c^2\gamma$ going on a trajectory with curvature $1/\rho$ through a magnet is

$$P_\gamma = \frac{2cr_e m_0 c^2 \gamma^4}{3\rho^2}, \tag{87}$$

where r_e is the classical electron radius. The spectral distribution in photon energy ϵ is

$$\frac{d\dot{n}(\epsilon)}{d\epsilon} = \frac{P_\gamma}{\epsilon_c^2} \frac{\epsilon_c}{\epsilon} S(\epsilon/\epsilon_c) \text{ with } S(\epsilon/\epsilon_c) = \frac{9\sqrt{3}}{8\pi} \frac{\epsilon}{\epsilon_c} \int_{\epsilon/\epsilon_c}^{\infty} K_{5/3}(x) dx, \tag{88}$$

where we used the critical energy ϵ_c

$$\epsilon_c = \frac{3}{4\pi} \frac{\lambda_{Comp}}{\rho} m_0 c^2 \gamma^3 \ , \ \ \lambda_{Comp} = \frac{h}{m_0 c}, \tag{89}$$

the Compton wave length λ_{Comp}, Planck's constant h and the modified Bessel function $K_{5/3}$ of order 5/3. We have to know also the variance of this spectrum

$$\int_0^{\infty} \epsilon^2 \dot{n}(\epsilon) d\epsilon = \frac{55}{8\, 3^{3/2}} P_\gamma \epsilon_c \tag{90}$$

for which we give here only the result.

Next we need the response of the particle to a δ-pulse excitation which we can get from Eq. (74) by setting the initial energy deviation to the loss due to the emission of a photon $\Delta E = -\epsilon$ and assume weak damping

$$g(t - t_0) = \Delta E = -\epsilon e^{-\alpha_E (t - t_0)} \cos(\omega_{s0}(t - t_0)) \tag{91}$$

the square of which can easily be integrated

$$\int_{t_0}^{\infty} g^2(t-t_0)dt = \int_0^{\infty} g^2(t)dt = \frac{1}{4\alpha_E}. \tag{92}$$

To get the variance of the energy distribution we have to insert the variance of the photon energy spectrum and the integral over the square of the response function into Campbell's theorem

$$\sigma_E^2 = \langle \Delta E^2 \rangle = \int_0^{\infty} \epsilon^2 \dot{n}(\epsilon) d\epsilon \int_0^{\infty} g^2(t)dt = \frac{55}{8\,3^{3/2}} \frac{P_\gamma \epsilon_c}{4\alpha_E} \tag{93}$$

which is the well known expression for the energy spread in an electron storage ring.

References

[1] J. Jowett; Phys. of Part. Accel, Proc. SLAC 85 and Fermilab 84 Schools, AIP Conf. Proc. 153 (1987) 877.

[2] C.N. Lashmore-Davis; Proc. 1985 CAS CERN Accel. School on Advanced Accel. Physics, CERN 87-03 Vol. I (1987) 219.

[3] J.S. Bell; Proc. 1985 CAS CERN Accel. School on Advanced Accel. Physics, CERN 87-03 Vol. I (1987) 5.

[4] A. Hofmann, F. Pedersen; Proc. of the 1979 Part. Accel. Conf., IEEE Trans. on Nucl. Sci. NS 26-3, 3526 (1979).

[5] A.W. Chao; Lecture Notes in Physics 296, Proc. of a Topical Course Frontiers of Particle Beams, Joint US-CERN School of Particle Accel. 1986, Springer (1988) 51.

[6] J.D. Lawson; CERN 76-09 (1976).

[7] Bruck; Accelerateurs Circulaires de Particules, Presses Universitaires de France 1966.

[8] A. Hofmann, S. Myers; Proc. 11th Int. Conf. on High Energy Accel., CERN 1980, (Birkhäuser, Basel, 1980), 610.

[9] A. Hofmann, J.R. Maidment; CERN LEP-Note 168 (1979).

[10] J. Haïssinski; Il Nuovo Cimento, Vol. 18 B (1973) 72.

[11] J.L. Laclare; Proc. 1985 CAS CERN Accel. School on Advanced Accel. Physics, CERN 87-03 Vol. I (1987) 264.

[12] J. Jowett; Proc. 1985 CAS CERN Accel. School on Advanced Accel. Physics, CERN 87-03 Vol. II (1987) 570.

[13] B. Zotter; CERN Academic Training lecture notes (1981).

[14] A. Hofmann, J. Jowett, S. Myers; Proc. of the 1981 Part. Accel. Conf., IEEE Trans. on Nucl. Sci. NS 28-3, 2392 (1981).

[15] S.O. Rice; Selected Papers on Noise and Stochastic Processes p. 133, edited by N. Wax; Dover 1954.

[16] M. Sands; Phys. Rev. 87 (1955) 470.

LANDAU DAMPING

A. Hofmann,
CERN, Geneva, Switzerland

Abstract

The concept of Landau damping is introduced from an experimental point of view. We start by studying the response of a beam to an external excitation. We are mainly interested in the center-of-mass motion after the excitation and much less in the one of the individual particles. For a harmonic excitation this response is called transfer function and for a pulse excitation it is called Green function. This response will depend on the external focusing forces acting on the beam and also in the distribution of the resonance frequencies of the individual particles in the beam. A large spread in these frequencies will lead to phase mixing after a pulse excitation and can result in a decrease of the center-of-mass motion. Once the response of the beam is known we can study the stability of a beam in the presence of an impedance. The latter is caused by the beam surroundings in which the charge and current of the beam induce some fields. These can act back on the beam and provide a feed-back mechanism which can lead to an instability. Since the impedance does not resolve the individual particles the induced fields are only sensitive to the motion of the center of mass or of a simple mode. From the known response to an excitation we can calculate the behavior of the beam in the presence of fields induced in the impedance. If a large spread of frequencies provides a fast decay of the center-of-mass response the effect of the impedance can be compensated and the instability is suppressed. This is called Landau damping. We treat first the transverse stability of an unbunched beam. A spread in betatron frequency is provided by the momentum spread of the beam combined with the momentum dependence of the betatron frequency. The coherent (center-of-charge) response of such a beam to a transverse excitation is calculated. For the case of a harmonic excitation a resistive term is found which absorbs energy while for a pulse excitation a decaying coherent oscillation is obtained. This effect is responsible for the Landau damping present in the case where such a beam interacts with the transverse impedance of its surroundings. Next we consider the longitudinal stability of this unbunched beam. We calculate its response to an energy disturbance of all the particles and get the longitudinal transfer function for a longitudinal excitation and the longitudinal Green function for a pulse excitation. This case is more complicated than the transverse one and is best treated with the Vlasov equation. From the known response we obtain a criterion for the longitudinal stability of an unbunched beam.

1 TRANSVERSE LANDAU DAMPING FOR AN UNBUNCHED BEAM

1.1 Introduction

We consider now a coasting (unbunched beam) and investigate its transverse stability. There is no clear agreement as to which effects can be labelled as Landau damping. Let us take the case of a particle beam which has a spread in betatron frequency. If we give this whole beam a delta pulse deflection it will execute a betatron oscillation. In the absence of any incoherent damping (such as established by synchrotron radiation)

the individual particles will continue to oscillate. However, due to their different individual (incoherent) betatron oscillation frequencies the center-of-mass motion (coherent oscillation) will usually have a decreasing amplitude. This effect is called phase mixing or filamentation. Since in most cases we observe only the coherent oscillation, we have the impression that the excited oscillation is damped. In reality only the coherent or the center-of-mass motion is damped while the incoherent oscillations of the particles are unchanged. Should this effect be called Landau damping? Many people would say no. However, this effect of phase mixing is closely related to Landau damping and we have to understand it in order to investigate this damping mechanism. Landau damping is one of those effects which can be treated in different ways, each giving a correct description of the effect [1, 2, 3]. In this report we will look at it from an experimental point of view and relate it to the beam response to a harmonic or pulse excitation which can actually be measured. This will be used later to treat the stability provided by Landau damping for a beam interacting with the impedance or its surroundings. Since the particles in a beam behave in many respects like harmonic oscillators we start our investigation with the response of a set of such oscillators to an excitation.

1.2 Response of oscillators to a harmonic excitation

1.2.1 Introduction

We consider a set of harmonic oscillators having different individual resonant frequencies ω_β. The distribution of the particles with respect to these frequencies, i.e. the number of particles per frequency band, is defined as

$$f(\omega_\beta) = \frac{1}{N}\frac{dN_\beta}{d\omega_\beta} \text{ with } \int_{-\infty}^{\infty} f(\omega_\beta)d\omega_\beta = 1 \text{ and } f(\omega) = f(-\omega) \tag{1}$$

and assumed to be symmetric. We use here positive and negative frequencies for later applications involving Fourier transforms. It should, therefore, be noted that the distribution $f(\omega_\beta)$ appears at positive and negative ω_β. We will assume the total number of particles N to be large such that the distribution $f(\omega_\beta)$ can be regarded as being continuous. Next we apply a harmonic force to all the particles resulting in an acceleration

$$G(t) = \hat{G}e^{-i\omega t} \tag{2}$$

with a frequency ω being inside or close to the frequency distribution. All particles will undergo a forced oscillation with the frequency of excitation ω but the phase of the individual particle oscillation will depend on the difference $\omega - \omega_\beta$ between the exciting frequency and the individual resonant frequency.

1.2.2 Response calculated by complex notation after exciting for a long time

We calculate the response of the oscillators first by using a complex notation giving the response after a long time. This results in a very compact derivation but at the cost of a loss of physical transparency. The response of a single oscillator with resonant frequency ω_β to a harmonic excitation with frequency ω is given by the differential equation

$$\frac{d^2y}{dt^2} + \omega_\beta^2 y = \hat{G}e^{-i\omega t}. \tag{3}$$

Seeking solutions of the form $y = \hat{y}\exp(-i\omega t)$ leads to

$$-\omega^2\hat{y} + \omega_\beta^2\hat{y} = \hat{G} \ , \ \hat{y} = \frac{\hat{G}}{\omega_\beta^2 - \omega^2}. \tag{4}$$

Factorizing the denominator and expressing it in partial fractions gives

$$\hat{y} = \frac{\hat{G}}{(\omega_\beta - \omega)(\omega_\beta + \omega)} = \frac{\hat{G}}{2\omega}\left[\frac{1}{(\omega_\beta - \omega)} - \frac{1}{\omega_\beta + \omega}\right]. \tag{5}$$

We take now the whole set of oscillators having the individual resonant frequencies ω_β with distribution $f(\omega_\beta)$ which we consider to be continuous. We do not want to know the response of the individual oscillators but only the response of their center of mass i.e. the dipole moment divided by the number of particles. In an accelerator this is usually the only quantity which can be measured by position monitors or seen by the transverse impedance. This center-of-mass excursion $\langle y \rangle$ of the response is obtained by averaging the individual responses Eq. (5) weighted with the distribution $f(\omega_\beta)$

$$\langle y(\omega) \rangle = \frac{\hat{G}e^{-i\omega t}}{2\omega}\left[\int_{-\infty}^{\infty} \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta - \int_{-\infty}^{\infty} \frac{f(\omega_\beta)}{\omega_\beta + \omega} d\omega_\beta\right]. \tag{6}$$

The dipole moment of the beam is simply $D = N\langle y \rangle$. Using the symmetry property of the distribution $f(\omega_\beta) = f(-\omega_\beta)$ gives

$$\int_{-\infty}^{\infty} \frac{f(\omega_\beta)}{(\omega_\beta - \omega)} d\omega_\beta = -\int_{-\infty}^{\infty} \frac{f(\omega_\beta)}{\omega_\beta + \omega} d\omega_\beta \tag{7}$$

which results for the response

$$\langle y(\omega) \rangle = \frac{\hat{G}e^{-i\omega t}}{\omega} \int_{-\infty}^{\infty} \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta. \tag{8}$$

If the exciting frequency ω lies inside the distribution $f(\omega_\beta)$ the integration contains a pole which leads to a residue and a principle value integral

$$\langle \hat{y}(\omega) \rangle = \frac{\hat{G}}{\omega}\left[\pm i\pi f(\omega) + \mathrm{PV}\int \frac{f(\omega_\beta)}{(\omega_\beta - \omega)} d\omega_\beta\right]. \tag{9}$$

The integral has an imaginary part of ambiguous sign and a real principle value integral defined as

$$\mathrm{PV}\int_{-\infty}^{\infty} \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta = \lim_{\epsilon \to 0}\left[\int_{-\infty}^{-\epsilon} \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta + \int_{\epsilon}^{\infty} \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta\right]. \tag{10}$$

We have to investigate the physical meaning of this. To clarify this point we calculate the center-of-charge velocity $\langle \dot{y} \rangle$ which is related to the center-of-charge position obtained above by

$$\langle \dot{y} \rangle = -i\omega\langle y \rangle e^{-i\omega t} = \hat{G}e^{-i\omega t}\left[\pi f(\omega) - i\mathrm{PV}\int \frac{f(\omega)}{\omega_\beta - \omega} d\omega_\beta\right]. \tag{11}$$

This last equation gives a relation between the velocity response of the center-of-charge $\langle \dot{y} \rangle$ and the applied acceleration G. This response has a real part for which the velocity and the acceleration are in phase leading to an absorption of energy, and an imaginary part where velocity and acceleration are out of phase indicating no energy exchange. For this reason the two parts are also called resistive and reactive. We have chosen the positive sign for the real part assuming that the beam absorbs energy from the exciting acceleration. However, this point has to be clarified later. Since the real part absorbs energy it can lead to damping of the coherent oscillation which is (often) called Landau damping. It is interesting to note that this term is proportional to the density of the distribution $f(\omega)$ at the exciting frequency ω. It vanishes, therefore, if the exciting frequency lies outside the incoherent frequency distribution.

1.3 Time evolution of the response using real notation

We saw that the resistive part of the beam response to a harmonic excitation is due to the residue of the integral. Since it is this term which leads to Landau damping we would like to better understand the underlying physics. For this purpose we now discuss the time evolution of the response using real notation [4]. This leads to derivations which are more lengthy but more transparent. The differential equation describing the excitation of the beam is in analogy with Eq. (3)

$$\frac{d^2y}{dt^2} + \omega_\beta^2 y = \hat{G}\sin(\omega t). \tag{12}$$

This equation has a homogeneous solution

$$y_h = A\sin(\omega_\beta t) + B\cos(\omega_\beta t) \tag{13}$$

and a particular solution

$$y_p = \frac{\hat{G}}{\omega_\beta^2 - \omega^2}\sin(\omega t). \tag{14}$$

The general solution is a combination of the two

$$y = A\sin(\omega_\beta t) + B\cos(\omega_\beta t) + \frac{\hat{G}}{\omega_\beta^2 - \omega^2}\sin(\omega t). \tag{15}$$

To determine the integration constants A and B we have to specify the initial conditions. We take the case where the oscillators are at rest until the time $t = 0$ when the excitation starts

$$y(0) = \dot{y}(0) = 0 \;\rightarrow\; A = -\frac{\omega}{\omega_\beta}\frac{\hat{G}}{(\omega_\beta^2 - \omega^2)}\;;\;\; B = 0. \tag{16}$$

This gives for the general solution

$$\begin{aligned} y &= \frac{\hat{G}}{\omega_\beta^2 - \omega^2}\left(\sin(\omega t) - \frac{\omega}{\omega_\beta}\sin(\omega_\beta t)\right) \\ &= \frac{\hat{G}}{2\omega}\left(\frac{1}{(\omega_\beta - \omega)} - \frac{1}{\omega_\beta + \omega}\right)\left(\sin(\omega t) - \frac{\omega}{\omega_\beta}\sin(\omega_\beta t)\right). \end{aligned} \tag{17}$$

We differentiate with respect to t to get the velocity of the single oscillator response

$$\dot{y} = \frac{\hat{G}}{2}\left(\frac{\cos(\omega t) - \cos(\omega_\beta t)}{(\omega_\beta - \omega)} - \frac{\cos(\omega t) - \cos(\omega_\beta t)}{\omega_\beta + \omega}\right) \tag{18}$$

and rewrite this equation by substituting ω_β with $\omega_\beta = \omega + (\omega_\beta - \omega)$ in the first term and $\omega_\beta = (\omega_\beta + \omega) - \omega$ in the second term of the parenthesis,

$$\begin{aligned} \dot{y} &= \frac{\hat{G}}{2}\left[\cos(\omega t)\left(\frac{1 - \cos((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} - \frac{1 - \cos((\omega_\beta + \omega)t)}{\omega_\beta + \omega}\right)\right. \\ &\left. + \sin(\omega t)\left(\frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} + \frac{\sin((\omega_\beta + \omega)t)}{\omega_\beta + \omega}\right)\right]. \end{aligned} \tag{19}$$

The above equation gives the velocity response for a single oscillator with resonant frequency ω_β. For the first term in the square bracket velocity and acceleration are basically

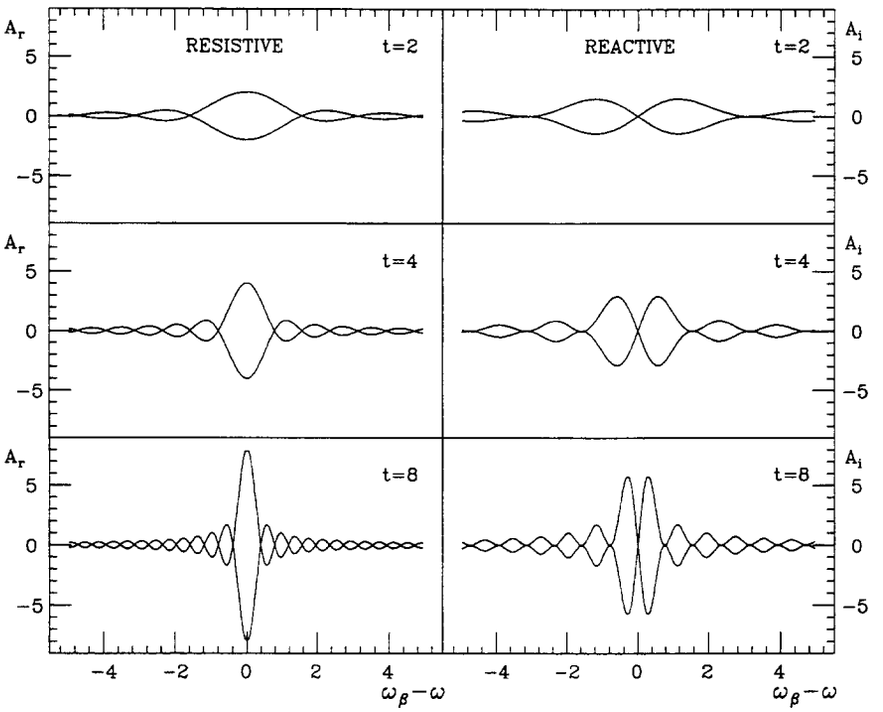


Figure 1: Time evolution of the beam response to a harmonic excitation

out of phase while for the second term they are in phase. This statement is not exact since both terms also contain an oscillatory term with frequency $(\omega_\beta - \omega)$ which has to be discussed further. Before we integrate to obtain the center-of-mass response we discuss these two components of the single oscillator response Eq. (20). We concentrate on a region in the vicinity of the exciting positive frequency ω, i.e. $\omega_\beta \approx \omega$ where the first term inside each of the two round brackets is dominant. A corresponding discussion for $\omega_\beta \approx -\omega$ could easily be carried out in addition. In Fig. 1 the envelope of the executed oscillation is shown as a function of the difference $(\omega_\beta - \omega)$ between the resonant and the exciting frequency for different times t after the start of the excitation. As this time increases particles oscillating with opposite phase are close together in frequency leading to some cancellation in the integration to follow. For the resistive (in phase) term the oscillators with resonant frequency close to the exciting frequency gain large amplitudes.

To calculate the center of mass response we have to integrate the single particle response Eq. (20) over the resonant frequency ω_β weighted with the distribution $f(\omega_\beta)$. Using again the symmetry property of the distribution $f(\omega_\beta) = f(-\omega_\beta)$ we can show that

$$\int_{-\infty}^{\infty} \frac{1-\cos((\omega_\beta-\omega)t)}{(\omega_\beta-\omega)} d\omega_\beta = -\int_{-\infty}^{\infty} \frac{1-\cos(\omega_\beta+\omega t)}{\omega_\beta+\omega} d\omega_\beta \tag{20}$$

and

$$\int_{-\infty}^{\infty} \frac{\sin((\omega_\beta-\omega)t)}{(\omega_\beta-\omega)} d\omega_\beta = \int_{-\infty}^{\infty} \frac{\sin(\omega_\beta+\omega t)}{\omega_\beta+\omega} d\omega_\beta \tag{21}$$

which gives for the velocity response

$$\begin{aligned} \langle \dot{y} \rangle &= \frac{\hat{G}\cos(\omega t)}{2\pi} \int_{-\infty}^{\infty} f(\omega_\beta) \frac{1-\cos((\omega_\beta-\omega)t)}{\omega_\beta-\omega} d\omega_\beta \\ &+ \frac{\hat{G}\sin(\omega t)}{2\pi} \int_{-\infty}^{\infty} f(\omega_\beta) \frac{\sin((\omega_\beta-\omega)t)}{\omega_\beta-\omega} d\omega_\beta. \end{aligned} \tag{22}$$

We start with the out of phase (reactive) part which has an oscillatory term of the form $(\cos((\omega_\beta - \omega)t))$. As the time t increases this term will have opposite phase for smaller and smaller frequency differences in ω_β. The integration over ω_β will therefore vanish and we can replace the numerator of the reactive term by unity $(1 - \cos((\omega_\beta - \omega)t) \rightarrow 1)$ except for the oscillators with $\omega_\beta \approx \omega$. This central part $\omega_\beta \approx \omega$ becomes more and more narrow as the time t increases and we can replace the integral over the reactive term by the principle value integral Eq. (10)

$$\lim_{t\to\infty} \int f(\omega_\beta) \frac{1 - \cos((\omega_\beta - \omega)t)}{\omega_\beta - \omega} d\omega_\beta = \int \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta \tag{23}$$
$$= \lim_{\epsilon\to 0} \left(\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty} \right) = \mathrm{PV} \int \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta.$$

The resistive (in phase) term of Eq. (20) contains an oscillatory term under the integral of the form

$$\frac{\sin((\omega_\beta - \omega)t)}{\omega_\beta - \omega} = t \frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)t} = t\,\mathrm{sinc}((\omega_\beta - \omega)t). \tag{24}$$

For $\omega_\beta \neq \omega$ and large t the above expression oscillates with opposite phase for small changes in ω_β. The integration over this frequency will vanish as long as the distribution $f(\omega_\beta)$ is sufficiently smooth. For $\omega_\beta \approx \omega$ and very large t the function $\mathrm{sinc}((\omega_\beta - \omega)t)$ will be about unity and the above expression grows with increasing time t without limits. Furthermore the integral [5]

$$\int_{-\infty}^{\infty} \frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} d\omega_\beta = \pi \tag{25}$$

is independent of t. We can therefore replace Eq. (24) by the δ-function

$$\lim_{t\to\infty} \frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} = \pi\delta(\omega_\beta - \omega). \tag{26}$$

Collecting the results obtained for the reactive and resistive part of the beam response we get

$$\lim_{t\to\infty} \langle \dot{y} \rangle = \hat{G} \left[\sin(\omega t)\pi f(\omega) + \cos(\omega t)\mathrm{PV} \int \frac{f(\omega_\beta)}{(\omega_\beta - \omega)} d\omega_\beta \right]. \tag{27}$$

This is the same result as that already obtained more quickly using complex notation. However, in this subsection we learned basically three things:

- The result Eq. (11) is only correct if the excitation has lasted for a long time. How long this time has to be depends on the resolution with which the distribution $f(\omega_\beta)$ has to be considered. If this distribution does not change significantly over a frequency range of $\Delta\omega_\beta$ it is sufficient to excite for a time $t \gg 1/\Delta\omega_\beta$.
- The sign of the residue can be determined from the initial conditions. Usually one excites a set of oscillators being initially at rest. However, it is in principle possible to have a set of particles oscillating initially with a particular distribution in amplitude and phase such that the exciting acceleration takes energy out of the beam. Since in the complex notation the initial conditions were not specified both possibilities are contained in the beam response equation.
- From Fig. 1 it is clear that a few oscillators, having resonant frequencies close to the exciting frequency $\omega_\beta \approx \omega$, attain large amplitudes. The energy absorbed by the beam from the exciter goes, therefore, into large oscillation amplitudes obtained by a small fraction of the oscillators.

1.4 Response to a δ-pulse excitation

We take now the set of oscillators with resonant frequencies ω_β and distribution $f(\omega_\beta)$ and at the time $t = 0$ give every particle a kick such that it has an initial velocity $\dot{y}(0) = \dot{y}_0$. Each particle will execute a harmonic oscillation with it own resonant frequency ω_β

$$\dot{y}(t) = \dot{y}_0 \cos(\omega_\beta t)\,,\ y(t) = \frac{\dot{y}_0}{\omega_\beta} \sin(\omega_\beta t). \tag{28}$$

To get the response of the center-of-charge we have to average the single particle response over the distribution [6]

$$\langle \dot{y}(t) \rangle = \dot{y}_0 \int_{-\infty}^{\infty} f(\omega_\beta) \cos(\omega_\beta t) d\omega_\beta. \tag{29}$$

This is the inverse cosine Fourier transform of the oscillator distribution $f(\omega_\beta)$ in frequency, a result well known in fields like filter theory etc.

1.5 Relation between the two responses

For the velocity response to a harmonic excitation we obtain from the previous section Eq. (11)

$$r(\omega) = \frac{\langle \dot{y}(\omega) \rangle}{G} = \left[\pi f(\omega) - i\mathrm{PV} \int \frac{f(\omega_\beta)}{(\omega_\beta - \omega)} d\omega_\beta \right] = r_r(\omega) + i r_i(\omega). \tag{30}$$

This response of an oscillator set to a harmonic excitation is called the *transfer function.*

The pulse response, also called *Green function*, is according to Eq. (29)

$$g(t) = \frac{\langle \dot{y}(t) \rangle}{\dot{y}_0} = \int f(\omega_\beta) \cos(\omega_\beta t) d\omega_\beta \tag{31}$$

Comparing these two responses one suspects that they are related to each other by the Fourier transform and its inverse

$$\begin{aligned} F(\omega) &= \mathcal{F}[f(t)] = \int f(t) \mathrm{e}^{i\omega t} dt = \int f(t) \cos(\omega t) dt + i \int f(t) \sin(\omega t) dt \\ f(t) &= \mathcal{F}^{-1}[F(\omega)] = \frac{1}{2\pi} \int F(\omega) \mathrm{e}^{-i\omega t} d\omega. \end{aligned} \tag{32}$$

The integration goes usually from $-\infty$ to ∞, however for our case it has to cover just the distribution $f(\omega_\beta)$ for the Fourier transform and it goes from 0 to ∞ for its inverse since the pulse response vanishes for negative times. Since we used an excitation of the form $G(t) = \hat{G}\mathrm{e}^{-i\omega t}$ we use in the Fourier transform $\mathrm{e}^{=i\omega t}$. The factor in front of the integration is chosen for convenience to be unity for the transform and $1/2\pi$ for the inverse. This choice is arbitrary but the product of the two factors has to be $1/2\pi$.

We take the cosine Fourier transform of the Green function

$$\mathcal{F}_{\cos}[g(t)] = \int_0^{\infty} g(t) \cos(\omega t) dt = \int_{-\infty}^{\infty} f(\omega_\beta) \int_0^{\infty} \cos(\omega_\beta t) \cos(\omega t) dt d\omega_\beta. \tag{33}$$

The integration over t results in the δ-function

$$\int_0^{\infty} \cos(\omega_\beta t) \cos(\omega t) d\omega_\beta = \frac{1}{2} \int_{-\infty}^{\infty} \cos(\omega_\beta t) \cos(\omega t) d\omega_\beta = \pi \delta(\omega_\beta - \omega). \tag{34}$$

This gives

$$\int_0^\infty g(t)\cos(\omega t)dt = \pi \int_{-\infty}^\infty f(\omega_\beta)\delta(\omega_\beta - \omega)d\omega_\beta = \pi f(\omega) = r_r(\omega), \tag{35}$$

For the imaginary part $r_i(\omega)$ we get

$$\begin{aligned} r_i &= -\mathrm{PV}\int_{-\infty}^\infty \frac{f(\omega_\beta)}{(\omega_\beta - \omega)}d\omega_\beta = -\frac{1}{\pi}\mathrm{PV}\int_{-\infty}^\infty \frac{r_r(\omega_\beta)}{(\omega_\beta - \omega)}d\omega_\beta \\ &= -\frac{1}{\pi}\mathrm{PV}\int_0^\infty g(t)\int_{-\infty}^\infty \frac{\cos(\omega_\beta t)}{(\omega_\beta - \omega)}d\omega_\beta dt. \end{aligned} \tag{36}$$

The integral over ω_β is [5]

$$\int_{-\infty}^\infty \frac{\cos(\omega_\beta t)}{(\omega_\beta - \omega)}d\omega_\beta = -\pi\sin(\omega t) \tag{37}$$

results for the imaginary part of the transfer function

$$r_i(\omega) = \int_0^\infty g(t)\sin(\omega t)dt = \mathcal{F}_{\sin}[g(t)], \tag{38}$$

which is the sine Fourier transform of the pulse response $g(t)$. We conclude therefore that the transfer function is the Fourier transformation of the Green function and vice versa

$$r(\omega) = \int_{-\infty}^\infty g(t)e^{i\omega t}dt = \int_0^\infty g(t)e^{i\omega t}dt \text{ and } g(t) = \frac{1}{2\pi}\int_{-\infty}^\infty r(\omega)e^{-i\omega t}d\omega \tag{39}$$

and the transfer function is related to the particle distribution $f(\omega_\beta)$ in frequency

$$r(\omega) = r_r(\omega) + ir_i(\omega). = \frac{\dot{y}(\omega)}{G} = \pi f(\omega) - i\mathrm{PV}\int \frac{f(\omega_\beta)}{(\omega_\beta - \omega)}d\omega_\beta \tag{40}$$

The relation between the real and imaginary part of the transfer function

$$r_i(\omega) = -\int_{-\infty}^\infty \frac{f(\omega_\beta)}{(\omega_\beta - \omega)}d\omega_\beta = -2\int_{-\infty}^\infty \frac{r_r(\omega_\beta)}{(\omega_\beta - \omega)}d\omega_\beta \tag{41}$$

is called a Hilbert transform and is a consequence of the fact that the pulse response $g(t)$ is real and that it vanishes for negative arguments.

The transfer function and the Green function of a beam can be measured directly [7, 8, 9]. It is a powerful method to find the distribution of frequencies and to estimate the stability of a beam.

1.6 Examples of transfer and Green functions

1.6.1 Symmetric distributions centered around the origin

We give now two examples of frequency distributions $f(\omega_\beta)$ and calculate their responses to a harmonic excitation (transfer function $r(\omega)$) and to a pulse excitation (Green function $g(t)$). In the figures we also give the stability diagram which will be discussed later.

We start with a *Gaussian* distribution

$$f(\omega_\beta) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \exp\left(-\frac{\omega_\beta^2}{2\sigma_\omega^2}\right) \text{ with } \int_{-\infty}^{\infty} f(\omega_\beta) d\omega_\beta = 1. \tag{42}$$

We calculate first the Green function

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \int_{-\infty}^{\infty} \exp\left(-\frac{\omega_\beta^2}{2\sigma_\omega^2}\right) \cos(\omega_\beta t) d\omega_\beta = \exp\left(-\frac{\sigma_\omega^2 t^2}{2}\right). \tag{43}$$

The transfer function is best obtained by a Fourier transform of the Green function. For the real part we find

$$r_r(\omega) = \int_0^{\infty} \exp\left(-\frac{\sigma_\omega^2 t^2}{2}\right) \cos(\omega t) dt = \frac{\sqrt{2\pi}}{2\sigma_\omega} \exp\left(-\frac{\omega^2}{2\sigma_\omega^2}\right). \tag{44}$$

The imaginary part is

$$r_i(\omega) = \int_0^{\infty} \exp\left(-\frac{\sigma_\omega^2 t^2}{2}\right) \sin(\omega t) dt = \frac{\sqrt{2}}{\sigma_\omega} \exp\left(-\frac{\omega^2}{2\sigma_\omega^2}\right) \int_0^{\omega/\sqrt{2}\sigma_\omega} \exp(t'^2) dt', \tag{45}$$

where the last integral is called Dawson integral [10]. The complex transfer function becomes

$$r(\omega) = \frac{\sqrt{2\pi}}{2\sigma_\omega} \exp\left(-\frac{\omega^2}{2\sigma_\omega}\right) \left[1 + i\frac{2}{\sqrt{\pi}} \int_0^{\omega/\sqrt{2}\sigma_\omega} \exp(t'^2) dt'\right]. \tag{46}$$

The transfer function can also be represented by giving its amplitude $A(\omega)$ and phase $\phi(\omega)$ instead of its real and imaginary part. Both representations of the transfer function as well as the Green function of a Gaussian distribution are shown in Fig. 2.

Next we investigate the response of a set of particles having a *Lorentz* distribution in frequency

$$f(\omega_\beta) = \frac{2a}{a^2 + \omega_\beta^2}. \tag{47}$$

The Green function $g(t)$ is

$$g(t) = \int_{-\infty}^{\infty} \frac{2a}{a^2 + \omega_\beta^2} \cos(\omega_\beta t) d\omega_\beta = \pi e^{-at}. \tag{48}$$

From this we get the transfer function $r(\omega)$ which can be expressed by its real and imaginary part or by its in amplitude $A(\omega)$ and phase $\phi(\omega)$

$$r(\omega) = 2\pi \frac{a + i\omega}{a^2 + \omega^2}, \; A(\omega) = \frac{2\pi}{\sqrt{a^2 + \omega^2}}, \; \tan(\phi) = \frac{\omega}{a}. \tag{49}$$

This Green function and the transfer function are shown in Fig. 3. It is interesting to note that for the Lorentz distribution the Green function is an exponential, i.e. after a pulse excitation the center of mass of this beam will move exponentially to zero.

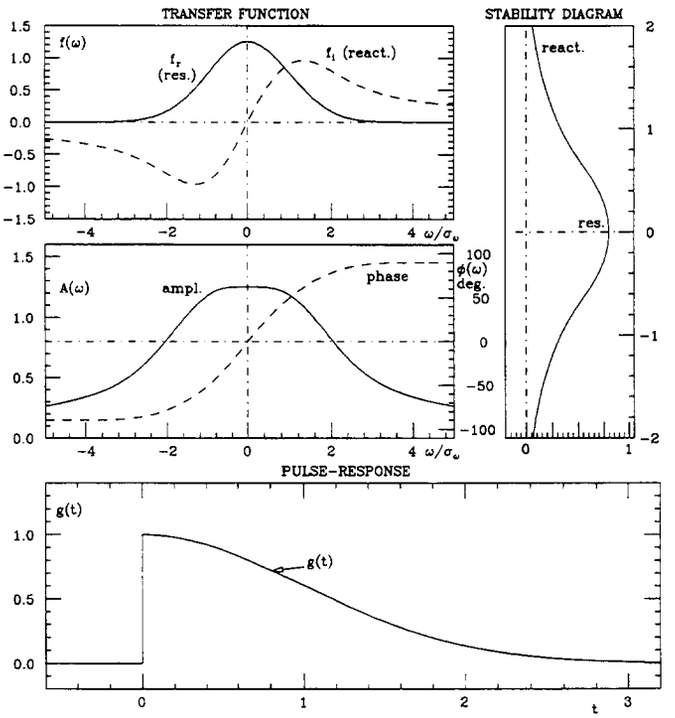


Figure 2: Response of oscillators with a Gaussian frequency distribution

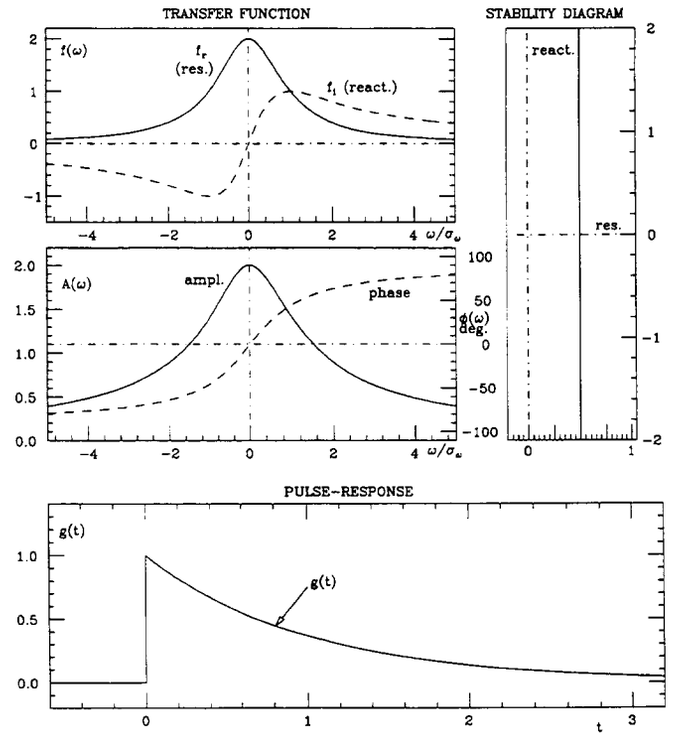


Figure 3: Response of oscillators with a Lorentz frequency distribution

1.6.2 Adaption to narrow, symmetric frequency distributions

For accelerator applications the frequency distributions are not centered around the origin and they are usually very narrow, such that the width of this distribution is much smaller than the frequency itself. This is certainly the case for synchrotron and betatron side bands. We will adapt our results for the beam response to this condition. We choose a certain central frequency $\omega_{\beta c}$ of the distribution and introduce $\Delta\omega_\beta = \omega_\beta - \omega_{\beta c}$ as a variable and express the distribution

$$f(\omega_\beta) \;\rightarrow\; F(\Delta\omega_\beta). \tag{50}$$

For simplicity we assume the distribution $F(\Delta\omega_\beta)$ to be symmetric in $\Delta\omega_\beta$ but the treatment could easily be generalized. To find the Green function we use the inverse Fourier transform Eq. (29)

$$\begin{aligned} g(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\Delta\omega_\beta)\cos(\omega_{\beta c}+\Delta\omega_\beta)d\Delta\omega_\beta \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\Delta\omega_\beta)\left[\cos(\omega_{\beta c}t)\cos(\Delta\omega_\beta t)-\sin(\omega_{\beta c}t)\sin(\Delta\omega_\beta t)\right]d\Delta\omega_\beta \\ &= \frac{1}{2\pi}\left[\cos(\omega_{\beta c}t)\int_{-\infty}^{\infty} F(\Delta\omega_\beta)\cos(\Delta\omega_\beta t)d\Delta\omega_\beta\right. \\ &\quad \left.-\sin(\omega_{\beta c}t)\int_{-\infty}^{\infty} F(\Delta\omega_\beta)\sin(\Delta\omega_\beta t)d\Delta\omega_\beta\right]. \end{aligned} \tag{51}$$

The integrations have to include the whole narrow distribution for the positive and negative frequencies. It can be seen that the second integral in the square bracket vanishes due to the symmetry of the distribution. This leads to the result

$$g(t) = \frac{1}{2\pi}\cos(\omega_{\beta c})\int F(\Delta\omega_\beta)\cos(\Delta\omega_\beta t)d\Delta\omega_\beta = \cos(\omega_{\beta c})E(t). \tag{52}$$

The pulse response of a narrow distribution consists of a fast oscillation with frequency $\omega_{\beta c}$ equal to the central frequency of the distribution modulated in amplitude by the envelope function $E(t)$. The latter is the cosine Fourier transform of the narrow distribution $F(\Delta\omega_\beta)$

$$E(t) = \frac{1}{2\pi}\int F(\Delta\omega_\beta)\cos(\Delta\omega_\beta t)d\Delta\omega_\beta. \tag{53}$$

It should be noted that this treatment is not exact since a central frequency which is not known has to be chosen. In most accelerator applications one is only interested in this envelope of the beam response to a δ-pulse excitation. We show the transfer function and this envelope of the Green function for a narrow Gaussian distribution in Fig. 4 and for a narrow Lorentz distribution in Fig. 5.

1.7 The beam response and Landau damping

We considered here a set of oscillators having a certain distribution $f(\omega_\beta)$ in frequency ω_β and calculate the response of their center-of-mass velocity $\langle\dot{y}\rangle$ to a harmonic excitation at frequency ω after a long time

$$\langle\dot{y}\rangle = \frac{\hat{G}e^{-i\omega t}}{2\pi}\left[\pi f(\omega) - i\mathrm{PV}\int\frac{f(\omega)}{\omega_\beta-\omega}d\omega_\beta\right]. \tag{54}$$

This response is also called the transfer function. It has a real part for which the velocity is in phase with the applied acceleration and absorbs energy. It is proportional to the

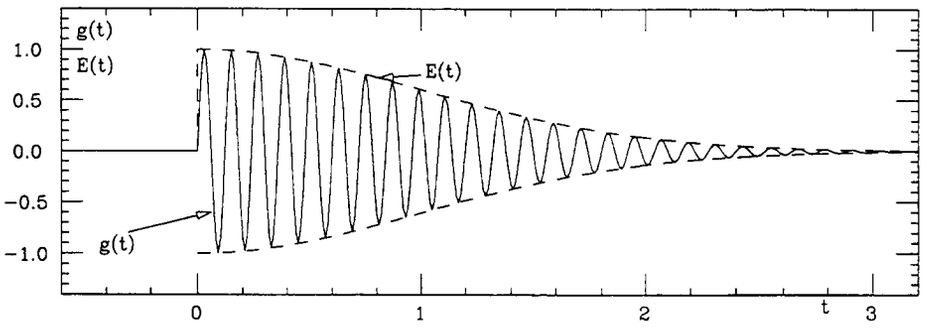


Figure 4: Response of a narrow Gaussian distribution to a pulse excitation

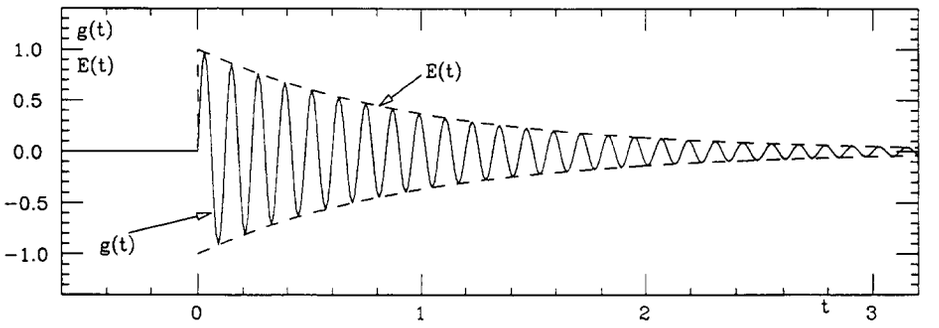


Figure 5: Response of a narrow Lorentz distribution to a pulse excitation

frequency distribution $f(\omega)$ at the excitation frequency ω. By looking at the time development of this beam response we found the oscillators with frequencies ω_β close to the excitation frequency ω reach rather large amplitudes which explains the absorbed energy.

We also investigated the response of the oscillators to a short pulse which gives each particle an initial velocity. The center-of-mass motion will be damped in a (usually) non-exponential fashion due to the spread of the individual oscillating frequencies of the particles. This process is often called phase mixing or filamentation.

These beam responses by themselves are usually not referred to as Landau damping, however they contain the essential physics of this damping mechanism. We will see in the next section that a beam of particles represents a set of oscillators like the one considered in the previous sections for the beam response. The electromagnetic fields created by this beam induce currents in the vacuum chamber surrounding the beam. These induced currents then create fields which act back on the particles. In many cases this interaction between the beam and its surroundings can increase an initial oscillation of the beam and lead to an instability. The energy necessary to drive this oscillation is taken from the forward motion of the beam. Due to the relatively large distance between the beam and the surroundings the induced current will practically only depend on the center-of-charge motion of the beam. This point is important in understanding Landau damping. The fact that the center-of-charge motion decays due to the spread in frequency can prevent an instability which would otherwise occur. In such a case the instability is indeed prevented, i.e. it never starts. If due to statistical fluctuation of the particles in the beam a very small center-of-mass motion occurs it will be damped at infinitesimal amplitudes before any sizable energy can go into the particles. This is different from the beam response to a continuous harmonic excitation we investigated before where an external excitation continued to put energy into the beam for a long time. Such a situation occur in beams only under special conditions, like e.g. in case of phase noise in the RF voltage which keeps exciting synchrotron oscillations. With a large frequency spread in the beam this

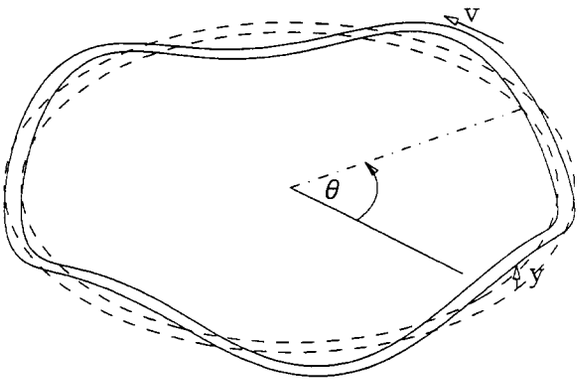


Figure 6: Oscillation modes of a coasting beam

will not result in a large coherent amplitude but in an increase of the incoherent oscillation amplitudes of the individual particles and lead to an increase of the bunch length. The large frequency spread (Landau damping) is a disadvantage in this case and it would be better to work with a small frequency spread and damp the resulting coherent oscillation with a feed back system.

1.8 Stability of a coasting beam due to Landau damping

1.8.1 Transverse oscillation modes of a coasting beam

We consider now a coasting (unbunched) beam circulating without energy loss or gain in a storage ring. We assume that this beam has a central momentum p_0 with a corresponding revolution frequency ω_{0c}. The particles themselves have a distribution in momentum with deviations Δp from the central values with distribution function

$$F_p(\Delta p) = \frac{1}{N}\frac{dN}{dp}. \tag{55}$$

The beam executes at the same time a vertical betatron oscillation such that the motion of each particle consists of a rotation $\theta = \omega_0 t$, and a vertical oscillation $y = \hat{y}\cos(Q\omega_0 t)$; Fig. 6. This beam as a whole can execute such a motion in a very large number of modes as long as we do not fix a phase relation between the betatron oscillation of the individual particles. We are, of course, interested in modes which are very simple, e.g. where all the particles go up and down together or with a simple phase relation. Such simple modes contain only relatively low frequencies and are more likely to interact with realistic impedances of the beam surroundings. We classify the modes as closed waves with a different number n of undulations around the circumference. They can be described as

$$y_n = \hat{y}\cos(n\theta - \omega t)\,, \;\; or \;\; y_n = \hat{y}e^{i(n\theta-\omega t)}. \tag{56}$$

For $n = 0$ all particles move together. If the motion is frozen by fixing the time t, the mode number n gives the number of waves around the ring; ($n = 4$ for Fig. 6).

The mode number n can take positive or negative values depending on a phase advance or a phase lag of the oscillation executed by a particle in front of a reference particle. These two classes of modes are often called fast or forward, and slow or backward, waves. We now wish to determine the frequency ω which is seen by a stationary observer at a fixed location θ. We know that the vertical oscillation frequency is $Q\omega_0$ for an individual particle. Such a particle executes the azimuthal motion $\theta = \theta_0 + \omega_0 t$. This

gives

$$e^{i(n\theta_0+n\omega_0 t-\omega t)} = e^{n\theta_0 - iQ\omega_0 t}$$

or

$$\omega = (n+Q)\omega_0 = \omega_\beta, \tag{57}$$

where n goes through positive and negative integers. It is convenient to divide these modes into fast waves

$$\omega_{\beta f} = (n_f + Q)\omega_0\,, \;\; n_f > -Q \tag{58}$$

and slow waves

$$\omega_{\beta s} = (n_s - Q)\omega_0\,, \;\; n_s > Q. \tag{59}$$

The spread in momentum of the particles in the beam results in a spread in the betatron frequency ω_β through two effects. The revolution frequency ω_0 depends on Δp

$$\Delta\omega_0 = -\omega_0\eta\frac{\Delta p}{p}\,, \;\; \eta = \alpha - \frac{1}{\gamma^2}. \tag{60}$$

Furthermore the betatron tune Q depends on the momentum through the chromaticity ξ

$$\Delta Q = \xi Q\frac{\Delta p}{p} = Q'\frac{\Delta p}{p}\,, \;\; \xi = \frac{\Delta Q/Q}{\Delta p/p}. \tag{61}$$

This leads to a spread in betatron frequency observed at a stationary location

$$\Delta\omega_{\beta f} = (Q' - \eta(n_f + Q))\omega_0\frac{\Delta p}{p}\,, \;\; \Delta\omega_{\beta s} = (Q' - \eta(n_s - Q))\omega_0\frac{\Delta p}{p}. \tag{62}$$

1.9 Response of the beam to a transverse excitation

We consider now an experiment where a periodic transverse force is applied to the beam by an electromagnetic device of length L_G. This force results in an acceleration G averaged over one revolution

$$G = \frac{F}{m} = \hat{G}e^{-i\omega t} = \frac{e[\vec{E} + [\vec{\beta}c \times \vec{B}]]_T}{m_0\gamma}\frac{L_G}{2\pi R}. \tag{63}$$

The equation of motion for a particle with betatron wave number Q and revolution frequency ω is

$$\frac{d^2y}{dt^2} + \omega_0^2 Q^2 y = \hat{G}e^{-i\omega t}.$$

We are seeking a solution where a string of particles appear to a stationary observer as

$$y = \hat{y}e^{i(n\theta-\omega t)}. \tag{64}$$

Using this and the relation between absolute and partial differentiation we get

$$\frac{dy}{dt} = \frac{\partial y}{\partial t} + \frac{\partial y}{\partial\theta}\dot{\theta} = \frac{\partial y}{\partial t} + \frac{\partial y}{\partial\theta}\omega_0 = i(n\omega_0 - \omega)y \tag{65}$$

and the equation of motion becomes

$$[-(n\omega_0 - \omega)^2 + \omega_o^2 Q^2]\hat{y}e^{i(n\theta-\omega t)} = \hat{G}e^{-i\omega t}. \tag{66}$$

We solve this for $\hat{y}$ and assume that $\theta = 0$ which means that the oscillation is observed close to the exciter

$$\hat{y} = \frac{\hat{G}}{\omega_0^2 Q^2 - (n\omega_0 - \omega)^2} = \frac{-G_0}{(\omega - \omega_0(n+Q))(\omega - \omega_0(n-Q))} \tag{67}$$

$$\frac{\hat{y}}{\hat{G}} = \frac{-1}{(\omega_{\beta f} - \omega)(\omega_{\beta s} - \omega)}. \tag{68}$$

This is the response of a single particle. It is only large if the exciting frequency ω is either close to the frequency $\omega_{\beta s}$ of the fast, or to the frequency $\omega_{\beta s}$ of the slow, wave. We will approximate for these two cases and get for the single particle responses

$$\left(\frac{\hat{y}}{\hat{G}}\right)_f \sim \frac{-1}{2Q\omega_0}\frac{1}{(\omega_{\beta f} - \omega)}\ , \quad \left(\frac{\hat{y}}{\hat{G}}\right)_s \sim \frac{1}{2Q\omega_0}\frac{1}{(\omega_{\beta s} - \omega)}. \tag{69}$$

This approximation takes only the response to the mode n (being closest in frequency) while the contributions due to the other modes are ignored [11]. In this form the single particle response is the same as that for the case of a set of oscillators used earlier. Therefore, we can easily find the response for the center-of-charge for the fast wave case by an integration over the single particle response weighted with the distribution $f(\omega_{\beta f})$ which is normalized such that $\int f(\omega_{\beta f}) d\omega_{\beta f} = 2\pi$

$$\begin{aligned} \langle \hat{y} \rangle_f &= \int f(\omega_{\beta f}) \hat{y}(\omega_{\beta f}) d\omega_{\beta f} = -\frac{\hat{G}}{2Q\omega_0} \int \frac{f(\omega_{\beta f})}{\omega_{\beta f} - \omega} d\omega_{\beta f} \\ &= \frac{\hat{G}}{2Q\omega_0}\left[\mathrm{PV} \int \frac{f(\omega_{\beta f})}{\omega_{\beta f} - \omega} d\omega_{\beta f} \pm i\pi f(\omega)\right], \end{aligned} \tag{70}$$

and for the velocity response

$$r_f = \frac{\langle \dot{y}_f \rangle_f}{G} = -\frac{i\omega \langle y \rangle_f}{G} = -\frac{\omega}{2Q\omega_0}[\pi f(\omega) - i\mathrm{PV} \int \frac{f(\omega_{\beta f})}{\omega_{\beta f} - \omega} d\omega_{\beta f}]. \tag{71}$$

$$r_s = \frac{\langle \dot{y}_f \rangle_s}{G} = -\frac{i\omega \langle y \rangle_s}{G} = \frac{\omega}{2Q\omega_0}[\pi f(\omega) - i\mathrm{PV} \int \frac{f(\omega_{\beta s})}{\omega_{\beta s} - \omega} d\omega_{\beta s}]. \tag{72}$$

The corresponding expression for the slow wave is obtained by changing the sign and replacing $\omega_{\beta f}$ by $\omega_{\beta s}$. In both cases the limits of the integration are such that only the side band in question is covered. The obtained response of a coasting beam to a transverse harmonic excitation with a frequency ω being close to the one of a betatron side band ω_β. In our approximation the integration has to cover just such a single side band and we neglect the influence of other side bands.

1.10 Transverse impedance

It is well known that a beam can excite longitudinal modes in a cavity which than react onto the beam. It is also possible that the beam excites so-called deflecting modes which give a transverse force to the beam. A simple case is illustrated in Fig. 7 where a bunch is going through a cavity with a displacement y from the axis. This can excite a mode having a longitudinal electric field which increases with distance from the cavity axis. A quarter of an oscillation later this mode has a transverse magnetic field which can deflect particles. The transverse impedance is defined as the integrated deflecting field per unit dipole moment of the exciting current

$$Z_T(\omega) = -i\frac{\int_0^{2\pi R}[\vec{E}(\omega) + [\vec{\beta}c \times \vec{B}(\omega)]]_T ds}{Iy(\omega)}. \tag{73}$$

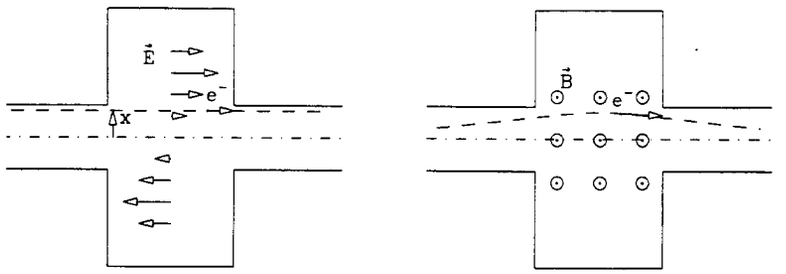


Figure 7: Transverse mode in a cavity during excitation and deflection

The right hand side is multiplied with i which indicates that the driving dipole moment Iy is out of phase with the deflecting field. The above expression is matched to a complex description of an oscillation with $e^{-i\omega t}$. If one uses the convention $e^{j\omega t}$ instead, i has to be replaced by $-j$. This impedance might become more clear if we relate the fields to the vertical velocity $\dot{y}$ rather than to the position y. With $y = \hat{y}e^{-i\omega t}$ we have $\dot{y} = -i\omega y$ and

$$Z_T(\omega) = -\omega \frac{\int_0^{2\pi R}[\vec{E}(\omega) + [\vec{\beta}c \times \vec{B}(\omega)]]_T ds}{I\dot{y}(\omega)}. \tag{74}$$

A real transverse impedance means that the transverse deflecting fields are in phase with the transverse velocity and transfer energy to or from the transverse motion of the beam.

1.11 Transverse beam stability criterion

We calculated the center-of-charge (coherent) response of the beam to a external transverse excitation with acceleration G

$$G = \frac{\langle e[\vec{E} + [\vec{\beta}c \times \vec{B}]]\rangle}{m_0\gamma} \tag{75}$$

which gives for the fast wave

$$\langle y\rangle = -\frac{\hat{G}e^{-i\omega t}}{2Q\omega_0}\int \frac{f(\omega_{\beta f})}{\omega_{\beta f} - \omega} d\omega_{\beta f}. \tag{76}$$

If the beam surroundings represent a transverse impedance Z_T

$$Z_T(\omega) = -i\frac{\int_0^{2\pi R}[\vec{E}(\omega) + [\vec{\beta}c \times \vec{B}(\omega)]]_T ds}{Iy(\omega)} \tag{77}$$

the oscillating beam induces fields in it which will act back on the beam. We assume now that the acceleration G is not external but due to the self fields induced

$$G = \frac{ieZ_T I\langle y\rangle}{\gamma m_0 2\pi R}. \tag{78}$$

Substituting this in the expression for the beam response we get for the fast and the slow wave

$$1 = -\frac{icIZ_T}{4\pi Q m_0 c^2\gamma}\int \frac{f(\omega_{\beta f})}{\omega_{\beta f} - \omega} d\omega_{\beta f} \text{ and } 1 = \frac{icIZ_T}{4\pi Q m_0 c^2\gamma}\int \frac{f(\omega_{\beta s})}{\omega_{\beta s} - \omega} d\omega_{\beta s}. \tag{79}$$

This gives a condition for which the beam is just at the limit of stability [12, 13]. By replacing the external excitation with the self induced one we assume that an oscillation, once started, is just kept going by the self forces. To make it more applicable we introduce some normalization to separate the term which depends on the different beam parameters from the one which is just given by the form of the distribution. We introduce the half width at half height S of the distribution $f(\omega_\beta)$ and normalize the two frequencies ω and ω_β with it

$$\xi_f = \frac{\omega_{\beta f}}{S} \;\; \xi_s = \frac{\omega_{\beta s}}{S}\,, \;\; \xi = \frac{\omega}{S}\,, \;\; f(\xi_f) = Sf(\omega_{\beta f})\,, \;\; f(\xi_s) = Sf(\omega_{\beta s}) \tag{80}$$

and get

$$1 = \frac{iecIZ_T}{4\pi EQS}\int \frac{f(\xi_f)}{\xi - \xi_f}d\xi_f \;\text{ and }\; 1 = -\frac{iecIZ_T}{4\pi EQS}\int \frac{f(\xi_s)}{\xi - \xi_s}d\xi_s. \tag{81}$$

This expression allows us to find, for given beam parameters, for each value of the driving frequency ω (or its normalized value ξ), the maximum impedance Z_T which still does not lead to an instability. In this expression the integral and the impedance are complex numbers. It is therefore convenient to visualize the stability criterion by mapping the impedance to the frequency ω or ξ_1. As long as this frequency has no imaginary part there is no growing instability since we expressed the oscillation as $e^{-i\omega t}$. We write the above condition slightly differently

$$\frac{ecIZ_T}{4\pi EQS} = V_T + iU_T = \frac{1}{i\int \frac{f(\xi_f)}{\xi-\xi_f}d\xi_f} \;\text{ and }\; \frac{ecIZ_T}{4\pi EQS} = V_T + iU_T = -\frac{1}{i\int \frac{f(\xi_s)}{\xi-\xi_s}d\xi_s}. \tag{82}$$

Plotting the above equality for a real frequency ξ_1 gives the stability diagram. Unfortunately the above equations rarely give solutions which can be expressed in elementary function for realistic distributions. We give in Fig. 8 the example of a Gaussian distribution which we have calculated earlier. The diagram consists of two curves; at the right, one for the slow wave with positive values of the impedance; at the left, one for the fast wave with negative values for the resistive impedance. These curves represent a situation which is at the limit of stability. A slightly larger impedance than the one corresponding to this limit will lead to an instability. For the slow wave this means that values for the complex impedance Z_T giving a reduced impedance $U_T + iV_T$ lying on the right of the stability diagram lead to instability while for corresponding values lying on the left we still have stability. For the fast wave only negative impedances lead to instabilities if the resulting reduced impedance $U_T + iV_T$ lies on the left of the fast wave stability curve. To summarize the situation for both waves, we have stability as long as the reduced impedance $U_T + iV_T$ is inside the stability diagram bounded by the fast and the slow wave stability limit curves.

1.11.1 Landau damping for bunched beams and for longitudinal instabilities

Without going into any details concerning the Landau damping for bunched beams we should make a few remarks on the limitation of the treatment presented here. The Landau damping against transverse instabilities in an unbunched beam which has a betatron frequency spread due to the energy spread combined with sextupoles Eq. (62) is a special case. The betatron oscillation we excite will not influence the frequency spread in the beam. In other words, the frequency ω_β of a particle is given by parameters like Δp, η and Q' which are not influenced by the excitation of betatron oscillations. This situation is different and more complicated in cases where the betatron frequency spread is determined by octupole fields which give a dependence of the betatron frequency on amplitude. This is always the case for a bunched beam but it can also be the dominant effect in unbunched beams. Exciting a betatron oscillation will, at the same time, influence the betatron frequency distribution. Going through this calculation one finds that

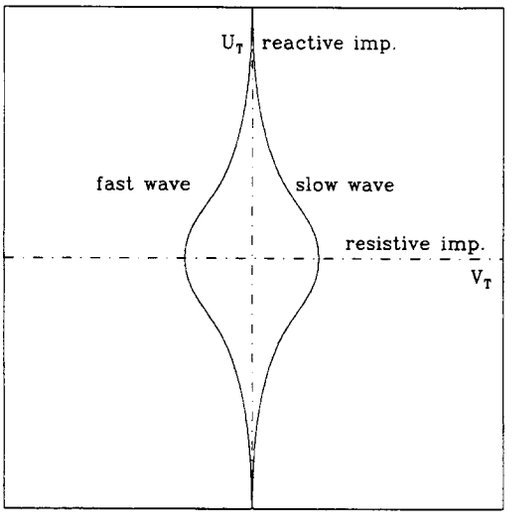


Figure 8: Stability diagram

the integrals, Eq. (8), determining the beam response and the stability diagram does not contain the distribution $f(\omega_\beta)$ but rather its derivative [12, 13]. The same situation is present for longitudinal stability for unbunched as we will see in the next section.

2 LONGITUDINAL LANDAU DAMPING OF AN UNBUNCHED BEAM

2.1 Introduction

We consider now the same coasting (unbunched) beam and investigate its longitudinal stability provided by Landau damping. Again we first have to study the response of the beam to an excitation in the longitudinal direction or in energy. In the transverse case treated before the frequency spread was determined by the energy distribution of the particles combined with the betatron frequency dependence on energy. The excitation affected the amplitude of the betatron oscillations but not their frequencies. The longitudinal response is more complicated. The spread in revolution frequencies is given by the energy distribution combined with the relation between energy and revolution frequency deviations. The beam is excited in energy which directly affects the frequency distribution. This situation is more complicated and the longitudinal response is usually calculated with the Vlasov equation. Before using this method we make a short study of the longitudinal dynamics and try to illustrate the beam response by a more qualitative description.

2.2 Longitudinal dynamics

We consider an unbunched (coasting) beam circulating in a storage ring with the nominal radius R_0, the nominal revolution frequency $\omega_0 = \beta_0 c/R_0$, the nominal energy and momentum $E_0 = m_0c^2\gamma_0$, $p_0 = m_0c\beta_0\gamma_0$. A particle with a different momentum $p = p_0 + \Delta p$ has a relative energy deviation

$$\frac{\Delta E}{E_0} = \frac{E - E_0}{E_0} = \beta_0^2 \frac{\Delta p}{p_0}. \tag{83}$$

This particle will move on an orbit which is different from the nominal one by the radial distance

$$\Delta R = D_x \frac{\Delta p}{p_0} = \frac{D_x}{\beta_0^2} \frac{\Delta E}{E_0}, \tag{84}$$

where D_x is the horizontal dispersion. The length L of the orbit has a relative difference compared to the nominal one by

$$\frac{\Delta L}{L_0} = \frac{\Delta p}{p_0} \int_0^{2\pi} \frac{D_x(\theta)}{\rho} d\theta = \alpha \frac{\Delta p}{p_0}, \tag{85}$$

where ρ is the radius of curvature. The revolution time $T = L/\beta c$ is also changed

$$\frac{\Delta T}{T_0} = \frac{\Delta L}{L_0} - \frac{\Delta \beta}{\beta_0} = \left(\alpha - \frac{1}{\gamma^2}\right) \frac{\Delta p}{p_0}, \tag{86}$$

where we used the relation

$$\frac{\Delta p}{p_0} = \gamma^2 \frac{\Delta \beta}{\beta}. \tag{87}$$

As a consequence the revolution frequency ω_r is changed by

$$\frac{\Delta \omega_r}{\omega_0} = -\frac{\Delta T}{T_0} = -\left(\alpha - \frac{1}{\gamma^2}\right) \frac{\Delta p}{p_0} = -\eta \frac{\Delta p}{p_0} = -\frac{\eta}{\beta^2} \frac{\Delta E}{E_0}, \tag{88}$$

where we introduced $\eta = \alpha - 1/\gamma^2$.

A realistic beam has many particles with a certain distribution

$$f(\Delta E, \theta) = \frac{1}{N} \frac{d^2 N}{d\theta dE} \quad \text{with} \quad \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} f(\theta, \Delta E) dE = 1 \tag{89}$$

in the energy deviation ΔE. In general this distribution can depend on the azimuthal angle θ. However, the equilibrium distribution is independent of θ, $g_0 = f_0(\Delta E)$. Instead of giving the distribution with respect to ΔE we can also give it with respect to the revolution frequency ω_r

$$F(\theta, \Delta \omega_r) = \frac{1}{N} \frac{d^2 N}{d\theta d\omega_r} = \frac{1}{N} \frac{dE}{d\omega_r} \frac{d^2 N}{d\theta dE} = \frac{\beta^2 E_0}{\eta \omega_0} f(\theta, \Delta E). \tag{90}$$

For convenience we assumed a symmetric equilibrium distribution is symmetric

$$f_0(-\Delta E) = f_0(\Delta E). \tag{91}$$

2.3 Illustration of the beam response to a pulse excitation

We take now a coasting beam having a certain energy distribution $f_0(\Delta E)$ with rms value σ_E. We introduce at the time $t = 0$ a disturbance in energy of the form

$$\delta E(\theta) = \delta E_0 \sin(n\theta) \tag{92}$$

for all particles in the beam. This is illustrated in Fig. 9 for the case $n = 2$. It shows the beam over one turn, $\theta = 0 \rightarrow 2\pi$ with the lines indicating points of equal density $f(\Delta E, \theta)$. Two particles are marked by a point to give their position and by an arrow indicating their motion with respect to a nominal particle at $\theta = \pi$ and $\Delta E = 0$. A cross section through the beam at $\theta = \pi$ is shown on the right. At the bottom we show the current $I(\theta)$ represented by this beam which is given by the integral

$$I(\theta) = \frac{N e \omega_0}{2\pi} \int f(\theta, \Delta E) d(\Delta E). \tag{93}$$

The first frame gives the situation just after the excitation. Since the disturbance affects the energy only, the current I is not changed by it and is still uniform $I(\theta) = I_0$ and equal to its average value. The dependence of the revolution frequency on energy deviation of the particles

$$\frac{\Delta\omega_r}{\omega_0} = -\frac{\eta}{\beta^2}\frac{\Delta E}{E} \tag{94}$$

results in a relative longitudinal displacement of regions with different energy as time goes on. This is illustrated in the second frame of the figure which shows the situation after N turns for which

$$2\pi N\frac{\eta}{\beta^2}\frac{\sigma E}{E_0} = 0.3 \tag{95}$$

which means that after N turns a particle with an energy deviation $\Delta E = \sigma_E$ has slipped by an angle 0.3 rad compared to a particle with nominal energy. Due to this sheering the original energy disturbance is partially converted into a modulation of the beam current. This gets even more pronounced a little later as shown in the third frame. However, as time goes on the sheering becomes more pronounced and the modulations at the different energy deviations ΔE overlap. This reduces the current modulation again which will slowly disappear as illustrated in the last two frames.

2.4 The Vlasov equation

The application of kinetic theory and the related Vlasov equation has been treated in earlier CAS courses [14, 15, 16]. Here, we only make a short and not rigorous derivation and put all the emphasis on the application to Landau damping. We start with Liouville's theorem which says that the phase space density $\psi(q,p)$ in a small volume around a given particle does not change with time

$$\frac{d\psi(q,p)}{dt} = 0. \tag{96}$$

Usually we associate q with a space coordinate and p with a momentum or a velocity but the situation can be more general. These coordinates q and p should be canonically conjugate which means that they are derived from a Hamiltonian $H(q,p)$ by the canonical equations

$$\dot{q} = \frac{\partial H(q,p)}{\partial p}\ ,\ \dot{p} = -\frac{\partial H(q,p)}{\partial q}. \tag{97}$$

The theorem of Liouville is visualized in Fig. 10 where the phase-space trajectories of three particles are drawn. Their positions and the surface element they determine are shown for two different times t. The form of this surface element changes but its area stays constant. According to Liouville's theorem the particles move in phase space like an incompressible liquid. This is also illustrated in the development of the energy disturbance shown in Fig. 9. This conservation of the phase-space density seems to contradict common experience with accelerators where one often talks about dilution of phase-space due to non-linearities, called filamentation. This paradox is illustrated in Fig. 11. On the left hand side we show a phase space occupied by particles and which has the simple form of a rectangle. The instruments observing the distribution of the particles have, in practice, a limited resolution indicated by the grid in the figure. For the left diagram the instruments would localize the beam within four such resolution elements. After the beam has gone through some non-linear elements (e.g. a mismatched bunch rotating in the non-linear RF wave form) the form of the phase-space occupied by the beam is distorted. Although its actual area has not changed the limited resolution of the instruments indicates now a distribution of particles over many more resolution grid elements. The situation is similar to a liquid which is transformed into foam. On a microscopic level the density is still the same but the global density is much less. For practical purposes the global phase-space

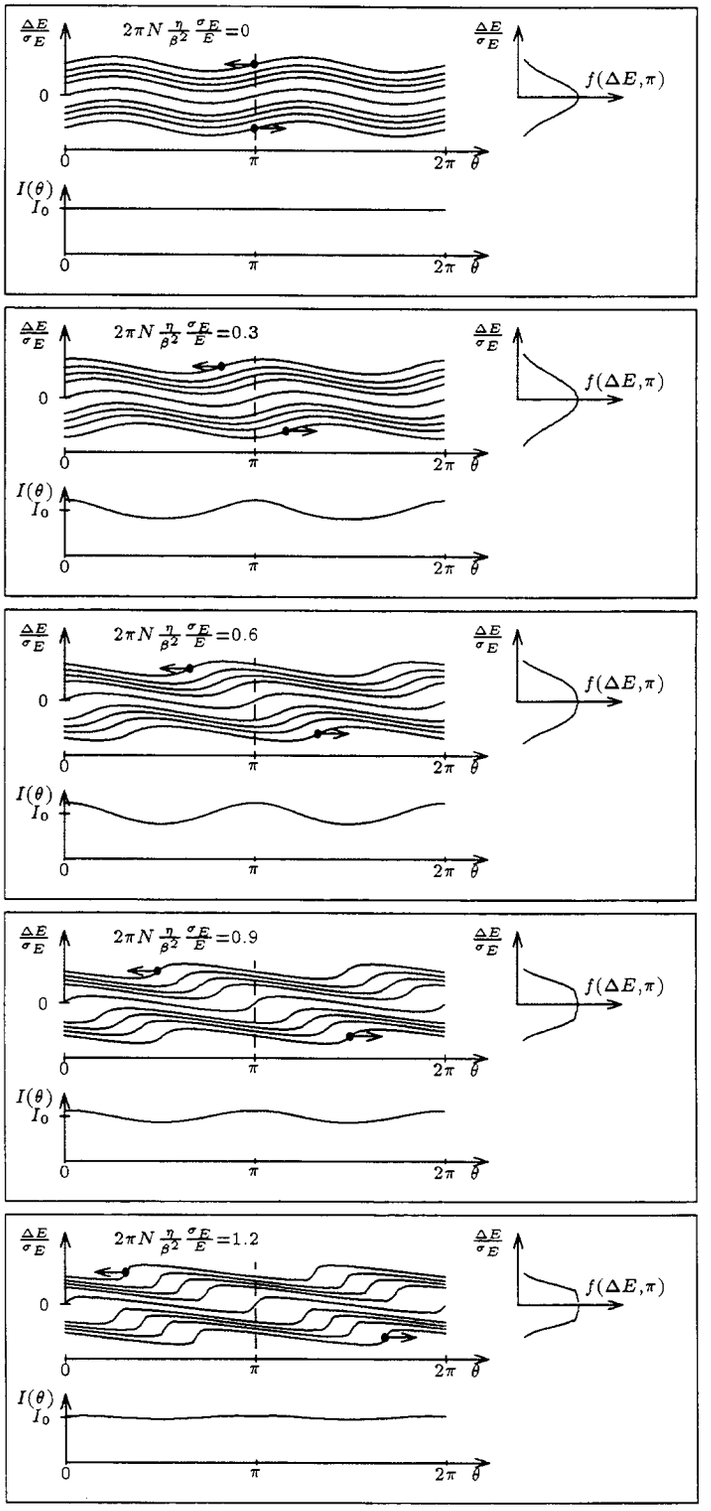


Figure 9: Development of a harmonic energy disturbance of an unbunched beam

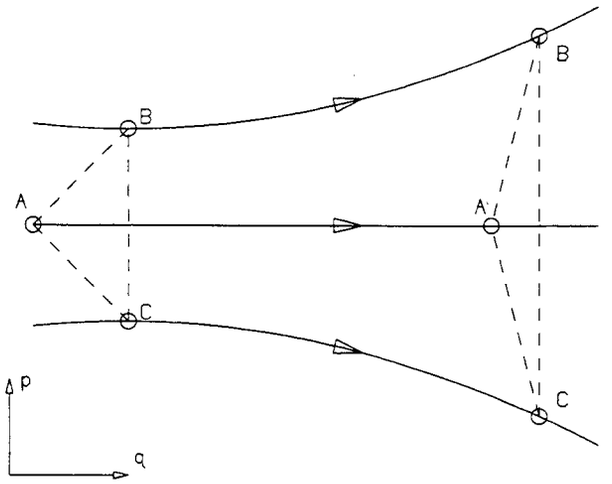


Figure 10: Phase space trajectories and Liouville's theorem

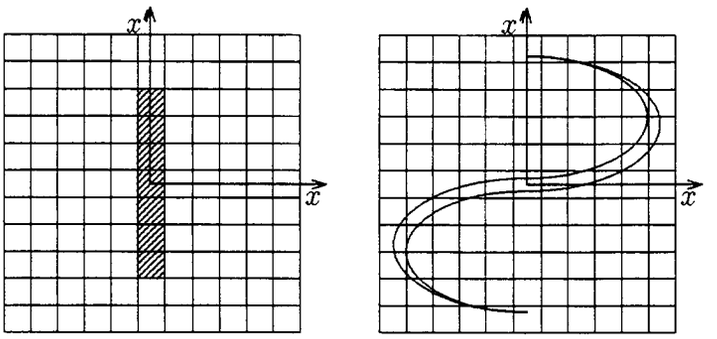


Figure 11: Resolution and phase-space area conservation [17]

density is of course also relevant. The constancy of the phase density $\psi(q,p)$ is expressed by the equation

$$\frac{d\psi(q,p)}{dt} = 0 \tag{98}$$

where the absolute differentiation indicates that one follows the particle while measuring the density of its immediate neighborhood. We would like to know the development of this density as seen by a stationary observer (like a beam monitor) which does not follow the particle. It depends now not only directly on the time t but also indirectly through the coordinates q and p of the moving particles which change with time. Therefore, we have to express the absolute differential with respect to t by the partial differentiations with respect to time as well as with respect to the coordinates q and p multiplied with the time derivative of these coordinates

$$\frac{d\psi(q,p)}{dt} = \frac{\partial\psi(q,p)}{\partial t} + \dot{q}\frac{\partial\psi(q,p)}{\partial q} + \dot{p}\frac{\partial\psi(q,p)}{\partial p} = 0. \tag{99}$$

This expression is the Vlasov equation in its most simple form and is nothing else but an expression for Liouville's conservation of phase-space density seen by a stationary observer. Rigorous derivation of the Vlasov equation can be found in the literature e.g. [14, 18]. It is interesting to apply the Vlasov equation to a stationary distributions which does not depend explicitly on time. For example a stationary bunch looks the same each revolution as it is observed through an intensity monitor. For this stationary case the

Vlasov equation becomes

$$\frac{\partial\psi(q,p)}{\partial t} = 0 \;\rightarrow\; \dot{q}\frac{\partial\psi(q,p)}{\partial q} + \dot{p}\frac{\partial\psi(q,p)}{\partial p} = 0. \tag{100}$$

Expressing $\dot{q}$ and $\dot{p}$ with the canonical equations we get

$$\frac{\partial\psi}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial\psi}{\partial p}\frac{\partial H}{\partial q} = 0. \tag{101}$$

From this it can be shown that a stationary distribution ψ is a function of the Hamiltonian H only $\psi = \psi(H)$. This fact is often useful in applications, specially in finding stationary distributions [19, 20]. In this presentation we apply it to determine the response of a beam to an external excitation. This will be one of the ingredients to treat the Landau damping of longitudinal instabilities.

Before we go to the practical applications we have to make some remarks on the coordinates q and p we use. In Liouville's theorem the phase-space area is only conserved if expressed in canonically conjugate variables q, p. The same criterion applies to the validity of the Vlasov equation. Examples of such pairs are position and momentum x, p or time and energy t, E. These variables are often not very practical for accelerator applications. We could for example choose the longitudinal position s and the corresponding momentum p_s as coordinates to describe the longitudinal phase-space. However the path length s around the ring is different for particles with different momenta. As a result, individual particles would enter a certain machine element, such as a quadrupole, at different values of the coordinate s. From this point of view other coordinate pairs in which the position of the machine element is given for all particles by one coordinate only are more suitable. Such a variable is for example the angle θ of the motion around the ring. However the conjugate variable is quite complicated [21]

$$W = \int_{E_0}^{(E_0+\Delta E} \frac{\Delta E}{\omega} d(\Delta E) \tag{102}$$

where ΔE is the deviation from the nominal energy E_0 and $\omega(\Delta E)$ is the revolution frequency which depends on this energy deviation. In our applications we only use terms up to first order in ΔE and use as an approximation θ and ΔE as variables. However we should always remember that this is a linear approximation in $\Delta E/E_0$ which can become very inaccurate close to transition energy. As a further convenience we will use the RF phase angle ϕ instead of θ to which it is related through $\phi = h\theta$ with h being the harmonic number.

2.5 Response of the beam to a longitudinal excitation

We follow here the methods used in [13, 22] and consider a storage ring containing a cavity which is excited with frequency $\omega \approx n\omega_0$ and amplitude U_0

$$U(t) = U_0 \mathrm{e}^{-i\omega t}. \tag{103}$$

At each revolution the energy of each particle is changed by

$$\Delta E = eU(t_i) = eU_0 \mathrm{e}^{-i\omega t_i} \tag{104}$$

due to the cavity voltage where t_i is the time the particle passes through the cavity. Assuming a relatively small energy change per turn we can make a smooth approximation as if the cavity voltage was distributed uniformly around the ring

$$\Delta\dot{E} = \dot{E} = \frac{\omega_r}{2\pi}\Delta E = \frac{e\omega_r U_0}{2\pi}\mathrm{e}^{i(n\theta-\omega t)}. \tag{105}$$

We use now the Vlasov equation to describe the development of the particle distribution. As mentioned before we should express it in canonically conjugate variables θ and W. However for our approximation of small energy deviations we use θ and ΔE as variables and get

$$\frac{\partial f(\theta,\Delta E)}{\partial t}+\frac{\partial f(\theta,\Delta E)}{\partial \theta}\dot{\theta}+\frac{\partial f(\theta,\Delta E)}{\partial E}\dot{E}=0 \tag{106}$$

The energy change $\dot{E}$ Eq. (105) has to be used in the Vlasov equation. We expect that it will establish an energy modulation of the particles of the same form resulting in a modification of the distribution function

$$f(\theta,\Delta E)=f_0(\Delta E)+f_1(\Delta E)\mathrm{e}^{i(n\theta-\omega t)} \tag{107}$$

which we assume to be small $f_1 \ll f_0$ such that quadratic terms in f_1 as well as terms containing products $f_1 U_0$ can be neglected. With this we get from the Vlasov equation

$$i(\omega-n\omega_r)f_1(\Delta E)\mathrm{e}^{i(n\theta-\omega t)}=-\frac{df_0(\Delta E)}{d(E)}\frac{e\omega_r U_0}{2\pi}\mathrm{e}^{i(n\theta-\omega t)}. \tag{108}$$

We can replace the revolution frequency ω_r of the individual particles on the right hand side by the nominal frequency ω_0 (but not on the left hand side) leading to

$$f_1(\Delta E)\mathrm{e}^{i(n\theta-\omega t)}=\frac{ie\omega_0\mathrm{e}^{i(n\theta-\omega t)}}{2\pi}\frac{\frac{df_0}{d\Delta E}}{\omega-n\omega_r} \tag{109}$$

We cannot observe the distribution $f(\theta,\Delta E)$ easily but we can measure the beam current $I(\theta,t)$ which is given by

$$I(\theta,t)=Ne\omega_0\int_{-\infty}^{\infty}f(\theta,\Delta E)d(\Delta E). \tag{110}$$

The above integration does not really go from $-\infty$ to ∞ but has just to cover the distribution $f(\theta,\Delta E)$. We split the current into the unperturbed part I_0 and a perturbation $I_1(\theta,t)$. For convenience we chose $\theta=0$ and get for the latter

$$I_1(0,t)=\hat{I}_1\mathrm{e}^{-i\omega t}=\frac{iNe^2\omega_0^2U_0\mathrm{e}^{-i\omega t}}{2\pi}\int\frac{\frac{df_0}{dE}}{\omega-n\omega_r}dE. \tag{111}$$

The response will be large if $\omega\approx n\omega_0$. The revolution frequency ω_r of the individual particles depends on its energy deviation ΔE

$$\frac{\Delta\omega_r}{\omega_0}=-\frac{\eta}{\beta^2}\frac{\Delta E}{E_0}. \tag{112}$$

The above integration might contain a pole. To investigate this we use ω_r as an independent variable

$$\hat{I}_1=\frac{iNe^2\omega_0^3\eta U_0}{2\pi\beta^2E_0}\int\frac{\frac{dF_0(\omega_r)}{d\omega_r}}{\omega-n\omega_r}d\omega_r. \tag{113}$$

Such integrals containing a pole have a principle value and a residue as we explained before

$$\int\frac{\frac{dF_0(\omega_r)}{d\omega_r}}{\omega-n\omega_r}d\omega_r=PV\int\frac{\frac{dF_0(\omega_r)}{d\omega_r}}{\omega-n\omega_r}d\omega_r\pm i\pi\frac{dF_0}{d\omega_r}(\omega). \tag{114}$$

The principle value integral is obtained by leaving a small gap around the pole during integration and letting this gap go to zero

$$PV\int\frac{\frac{dF_0(\omega_\tau)}{d\omega_\tau}}{\omega-n\omega_\tau}d\omega_\tau=\int_{-\infty}^{\omega/n-\epsilon}\frac{\frac{dF_0(\omega_\tau)}{d\omega_\tau}}{\omega-n\omega_\tau}d\omega+\int_{\omega/n+\epsilon}^{\infty}\frac{\frac{dF_0(\omega_\tau)}{d\omega_\tau}}{\omega-n\omega_\tau}d\omega\ \text{ with }\ \epsilon\to 0. \tag{115}$$

The residue is just the derivative $dF_0/d\omega_\tau$ taken at the frequency ω of excitation. We get for the perturbed current

$$\hat{I}_1(\omega)=\frac{Ne^2\omega_0^3\eta U_0}{2\pi\beta^2E_0}\left(\mp\pi\frac{dF_0}{d\omega_\tau}(\omega)+iPV\int\frac{\frac{dF_0(\omega_\tau)}{d\omega_\tau}}{\omega-n\omega_\tau}d\omega_\tau\right) \tag{116}$$

The current $\hat{I}_1(\omega)$ induced by the excitation has a real and an imaginary part. The real component of the current is in phase with the excitation voltage and therefore absorbs power. This can lead to damping, called Landau damping. The sign of this term seems to be ambiguous. This is due to the fact that we did not specify the initial condition. In a natural experiment there will be no perturbed current when one starts the excitation. It will appear only later as a consequence of the external voltage. In this case the positive sign should be taken. It is however also possible that the beam has already a modulation at the start and the external voltage can be such as to reduce this perturbation which would be expressed by the negative sign. We assume now the first case in which energy is absorbed by the beam and take the positive sign.

The imaginary part of the induced current is 90^0 out of phase compared to the excitation and does not lead to energy absorption and damping.

2.6 Response of the beam to a pulse excitation

Instead of investigating the beam to a harmonic excitation (transfer function) we can also establish a sudden energy modulation and observe the free motion of the particles thereafter. This is the response to a pulse excitation also called Green function.

We like to start with a simple form of an energy modulation of the beam and consider a harmonic dependence of the energy deviation around the ring which is closed in itself. At the time $t=0$ we give each particle an energy change of the form

$$\delta E(\theta)=\delta E_0\cos(n\theta). \tag{117}$$

This modulation is closed in itself around the ring after n periods. In a real machine one does not have cavities distributed around the ring and such a modulation can be obtained approximately by applying a harmonic voltage to a single cavity during one turn. With the particle distribution $f_0(\Delta E)$ before the pulse excitation we will have the modified distribution immediately after the excitation at the time $t=0^+$ assuming a small modulation

$$f(0^+,\theta,\Delta E)=f_0\left(\Delta E+\delta E_0\cos(n\theta)\right)=f_0(\Delta E)-\frac{df_0}{dE}\delta E_0\cos(n,\theta). \tag{118}$$

We assume that the induced current is observed at the location θ_0 and consider development of the disturbance in time. Each particle moves around the ring with the angular velocity ω_τ. The particle which was at the time $t=0$ at the position $-\theta$ will be at the point of observation θ_0 at the time $t=(\theta-\theta_0)/\omega_0$ The distribution function at the time t is therefore

$$f(\theta_0,\Delta E,t)=f_0(\Delta E)-\frac{df_0}{dE}\delta E_0\cos(n\theta_0-n\omega_\tau t). \tag{119}$$

To simplify the expression we assume that the beam is observed at the origin $\theta_0=0$

$$f(\Delta E,t)=f_0(\Delta E)-\frac{df_0}{dE}\delta E_0\cos(n\omega_\tau t). \tag{120}$$

The resulting current $I(t)$ is

$$\begin{aligned} I(t) &= Ne\int \omega_r f(\Delta E,t)dE \approx e\omega_0 \int f(\Delta E,t)dE \\ &= I_0 - Ne\omega_0 \int \frac{df_0}{dE}\delta E_0 \cos(n\omega_r t)dE = I_0 + I_1(t). \end{aligned} \tag{121}$$

We assume now that the relative energy spread is small and remember that the revolution frequency ω_r of the individual particles is a function of their deviation ΔE

$$\omega_r = \omega_0 + \Delta\omega_r \ , \ \Delta\omega_r = -\omega_0\eta\frac{\Delta E}{E}. \tag{122}$$

With this we get for the perturbed current.

$$\begin{aligned} I_1(t) &= \frac{Ne\omega_0\eta\delta E}{\beta^2 E_0}\int \frac{dF_0}{d\omega_r}\cos(n\omega_0 t + n\Delta\omega_r t)d\omega_r \\ &= \frac{Ne\omega_0\eta\delta E}{\beta^2 E_0}\int \frac{dF_0}{d\omega_r}\left[\cos(n\omega_0 t)\cos(n\Delta\omega_r t) - \sin(n\omega_0 t)\sin(n\Delta\omega_r t)\right]d\omega_r. \end{aligned} \tag{123}$$

Making use of the assumed symmetry

$$F_0(\Delta\omega_r) = F_0(-\omega_r) \ \rightarrow \ \frac{dF_0}{d\omega_r}(\Delta\omega_r) = -\frac{dF_0}{d\omega_r}(\Delta\omega_r) \tag{124}$$

we find that the first term of the above integration vanish and the perturbed current becomes

$$I_1(t) = -\frac{Ne\omega_0\eta\delta E}{\beta^2 E_0}\sin(n\omega_0 t)\int \frac{dF_0}{d\omega_r}\sin(n\Delta\omega_r t)d\omega_r = -\frac{Ne\omega_0\eta\delta E}{\beta^2 E_0}\sin(n\omega_0 t)F(t). \tag{125}$$

The factor in front of the integral of the above equation represents an oscillation with frequency $n\omega_0$. The integral $F(t)$ is slowly varying and represents the envelope of the oscillation. It is often called Green function.

Comparing the pulse response of the beam (Green function) with the response to a harmonic excitation (transfer function) (116) we find that it is proportional to the inverse sine Fourier transform of the latter. We would at first expect that they are related by the cosine Fourier transform. However the excitation with a pulse leads first to a modulation in energy which transforms into a current modulation only later. It can be shown that the complex transfer function is the Fourier transform of the Green function.

2.7 Example of a Gaussian distribution

We take the case of a Gaussian particle distribution in energy and revolution frequency deviation

$$F_0(\Delta\omega_r) = \frac{1}{\sqrt{2\pi}\sigma_\omega}\mathrm{e}^{-\frac{1}{2}\left(\frac{\Delta\omega_r}{\sigma_\omega}\right)^2}, \tag{126}$$

where σ_ω is the rms width of the distribution in ω_r. The response to a harmonic excitation is according to Eq. (116)

$$\begin{aligned} \hat{I}_1(\omega) &= \frac{Ne^2\omega_0^3\eta U_0}{2\pi\beta^2 E_0}\left(\pi\frac{dF_0}{d\omega_r}(\Delta\omega) + i\mathrm{PV}\int \frac{\frac{dF_0(\omega_r)}{d\omega_r}}{\omega - n\omega_r}d\omega_r\right) \\ &= \frac{e^2\omega_0^3\eta U_0}{2\pi\beta^2 E_0 n}\frac{N}{\sqrt{2\pi}\sigma_\omega^2}\left(-\pi\frac{\Delta\omega}{\sigma_\omega}\mathrm{e}^{-\frac{1}{2}\left(\frac{\Delta\omega_r}{\sigma_\omega}\right)^2} - i\mathrm{PV}\int \frac{\Delta\omega_r}{\sigma_\omega}\frac{\mathrm{e}^{-\frac{1}{2}\left(\frac{\Delta\omega_r}{\sigma_\omega}\right)^2}d\omega_r}{\omega - n\omega_r}\right). \end{aligned} \tag{127}$$

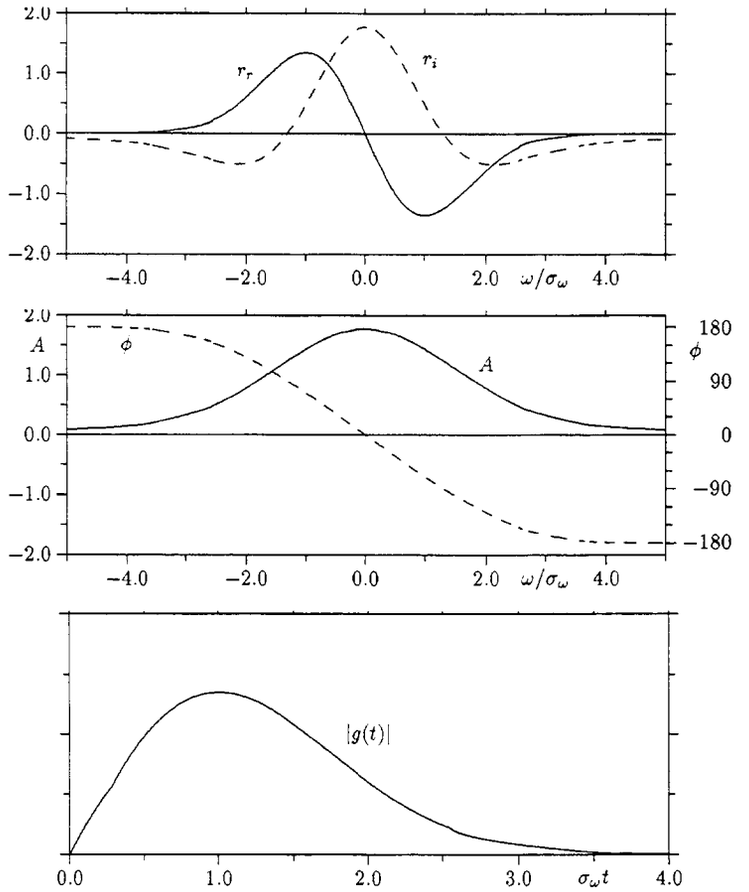


Figure 12: Longitudinal response of a coasting beam with Gaussian distribution

The principle value integral cannot be expressed in elementary functions. For the response to a pulse excitation we get

$$\begin{aligned} F(t) &= \int \frac{dF_0}{d\omega_r}\sin(n\Delta\omega_r t)d\omega_r = -\frac{N}{2\pi\sqrt{2\pi}\sigma_\omega}\int \frac{\Delta\omega_r}{\sigma_\omega}\mathrm{e}^{-\frac{1}{2}\left(\frac{\Delta\omega_r}{\sigma_\omega}\right)^2}\sin(\Delta\omega_r t)\frac{d\omega_r}{\sigma_\omega} \\ &= -\frac{N}{2\pi\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\sigma_\omega^2 t^2}t. \end{aligned} \tag{128}$$

These responses of a Gaussian beam to a harmonic and a pulse excitation are plotted in Fig. 12.

2.8 Longitudinal stability

The response to a longitudinal excitation $U_0\mathrm{e}^{-i\omega t}$ is a current perturbation Eq. (116)

$$\hat{I}_1 = \frac{iNe^2\omega_0^3\eta U_0}{2\pi\beta^2 E_0}\int \frac{\frac{dF_0(\omega_r)}{d\omega_r}}{\omega - n\omega_r}d\omega_r = \frac{Ne^2\omega_0^3\eta U_0}{2\pi\beta^2 E_0}\left(\mp\pi\frac{dF_0}{d\omega_r}(\omega) + iPV\int \frac{\frac{dF_0(\omega_r)}{d\omega_r}}{\omega - n\omega_r}d\omega_r\right). \tag{129}$$

In a storage ring the beam is surrounded by different components like the vacuum chamber, bellows, kicker magnets, cavities etc. which represent a certain longitudinal impedance $Z(\omega)$. The perturbed current I_1 induces longitudinal fields in this impedance which can act back on the beam. Integrating the field component at one frequency ω over the circumference gives the voltage

$$U_{ind.}(\omega) = -I_1(\omega)Z(\omega). \tag{130}$$

We consider now a situation in which the current perturbation I_1 established by the external voltage U_0 induces in the impedance a voltage which is equal to the applied voltage U_0. In this case the external voltage could be turned off and the current perturbation would continue to exist with constant amplitude. If the impedance is increased the induced voltage becomes larger than the one necessary to establish the original perturbation and the amplitude would grow. The case $U_{ind} = U_0$ represents therefore a stability limit. A slightly larger impedance would lead to growth of a small perturbation at the frequency ω. We get the stability limit by replacing U_0 by U_{ind} in Eq. (129)

$$I_1(\omega) = -\frac{iNe^2\omega_0^3 I_1(\omega)Z(\omega)}{2\pi\beta^2 E_0 n}\int \frac{\frac{dF_0}{d\omega_r}}{\frac{\omega}{n}-\omega_r}d\omega_r = \tag{131}$$

$$\frac{Ne^2\omega_0^3\eta Z(\omega)}{2\pi\beta^2 E_0}\left[\pi\frac{dF_0}{d\omega_r}(\omega) + iPV\int \frac{\frac{dF_0}{d\omega_r}}{\frac{\omega}{n}-\omega_r}d\omega_r\right] \tag{132}$$

or

$$1 = -\frac{iNe^2\omega_0^3\eta Z(\omega)}{2\pi\beta^2 E_0 n}\int \frac{\frac{dF_0}{d\omega_r}}{\frac{\omega}{n}-\omega_r}d\omega_r. \tag{133}$$

The integral and the impedance $Z(\omega)$ are complex. The above equation gives the maximum complex impedance for which there is still stability.

To bring the above equation into a form more adapted for practical applications we introduce some normalizations [23]. We use the *half-width* frequency spread at half-height and relate it to the corresponding spread in momentum

$$2S = -\eta\omega_0\frac{\Delta p}{p}, \tag{134}$$

and introduce the reduced frequencies

$$x = \frac{\omega_r - \omega_0}{S}\ ,\ x_1 = \frac{\omega - n\omega_0}{nS}, \tag{135}$$

the reduced distribution

$$f_0(x) = 2\pi S F_0(\omega_r)\ \text{ with }\ \int f_0(x)dx = 1 \tag{136}$$

and use the average current

$$I_0 = \frac{eN\omega_0}{2\pi}. \tag{137}$$

This gives for the stability condition

$$1 = -\frac{eI_0 Z(\omega)/n}{2\pi\beta^2 E_0\eta(\Delta p/p)^2}\left[\frac{\pi df_0}{dx}(x_1) - iPV\int\frac{\frac{df_0}{dx}}{x-x_1}dx\right] \tag{138}$$

which we can write in the form

$$V' + U' = \frac{eI_0}{2\pi\beta^2 E_0\eta(\Delta p/p)^2}\left(\frac{Z_r}{n} + i\frac{Z_i}{n}\right) = -\left(\pi\frac{df_0}{dx}(x_1) - iPV\int\frac{\frac{df}{dx}}{x-x_1}dx\right)^{-1}. \tag{139}$$

The Factor on the left hand side contains now all the global beam parameters and the impedance while the right hand side depends on the form of the distribution only.

We can now plot the above relation in a complex plane and obtain the stability diagram. This is done by calculating the U' and V' as a function of x_1, i.e. of the frequency

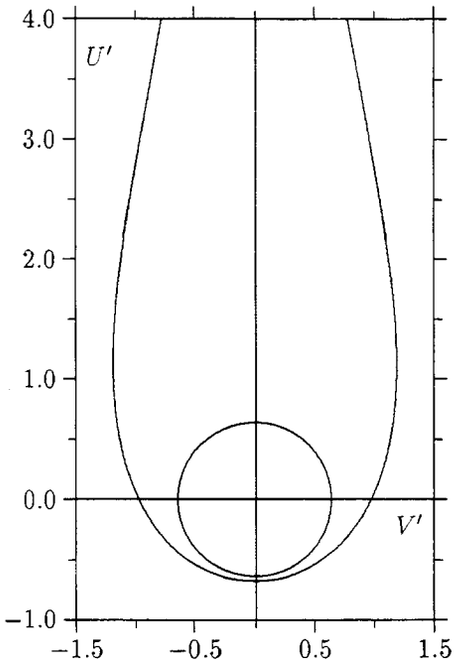


Figure 13: Statility diagram of a beam with Gaussian distribution

ω. The response of the beam to an excitation is only large when ω is close to a harmonic n of the revolution frequency ω_0. The quantity $U' + iV'$ representing the stability diagram is proportional to the inverse response. If the frequency ω approaches the harmonic of the revolution frequency a point on the stability diagram comes down from infinity, moves around the origin (while ω goes through $n\omega_0$) and moves then again to large values. The area surrounded by the stability diagram contains values of U' and V' for which the beam is stable. For a given beam current, energy, frequency spread and other beam parameters can therefore determine the maximum complex stability which still provides stability. In Fig. 13 the stability diagrams of a beam with a Gaussian distribution is plotted. By taking such diagrams for different distributions one can approximate them by a circle with a radius of about 0.6. This gives a very simple stability condition

$$\left|\frac{Z}{n}\right| \leq 0.6 \frac{2\pi\beta^2 E_0 \eta (\Delta p/p)^2}{eI_0} \tag{140}$$

which is called the Keil-Schnell criterion [24].

2.9 Discussion

We started by calculating the response of an unbunched beam to a longitudinal harmonic excitation with frequency ω. We found that this response has a part for which the velocity is in phase with the external acceleration. It can therefore absorb energy and lead to damping, called Landau damping. This result is not so easy to understand physically. A more detailed analysis indicates (shown for the transverse case) that the excitation leads to growing amplitudes of the few particles for which the natural revolution frequency ω_r (or a harmonic of it) is close the excitation frequency ω. The absorbed energy goes into the growing amplitudes of these particles.

We later considered the response of the particles in the beam to a pulse excitation which established an energy modulation around the ring. The following free oscillation executed by the particle with different frequencies lead to a decay of the coherent en-

ergy or current modulation. However the amplitude of the individual particles will stay constant. This response to a pulse excitation (Green function) is easier to understand than the response to a harmonic excitation (transfer function) but not so easy to check experimentally.

To compare this beam response to a longitudinal excitation with the transverse case we find an important difference. In the transfer or Green function the distribution f_0 appears for the transverse response while its derivative $df_0/d\omega$ is the relevant quantity for the longitudinal case. In the latter case the excitation affects the quantity ΔE which determines the spread while in the transverse case the distribution is given by parameters which are not affected by the excitation. The Landau damping of bunched beams can be treated in a similar way as the longitudinal Landau damping of the unbunched beam [2, 25, 26].

The reduction of the coherent amplitude after a pulse excitation and the resistive response of the beam to a harmonic excitation are the basis of Landau damping, but they represent two extreme cases. In the first case the continuous excitation leads to very large amplitudes of a few particles, in the second case most particles obtain a finite amplitude by the pulse. In reality Landau damping avoids an instability at infinitesimally small amplitudes. Let us assume, due to statistical fluctuation of the particles, a harmonic current modulation is established. It will induce a current in the beam surroundings which will in turn apply a voltage to the beam. Taking the picture of a harmonic excitation the energy provided by this voltage is absorbed by some gain in amplitude of a few particles and the modulation does not grow further. From the view of the response to a pulse excitation the frequency spread will smear out established modulation faster than the induced voltage can enhance it. Since all this happens at infinitesimally small amplitudes a beam stabilized by Landau damping does not grow in energy spread but can be kept stable for days as experience shows.

References

[1] H.G. Hereward, The elementary theory of Landau damping, CERN Report 65-20 (1965).

[2] H.G. Hereward, Landau damping by non-linearities, CERN/MPS/DL 69-11 (1969).

[3] R.D. Kohaupt, What is Landau damping? Plausibilities, fundamental thoughts, theory, DESY M-86-02 (1986).

[4] A. Hofmann, Landau damping, Proc. 1988 CAS CERN Accel. School on Advanced Accel. Physics, CERN 89-09 (1988).

[5] I.S. Gradshteyn, I.M. Ryzhik, Table of integrals series and products, Academic Press 1980, p. 406.

[6] A. Hofmann, Physics of beam instabilities, Lecture notes in Physics 296. Proc of a Topical Course 'Frontiers of Particle Beams' held by the Joint US-CERN School on Particle Accel. at South Padre Island, Texas, Oct. 1986, ed. M. Month and S. Turner, Springer (1988) 99.

[7] D. Möhl, A.M. Sessler, The use of RF knockout for determination of the characteristics of the transverse coherent instability of an intensive beam, Proc. 8th Intern. Conf. on High Energy Accel., CERN (1971) p. 334.

[8] A. Hofmann, B. Zotter, Measurement of beam stability and coupling impedance by RF excitation', Proc. of the 1977 Particle Accel. Conf., IEEE Trans. on Nucl. Sci. NS 24-3 (1977) 1487.

[9] D. Boussard, Schottky noise and beam transfer function diagnostics, CAS CERN Accelerator School, Advanced Accelerator Physics, 1985; edited by S. Turner. CERN 87-03, p. 416.

[10] N.N. Lebedev, Special functions and their applications, Dover, New York 1972, p. 19.

[11] S. van der Meer, A different formulation of the longitudinal and transverse beam response, CERN/PS/AA/80-4 (1980).

[12] L.J. Laslett, V.K. Neil, A.M. Sessler, Transverse resistive instabilities of intense coasting beams in particle accelerators, Rev. Sci. Instr. 36 No. 4 (1965) 436.

[13] K. Hübner, V. Vaccaro, Dispersion relations and stability of coasting particle beams, CERN/ISR-TH-RF/69-23 (1969).

[14] C.N. Lashmore-Davis, Kinetic theory and the Vlasov equation, Proc. 1985 CAS CERN Accel. School on Advanced Accel. Physics, CERN 87-03 Vol. I (1987) 219.

[15] J.S. Bell, Hamiltonian mechanics, Proc. 1985 CAS CERN Accel. School on Advanced Accel. Physics, CERN 87-03 Vol. I (1987) 5.

[16] A. Hofmann, Kinetic theory, Proc. 1987 CAS CERN Accel. School on Advanced Accel. Physics, CERN 89-01 (1989) 25.

[17] J.D. Lawson, Particle beams and plasmas, CERN 76-09 (1976).

[18] J. Jowett, Introductory statistical mechanics for electron storage rings, Phys. of Part. Accel, Proc. SLAC 85 and Fermilab 84 Schools, AIP Conf. Proc. 153 (1987) 864.

[19] A. Hofmann, F. Pedersen, Bunches with local elliptic energy distributions, Proc. of the 1979 Part. Accel. Conf., IEEE Trans. on Nucl. Sci. NS 26-3, 3526 (1979).

[20] A.W. Chao, Equations for multiparticle dynamics, Lecture Notes in Physics 296, Proc. of a Topical Course Frontiers of Particle Beams, Joint US-CERN School of Particle Accel. at South Padre Island, Texas, Oct. 1986, ed. M. Month and S. Turner, Springer (1988) 55.

[21] H. Bruck; Accelerateurs Circulaires de Particules, Presses Universitaires de France 1966.

[22] V.K. Neil, A.M. Sessler, Longitudinal resistive instabilities of intense coasting beams in particle accelerators, Rev. Sci. Instr. 36 No. 4 (1965) 429.

[23] K. Hübner, V.G. Vaccaro, Dispersion relations stability of coasting particle beams, CERN Report ISR-TH/70-44 (1970).

[24] E. Keil, W. Schnell, Concerning longitudinal stability in the ISR, CERN Report ISR-TH-RF/69-48 (1969).

[25] J. Gareyte, Landau damping of the longitudinal quadrupole mode in SPEAR II, SLAC SPEAR-207 (1977).

[26] R. Boni, S. Guiducci, M. Serio, F. Tazzioli, F.H. Wang, Investigations on beam longitudinal transfer function and coupling impedance in ADONE, Frascati LNF-Div. Report, RM-23,1981.

BEAM INSTABILITIES

A. Hofmann,
CERN, Geneva, Switzerland

Abstract
Collective effects are introduced in the example of the longitudinal Robinson instability which can be generalized to cover all coupled bunch modes. We first introduce the concept of the impedance by approximating a cavity resonance by a RLC circuit. The response of such a resonator to pulse excitation gives the wake potential or Green function while a harmonic excitation reveals the concept and properties of the impedance. The interaction of a stationary circulating bunch with an impedance leads to an energy loss and to a shift of the incoherent synchrotron frequency. A bunch executing a synchrotron oscillation has spectral lines and harmonics of the revolution frequency with side bands spaced by the synchrotron frequency. The voltage induced in an impedance by these spectral lines can act back on the beam. A simple expression involving the impedances at the upper and lower side bands gives the growth rate of the Robinson instability. This can be generalized for a more complicated impedance, for the case of many bunches and also for higher modes of longitudinal oscillations.

1 INTRODUCTION

The motion of a single particle in a storage ring is determined by the external guide fields created by the dipole and quadrupole magnets and the RF system, by the initial conditions, and by the synchrotron radiation. The many particles contained in a high intensity beam represent a sizeable charge and current which act as a source of electromagnetic fields called *self fields*. These fields are modified by the boundary conditions imposed by the beam surroundings (vacuum chambers, cavities, etc.) and act back on the beam. This can lead to a *frequency shift* (change of the betatron or synchrotron frequency), to an increase of a small disturbance of the beam, i.e. an *instability*, or a *change of the particle distribution*, e.g. bunch lengthening. These phenomena are called *collective effects* since they are caused by a collective action of the many particles in the beam.

As an example we consider bunches in a storage ring going through a cavity; Fig. 1. Each bunch induces electromagnetic fields in this cavity which oscillate and slowly decay

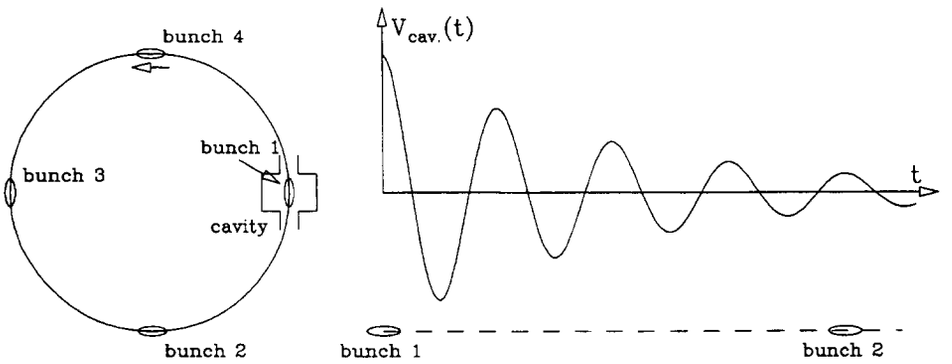


Figure 1: Field acting from one bunch to the next one

away. The next bunch, or the same bunch on the next turn, might find some field left and will be influenced by it. The phase of the field seen in the next turn can be such that a small initial synchrotron oscillation amplitude of the bunch is increased. In each turn the oscillation is amplified resulting in an exponentially growing instability.

In most cases the fields created by the beam are small compared to the guide fields and their effects can be treated as a *perturbation.* This is done in three steps:

a) First, the motion in the guide field and the stationary particle distribution are calculated.

b) A small disturbance of the bunch from its stationary motion is considered (betatron or synchrotron oscillation). The fields due to such a disturbance are determined taking the boundary condition imposed by the beam surroundings (impedance) into account.

c) The effect of these fields on the initial disturbance is investigated. If the amplitude of the disturbance is increased we have an *instability*, if it is decreased we have *damping*, or, if the frequency of the mode of oscillation is changed, we have a *frequency shift.*

For the case of small self-fields considered here the particle distribution in the bunch is given by external conditions (machine parameter, initial condition, synchrotron radiation) and is usually Gaussian in electron machines. As disturbances of the stationary distribution we consider some modes of oscillation which are orthogonal to each other such that the stability of each mode can be treated independently.

Strong self-fields, however, modify the particle distribution and also the modes of oscillation such that they are no longer independent. A self-consistent solution has to be found in this case and this can be a rather difficult task. It is usually only attempted for the case of bunch lengthening where the longitudinal particle distribution in the presence of an impedance is obtained. The beam position monitors and the transverse impedance can usually not "resolve" the local transverse particle distribution. The latter is therefore only in exceptional cases of importance for beam instabilities.

We distinguish between *single* and *multi-traversal* collective effects. For the first kind no memory of the induced field over one revolution or over the time interval between the passage of adjacent bunches is assumed. An example of a single-traversal effect is bunch lengthening. For multi-traversal effects the impedance has to have a memory such that one bunch can influence the next one or itself after one revolution. Such a memory can be provided by cavity-like objects with a relatively large quality factor Q.

Finally, we distinguish between *longitudinal* and *transverse* effects. In the first case a longitudinal impedance influences the synchrotron oscillation such that its amplitude grows or its frequency changes. The transverse impedance has a corresponding effect on the horizontal or vertical betatron oscillations.

Here, we will concentrate on longitudinal collective effects and derive the Robinson instability in some detail.

2 IMPEDANCE AND WAKE POTENTIAL OF A RESONATOR

2.1 Cavity resonance

Impedances and wake potentials have been treated extensively in the literature. We recapitulate here some of their essential properties on the simple case of a cavity resonance.

Cavities are the most likely objects to cause coupled-bunch mode instabilities since the induced fields oscillate for a relatively long time and provide a memory over the time interval between bunch passages. Such a cavity can be of a form which resembles an RCL circuit as shown in Fig. 2, and can be treated as such. The RCL circuit has a shunt impedance R_s, an inductance L and a capacity C. In a real cavity these three parameters cannot easily be separated. For this reason we use some other related parameters which can be measured directly: The *resonance frequency* ω_r, the *quality factor* Q and the

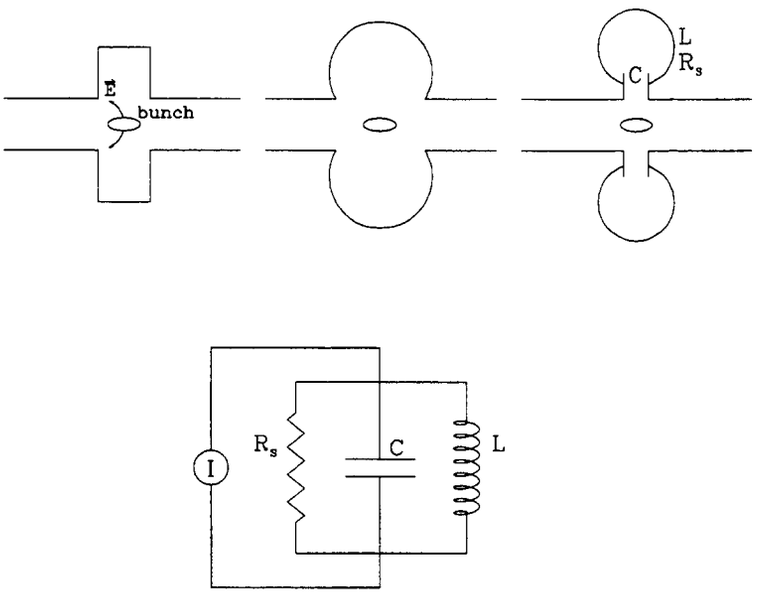


Figure 2: RLC circuit equivalent to a cavity resonance

damping rate α:

$$\omega_r = \frac{1}{\sqrt{LC}}\ ,\ \ Q = R_s\sqrt{\frac{C}{L}} = \frac{R_s}{L\omega_r} = R_s C\omega_r\ ,\ \ \alpha = \frac{\omega_r}{2Q}.$$

If this circuit is driven by a current I the voltages across each element are

$$V_R = I_R R_s\,,\ V_C = \frac{1}{C}\int I_C dt\,,\ V_L = L\frac{dI_L}{dt}$$

and have the relations to the currents

$$V_R = V_C = V_L = V\ ,\ \ I_R + I_C + I_L = I.$$

Differentiating with respect to t gives

$$\dot{I} = \dot{I}_R + \dot{I}_C + \dot{I}_L = \frac{\dot{V}}{R_s} + C\ddot{V} + \frac{V}{L}.$$

Using $L = R_s/(\omega_r Q)$ and $C = Q/(\omega_r R_s)$ gives the differential equation

$$\ddot{V} + \frac{\omega_r}{Q}\dot{V} + \omega_r^2 V = \frac{\omega_r R_s}{Q}\dot{I}.$$

The solution of the homogeneous equation is a damped oscillation

$$V(t) = \hat{V}\mathrm{e}^{-\alpha t}\cos\left(\omega_r\sqrt{1-\frac{1}{4Q^2}} + \phi\right)$$

or

$$V(t) = \mathrm{e}^{-\alpha t}\left(A\cos\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right) + B\sin\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right)\right).$$

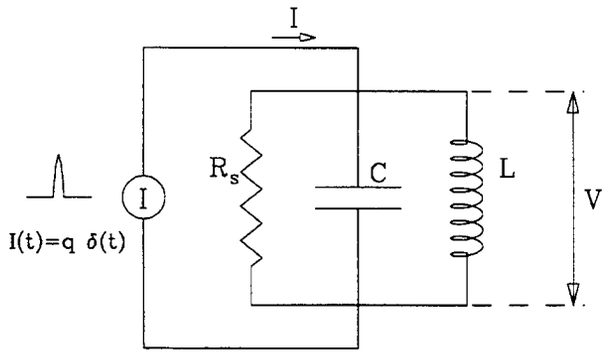


Figure 3: RLC circuit driven by a pulse

2.2 Wake potential

We now calculate the response of the RCL circuit (representing a cavity) to a delta function pulse (very short bunch), Fig. 3,

$$I(t) = q\delta(t).$$

The charge q induces a voltage in the capacity

$$V(0^+) = \frac{q}{C} = \frac{\omega_r R_s}{Q} q.$$

The energy stored in the capacitor

$$U = \frac{q^2}{C} = \frac{\omega_r R_s}{2Q} q^2 = \frac{V(0^+)}{2} q = k_{pm} q^2,$$

must be equal to the energy lost by the charge. Here we introduced the *parasitic mode loss factor* of a point charge

$$k_{pm} = \frac{U}{q^2} = \frac{\omega_r R_s}{2Q}$$

which is the energy loss normalized for the charge. The charged capacitor C will now discharge first through the resistor R_s and then also through the inductance L

$$\dot{V}(0^+) = -\frac{\dot{q}}{C} = -\frac{I_R}{C} = \frac{1}{C}\frac{V(0^+)}{R_s} = -\frac{\omega_r^2 R_s}{Q^2} q = \frac{2\omega_r k_{pm}}{Q} q.$$

The voltage in this resonance circuit has now the initial conditions

$$V(0^+) = 2k_{pm} q \text{ and } \dot{V}(0^+) = \frac{2\omega_r k_{pm}}{Q} q.$$

We take the solution of the homogeneous differential equation and its derivative

$$V(t) = \mathrm{e}^{-\alpha t}\left(A\cos\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right) + B\sin\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right)\right)$$

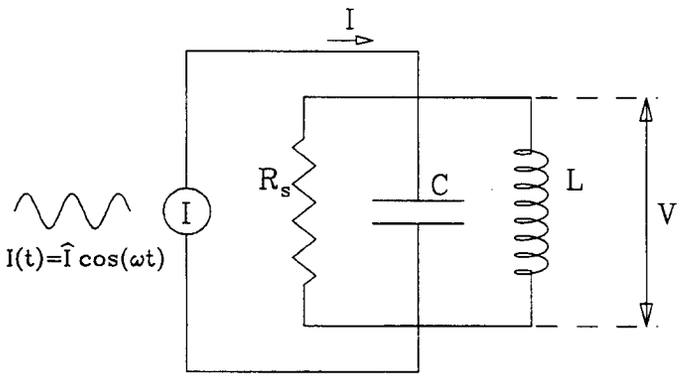


Figure 4: RLC circuit driven by a harmonic excitation

$$\dot{V}(t) = \mathrm{e}^{-\alpha t}\left(\left(-A\alpha + B\omega_r\sqrt{1-\frac{1}{4Q^2}}\right)\cos\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right) - \left(B\alpha + A\omega_r\sqrt{1-\frac{1}{4Q^2}}\right)\sin\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right)\right)$$

and satisfy the above initial conditions by

$$A = 2k_{pm}q \quad \text{and} \quad -A\alpha + B\omega_r\sqrt{1-\frac{1}{4Q^2}} = -\frac{2\omega_r k_{pm}}{Q}q.$$

We obtain the voltage in a resonator circuit excited at the time $t = 0$ by a δ-pulse $I(t) = q\delta t$

$$V(t) = 2qk_{pm}\mathrm{e}^{-\alpha t}\left(\cos\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right) - \frac{\sin\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right)}{2Q\sqrt{1-\frac{1}{4Q^2}}}\right).$$

This voltage is induced by a charge q going through the cavity at the time $t = 0$. A second point charge q' going through the cavity at a later time t will gain or loose the energy $U = q'V(t)$. This energy gain/loss per unit source and unit probe charge is called the *wake potential* of a point charge or also the *Green function* $G(t)$. For our resonator (cavity resonance) we have

$$G(t) = 2k_{pm}\mathrm{e}^{-\alpha t}\left(\cos\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right) - \frac{\sin\left(\omega_r\sqrt{1-\frac{1}{4Q^2}}\,t\right)}{2Q\sqrt{1-\frac{1}{4Q^2}}}\right)$$

which for a large quality factor $Q \gg 1$ simplifies to

$$G(t) \approx 2k_{pm}\mathrm{e}^{-\alpha t}\cos\left(\omega_r t\right).$$

2.3 Impedance

We use now a *harmonic* excitation of the circuit with a current $I = \hat{I}\cos(\omega t)$, Fig. 4. This is described by the differential equation

$$\ddot{V} + \frac{\omega_r}{Q}\dot{V} + \omega_r^2 V = -\frac{\omega_r R_s}{Q}\hat{I}\omega\sin(\omega t).$$

The solution of the homogeneous equation is a damped oscillation which disappears after some time. We are left with the particular solution of the form $V(t) = A\cos(\omega t) + B\sin(\omega t)$. Inserting this into the differential equation and separating cosine and sine terms gives

$$(\omega_r^2 - \omega^2)A + \frac{\omega_r\omega}{Q}B = 0 \text{ and } (\omega_r^2 - \omega^2)B - \frac{\omega_r\omega}{Q}A = -\frac{\omega_r\omega R_s}{Q}\hat{I}.$$

The voltage induced by the harmonic excitation of the resonator becomes

$$V(t) = \hat{I}R_s\frac{\cos(\omega t) - Q\frac{\omega_r^2-\omega^2}{\omega_r\omega}\sin(\omega t)}{1 + Q^2\left(\frac{\omega_r^2-\omega^2}{\omega_r\omega}\right)^2}.$$

This voltage has a cosine term which is *in phase* with the exciting current. It can absorb energy and is called the *resistive* term. The sine term of the voltage is *out of phase* with the exciting current and does not absorb energy, it is called the *reactive* term. The ratio between the voltage and current is called *impedance.* It is a *function of frequency* ω and has a resistive part $Z_r(\omega)$ and a reactive part $Z_i(\omega)$

$$Z_r(\omega) = R_s\frac{1}{1 + Q^2\left(\frac{\omega_r^2-\omega^2}{\omega_r\omega}\right)^2}\ ,\ Z_i(\omega) = -R_s\frac{Q\frac{\omega_r^2-\omega^2}{\omega_r\omega}}{1 + Q^2\left(\frac{\omega_r^2-\omega^2}{\omega_r\omega}\right)^2}.$$

The resonance excited by a current of the form $I(t) = \hat{I}\cos(\omega t)$ or $I(t) = \hat{I}\sin(\omega t)$ results in a voltage $V(t)$

$$I(t) = \hat{I}\cos(\omega t) \rightarrow V(t) = \hat{I}\left(Z_r(\omega)\cos(\omega t) + Z_i(\omega)\sin(\omega t)\right),$$

$$I(t) = \hat{I}\sin(\omega t) \rightarrow V(t) = \hat{I}\left(Z_r(\omega)\sin(\omega t) - Z_i(\omega)\cos(\omega t)\right).$$

2.4 Complex notation

We have used a harmonic excitation of the form

$$I(t) = \hat{I}\cos(\omega t) = \hat{I}\frac{e^{j\omega t} + e^{-j\omega t}}{2} \quad with\ \ 0 \le \omega \le \infty,$$

using positive frequencies only. A complex notation

$$I(t) = \hat{I}e^{j\omega t} \quad with \quad -\infty \le \omega \le \infty$$

involving positive and negative frequencies leads to more compact expressions and is often convenient. We take the differential equation

$$\ddot{V} + \frac{\omega_r}{Q}\dot{V} + \omega_r^2 = \frac{\omega_r R_s}{Q}\dot{I}$$

of the resonator voltage with the excitation $I(t) = \hat{I}\exp(j\omega t)$ and seek a solution of the form $V(t) = V_0\exp(j\omega t)$, where V_0 is in general complex and get

$$-\omega^2 V_0 e^{j\omega t} + j\frac{\omega_r - \omega}{Q}V_0 e^{j\omega t} + \omega_r^2 V_0 e^{j\omega t} = j\frac{\omega_r\omega R_s}{Q}\hat{I}e^{j\omega t}.$$

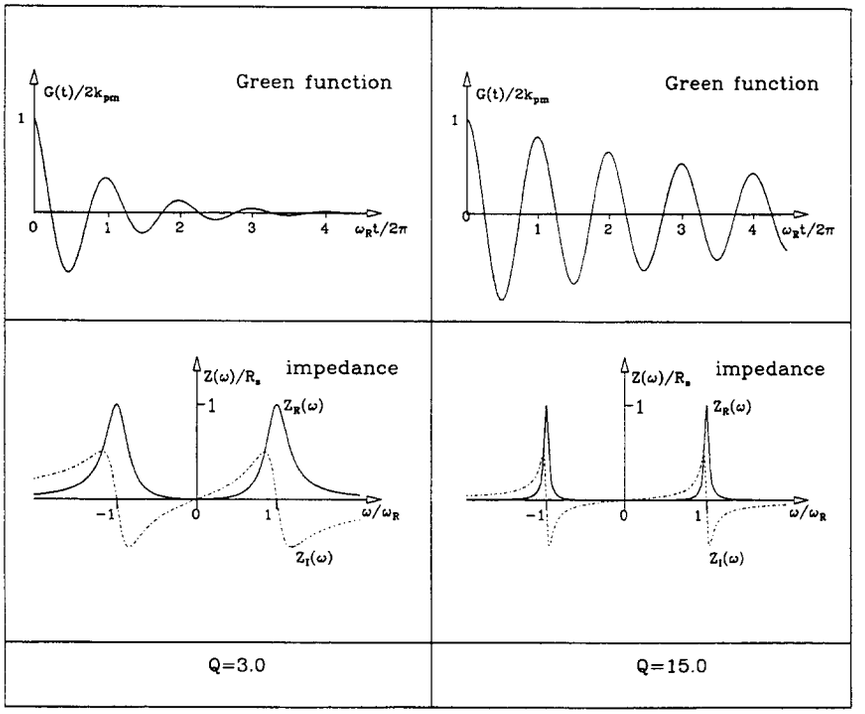


Figure 5: Green function and impedance of a resonance

The impedance, defined as the ratio V/I, is given by

$$Z(\omega) = \frac{V_0}{\hat{I}} = R_s \frac{j\frac{\omega_r \omega}{Q}}{\omega_r^2 - \omega^2 + jQ\frac{\omega_r \omega}{Q}} = R_s \frac{1 - jQ\frac{\omega^2 - \omega_r^2}{\omega\omega_r}}{1 + Q^2\left(\frac{\omega^2 - \omega_r^2}{\omega\omega_r}\right)^2} = Z_r(\omega) + jZ_i(\omega)$$

and has a real and an imaginary part, Fig. 5. For a large quality factor the impedance is only large for $\omega \approx \omega_r$ or $|\omega - \omega_r|/\omega_r = |\Delta\omega|\omega_r \ll 1$ and can be simplified to

$$Z(\omega) \approx R_s \frac{1 - j2Q\frac{\Delta\omega}{\omega_r}}{1 + 4Q^2\left(\frac{\Delta\omega}{\omega_r}\right)^2}.$$

The resonator impedance has some specific properties:

$$\begin{aligned} \omega &= \omega_r \rightarrow Z_r(\omega_r) \text{ has a maximum, } Z_i(\omega_r) = 0 \\ |\omega| &< \omega_r \rightarrow Z_i(\omega) > \omega_r \text{ (inductive)} \\ |\omega| &> \omega_r \rightarrow Z_i(\omega) < 0 \text{ (capacitive)} \end{aligned}$$

and some properties which apply to any impedance or wake potential

$$\begin{aligned} Z_r(\omega) &= Z_r(-\omega) \ , \ Z_i(\omega) = -Z_i(-\omega), \\ Z(\omega) &= \int_{-\infty}^{\infty} G(t)\mathrm{e}^{-j\omega t}dt, \ Z(\omega) = \text{Fourier transform of } G(t), \\ t &< 0 \rightarrow G(t) = 0 \text{ no fields before particle arrives.} \end{aligned}$$

Caution; sometimes one uses $I(t) = \hat{I}\mathrm{e}^{-i\omega t}$ instead of $I(t) = \hat{I}e^{j\omega t}$, this reverses the sign $Z_i(\omega)$.

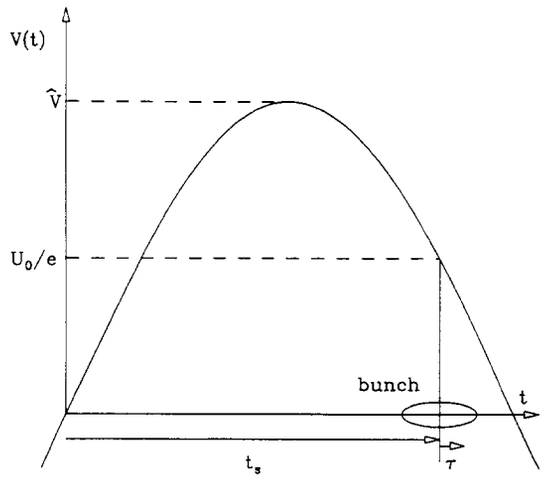


Figure 6: Longitudinal beam dynamics impedance

2.5 Review of the longitudinal dynamics

A particle with a momentum deviation Δp has a different closed orbit which is radially displaced by $\Delta x = D_x \Delta p/p$ with D_x being the dispersion. As a result the orbit length L, the revolution time T_0 and the revolution frequency ω_0 are changed

$$\frac{\Delta L}{L} = \alpha_c \frac{\Delta p}{p} \ , \quad \frac{\Delta \omega_0}{\omega_0} = -\frac{\Delta T_0}{T_0} = -\left(\alpha_c - \frac{1}{\gamma^2}\right)\frac{\Delta p}{p} = -\eta_c \frac{\Delta p}{p}$$

with α_c being the momentum compaction and $\eta_c = \alpha_c - 1/\gamma^2$. There is a transition energy $E_T = m_0 c^2 \gamma_T$ with $\gamma_T = 1/\alpha_c^2$ for which the dependence of the revolution frequency on momentum (or energy) changes sign

$$E > E_T \ \rightarrow \ \frac{1}{\gamma^2} < \alpha_c \ \rightarrow \ \eta_c > 1 \ \rightarrow \ \omega_0 \text{ decreases with } \Delta E$$

$$E < E_T \ \rightarrow \ \frac{1}{\gamma^2} > \alpha_c \ \rightarrow \ \eta_c < 1 \ \rightarrow \ \omega_0 \text{ increases with } \Delta E \, .$$

We will assume that the particles are ultra relativistic in which case $\Delta p/p \approx \Delta E/E = \epsilon$. In the presence of an RF system and of an energy loss per turn U due to synchrotron radiation or an impedance, a circulating particle has a gain or loss δE in energy of

$$\delta E = e\hat{V} \sin(h\omega_0(t_s + \tau)) - U$$

or in relative energy $\delta E/E = \delta\epsilon$

$$\frac{\delta E}{E} = \delta\epsilon = \frac{e\hat{V} \sin(h\omega_0(t_s + \tau))}{E} - \frac{U}{E}.$$

with t_s being the synchronous arrival time of the particle in the cavity and τ the deviation from it as illustrated in Fig. 6. We introduce the synchronous phase angle $\phi_s = h\omega_0 t_s$ and assume $\tau \ll T_0$ which allows us to develop the trigonometric function

$$\delta\epsilon = \frac{e\hat{V} \sin(\phi_s)}{E} + \frac{h\omega_0 e\hat{V} \cos\phi_s}{E}\tau - \frac{U}{E}.$$

The energy gain per turn is very small $\delta E \ll E$ and we can make a smooth approximation

$$\frac{\delta E}{E} = \dot{\epsilon} T_0 = \dot{\epsilon} \frac{2\pi}{\omega_0}$$

$$\dot{\epsilon} = \frac{\omega_0 e\hat{V} \sin\phi_s}{2\pi E} + \frac{\omega_0^2 he\hat{V}\cos\phi_s}{2\pi E}\tau - \frac{\omega_0}{2\pi}\frac{U}{E}. \tag{1}$$

The energy loss U suffered by a particle is in general a function of its relative energy deviation ϵ and its deviation τ from the synchronous and, to first order, can be developed as

$$U(\epsilon,\tau) \approx U_0 + \frac{\partial U}{\partial E}\Delta E + \frac{\partial U}{\partial t}\tau.$$

This leads to an expression for the time derivative of the energy loss

$$\dot{\epsilon} = \frac{\omega_0 e\hat{V}\sin\phi_s}{2\pi E} + \frac{\omega_0^2 he\hat{V}\cos\phi_s}{2\pi E}\tau - \frac{\omega_0}{2\pi}\frac{U_0}{E} - \frac{\omega_0}{2\pi}\frac{dU}{dE}\epsilon - \frac{\omega_0}{2\pi}\frac{1}{E}\frac{dU}{dt}\tau.$$

To have equilibrium for the synchronous particle $\epsilon = 0\,,\ \tau = 0$ we have

$$U_0 = e\hat{V}\sin\phi_s.$$

With this and using $\dot{\tau} = \omega_0 \Delta T_0/2\pi = \eta_c \epsilon$ we get a system of two first order differential equations

$$\begin{aligned} \dot{\epsilon} &= \omega_0^2 \frac{he\hat{V}\cos\phi_s}{2\pi E}\tau - \frac{\omega_0}{2\pi}\frac{dU}{dE}\epsilon - \frac{1}{E}\frac{\omega_0}{2\pi}\frac{dU}{dt}\tau \\ \dot{\tau} &= \eta_c \epsilon\,. \end{aligned}$$

They can be combined into one second-order equation

$$\ddot{\epsilon} + \frac{\omega_0}{2\pi}\frac{dU}{dE}\dot{\epsilon} - \frac{\omega_0^2 h\eta_c e\hat{V}\cos\phi_s}{2\pi E}\epsilon - \frac{\eta_c}{E}\frac{\omega_0}{2\pi}\frac{dU}{dt}\epsilon = 0$$

which is the equation of a damped oscillation. Using

$$\omega_{s0}^2 = -\omega_0^2 \frac{h\eta_c e\hat{V}\cos\phi_s}{2\pi E}\ ,\ \alpha_s = \frac{1}{2}\frac{\omega_0}{2\pi}\frac{dU}{dE}\,,$$

seeking a solution of the form $e^{j\omega t}$, and assuming $\alpha_s \ll \omega_{s0}$ we get

$$-\omega^2 + j\omega\alpha_s + (\omega_{s0}^2 + \frac{\omega_0}{2\pi}\frac{\eta_c}{E}\frac{dU}{dt}) = 0$$

$$\omega = j\alpha_s \pm \sqrt{(\omega_{s0}^2 + \frac{\omega_0}{2\pi}\frac{\eta_c}{E}\frac{dU}{dt}) - \alpha_s^2} \approx j\alpha_s \pm (\omega_{s0} + \frac{1}{2}\frac{\omega_0}{2\pi}\frac{\eta_c}{\omega_{s0}E}\frac{dU}{dt})\,.$$

Calling

$$\Delta\omega_i = \frac{1}{2}\frac{\omega_0}{2\pi}\frac{\eta_c}{\omega_{s0}E}\frac{dU}{dt}$$

gives

$$\epsilon = A\left(e^{(-\alpha_s + j(\omega_{s0}+\Delta\omega_i)t} + Be^{(-\alpha_s - j(\omega_{s0}+\Delta_i)t}\right).$$

For the initial conditions $\epsilon(t) = \hat{\epsilon},\ \dot{\epsilon}(0) = -\alpha_s\hat{\epsilon}$ we get $A = B = \hat{\epsilon}/2$ and

$$\epsilon(t) = \hat{\epsilon}\ e^{-\alpha_s t}\cos((\omega_{s0} + \Delta\omega_i)t).$$

In the absence of any energy loss U we have

$$\epsilon(t) = \hat{\epsilon}\cos(\omega_{s0}t + \phi)$$

with

$$\omega_{s0}^2 = -\omega_0^2 \frac{h\eta_c e\hat{V}\cos\phi_s}{2\pi E}$$

In order to get a stable oscillation we need $\omega_{s0}^2 > 0$ which leads to the conditions

$$E > E_T\ \eta_c < 0\ \rightarrow \cos\phi_s < 0\ ,\ \ E < E_T\ \eta_c > 0\ \rightarrow \cos\phi_s > 0.$$

For stability in the presence of an energy loss U we need in addition

$$\alpha_s = \frac{1}{2}\frac{\omega_0}{2\pi}\frac{dU}{dE} > 0.$$

In other words the energy loss U has to increase for a positive energy deviation of the beam.

3 A STATIONARY BUNCH INTERACTING WITH AN IMPEDANCE

3.1 Spectrum of a stationary bunch

We consider now a bunch which represents in a single traversal the current $I(t)$. Its Fourier transform is

$$\tilde{I}(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} I(t)\mathrm{e}^{-j\omega t}dt. \tag{2}$$

We assume the form of the bunch to be symmetric

$$I(-t) = I(t)$$

which leads to a Fourier transform having only a real part and being symmetric in ω

$$\tilde{I}(-\omega) = \tilde{I}(\omega).$$

This assumption is used for convenience to reduce the number of terms which have to be carried along in some calculation. Since, in most practical applications, the bunches are to a good approximation symmetric, this represents a minor restriction which could easily be removed. The currents of a bunch with Gaussian distribution in time and frequency domain are illustrated in Fig. 7 and given by the expressions

$$I(t) = \frac{q}{\sqrt{2\pi}\sigma_t}\mathrm{e}^{-\frac{t^2}{2\sigma_t^2}}\ \ ;\ \ \tilde{I}(\omega) = q\mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \tag{3}$$

where $q = N_b e$ is the total charge of the N_b particles in a bunch. The rms width of the bunch and its spectrum are σ_t and σ_ω which are related by

$$\sigma_t = \frac{1}{\sigma_\omega}.$$

Next we investigate the case of a circulating bunch having repetitive passages at a given location. For a stationary bunch having no synchrotron oscillations the observed current is of the form

$$I_k(t) = \sum_{k=-\infty}^{\infty} I(t - kT_0).$$

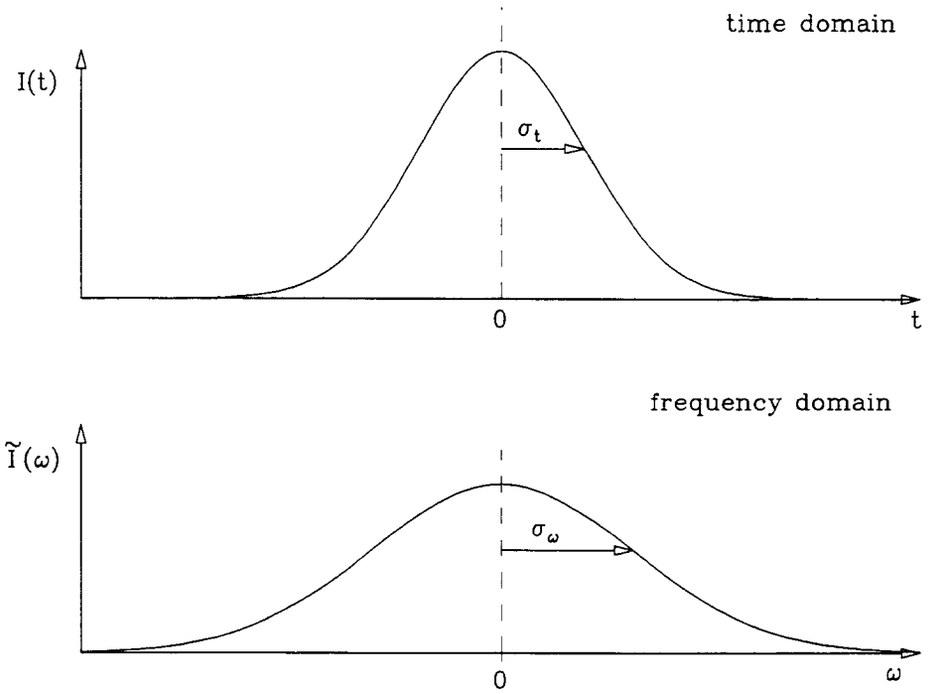


Figure 7: Single passage of a bunch in time and frequency domain

This is not a very useful expression for applications. Since the current is periodic it is natural to express it in a Fourier series

$$I_k(t) = I_0 + 2\sum_1^\infty I_p \cos(p\omega_0 t) \tag{4}$$

with

$$I_p = \frac{1}{T_0}\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} I(t)\cos(p\omega_0 t)dt \text{ and } I_0 = \langle I\rangle = \frac{1}{T_0}\int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} I(t)dt = \frac{q}{T_0}. \tag{5}$$

Comparing the Fourier transform Eq. (2) with the terms of the Fourier series Eq. (5) we find the relation

$$I_p = \frac{\omega_0}{\sqrt{2\pi}}\tilde{I}(p\omega_0).$$

For a Gaussian bunch Eq. (3) we get

$$I_p = \frac{q}{T_0} e^{-\frac{p^2\omega_0^2}{2\sigma_\omega^2}}.$$

At low frequencies we have $I_p \approx I_0$. The current in time and frequency domain of such a circulating bunch is shown in Fig. 8.

It should be noted that for the frequency component I_p of the current we use a value adapted to positive and negative frequencies. If only positive frequencies are used the current component at $p\omega_0$ is $2I_p$.

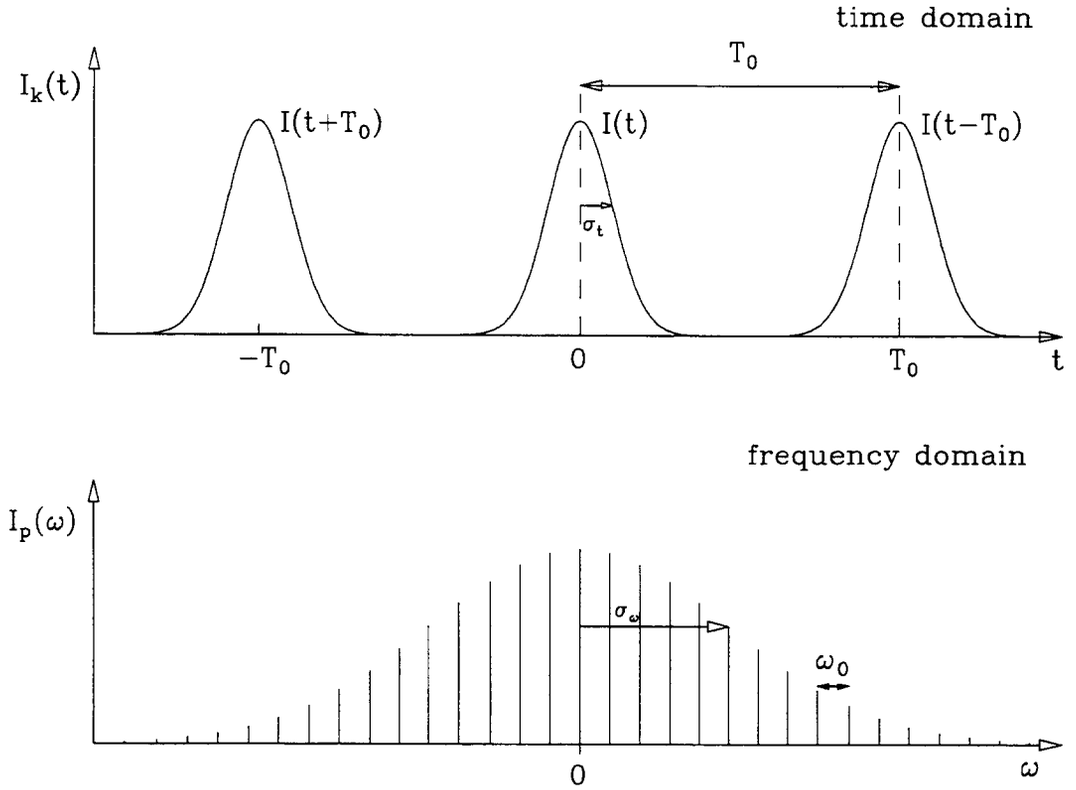


Figure 8: Multiple passage of a bunch in time and frequency domain

3.2 Voltage induced by the stationary bunch

In the presence of a cavity resonance or any *general impedance* $Z(\omega)$ the circulating stationary bunch induces a voltage which is in frequency domain

$$\tilde{V}_k(\omega) = \tilde{I}_k(\omega)Z(\omega) = \omega_0 \sum_{p=-\infty}^{\infty} \tilde{I}(\omega)\delta(\omega - p\omega_0)Z(\omega)$$

and in time domain

$$\begin{aligned} V_k(t) &= \frac{\omega_0}{\sqrt{2\pi}} \sum_{p=-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{I}(\omega)\delta(\omega - p\omega_0)Z(\omega)e^{j\omega t}d\omega \\ &= \frac{\omega_0}{\sqrt{2\pi}} \sum_{p=-\infty}^{\infty} \tilde{I}(p\omega_0)Z(p\omega_0)e^{jp\omega_0 t} = \sum_{p=-\infty}^{\infty} I_p Z(p\omega_0)e^{jp\omega_0 t}. \end{aligned}$$

By combining positive and negative frequencies and observing the symmetry conditions $Z_r(-\omega) = Z_r(\omega)$, $Z_i(-\omega) = -Z_i(\omega)$ and the fact that $Z(0) = 0$ we get a real expression

$$V_k(t) = 2\sum_{p=1}^{\infty} I_p \left(Z_r(p\omega_0)\cos(p\omega_0 t) - Z_i(p\omega_0)\sin(p\omega_0 t) \right). \tag{6}$$

We calculate the average induced voltage $\langle V \rangle$ seen by the particle in the bunch

$$\langle V \rangle = \frac{1}{I_0 T_0} \int_{-T_0/2}^{T_0/2} I_k(t)V_k(t)dt.$$

With the expressions Eq. (4) for the current and Eq. (6) for the voltage we get

$$\langle V\rangle = \frac{4}{I_0 T_0}\sum_{p=1}^{\infty}\sum_{p'=1}^{\infty} I'_p I_p \left(Z_r(p\omega_0 \int_{-T_0/2}^{T_0/2} \cos(p'\omega_0 t)\cos(p\omega_0 t)dt - Z_r(p\omega_0 \int_{-T_0/2}^{T_0/2} \cos(p'\omega_0 t)\sin(p\omega_0 t)dt \right).$$

The first integral vanishes except for $p' = p$ in which case it has the value $T_0/2$, and the second integral always vanishes. This leads to

$$\langle V\rangle = \frac{1}{I_0}\sum_{p=-\infty}^{\infty} |I_p|^2 Z_r(p\omega_0) = \frac{2}{I_0}\sum_{p=1}^{\infty} |I_p|^2 Z_r(p\omega_0). \tag{7}$$

We will also need the average voltage slope

$$\langle \frac{dV}{dt}\rangle = \frac{1}{I_0 T_0}\int_{-T_0/2}^{T_0/2} I_k(t)\frac{dV_k(t)}{dt}dt.$$

With the same method we used for the average voltage we obtain

$$\langle \frac{dV}{dt}\rangle = -\frac{\omega_0}{I_0}\sum_{p=-infty}^{\infty} p|I_p|^2 Z_i(p\omega_0) = -\frac{2\omega_0}{I_0}\sum_{p=1}^{\infty} p|I_p|^2 Z_i(p\omega_0).$$

3.3 Energy loss per turn of a stationary circulating bunch

The energy W_b lost by the whole circulating stationary bunch in one turn due to the impedance $Z(\omega)$ can be obtained from the average voltage Eq. (7)

$$W_b = q\langle V\rangle = \frac{2q}{I_0}\sum_{p=1}^{\infty} |I_p|^2 Z_r(p\omega_0)$$

where $q = eN_b$ is the total charge of the bunch. The average energy loss U of a particle in the bunch is

$$U = \frac{W_b}{N_b} = \frac{2e}{I_0}\sum_{p=1}^{\infty} |I_p|^2 Z_r(p\omega_0) = \frac{2T_0}{N_b}\sum_{p=1}^{\infty} |I_p|^2 Z_r(p\omega_0).$$

We can normalize the loss W_b by the square of charge (the charge inducing the voltage and the same charge suffering an energy loss) to get the so-called parasitic mode loss factor or a bunch

$$k_{pm} = \frac{W_b}{q^2} = \frac{U}{eq} = \frac{2}{qI_0}\sum_{p=1}^{\infty} |I_p|^2 Z_r(p\omega_0) = \frac{2T_0}{q^2}\sum_{p=1}^{\infty} |I_p|^2 Z_r(p\omega_0).$$

This parameter depends on the bunch length. For a short bunch the spectrum extends to higher frequencies. The parameter k_{pm} is therefore expected to increase with decreasing bunch length.

If the impedance is broad band and does not contain resonances of bandwidth smaller than the revolution frequency, the above sum can be approximated by an integral

$$k_{pm} \approx \frac{1}{q^2}\int_{-\infty}^{\infty} |\tilde{I}(\omega)|^2 Z_r(\omega)d\omega.$$

3.4 Incoherent synchrotron frequency shift

We take now the case of a stationary bunch in the presence of an impedance $Z(\omega) = Z_r(\omega) + jZ_i(\omega)$. As we saw before, the bunch induces an average voltage in the resistive part of the impedance

$$\langle V \rangle = \frac{2}{I_0} \sum_{p=1}^{\infty} |I_p|^2 Z_r(p\omega_0) \tag{8}$$

and an averaged voltage slope in the reactive part

$$\langle \frac{dV}{dt} \rangle = -\frac{2\omega_0}{I_0} \sum_{p=1}^{\infty} p|I_p|^2 Z_i(p\omega_0), \tag{9}$$

both being independent of the energy. We have to include these voltages in the equation of the synchrotron motion

$$\dot{\epsilon} = \frac{e\hat{V}\sin\phi_s\omega_0}{2\pi E} + \frac{\omega_0^2 he\hat{V}\cos\phi_s}{2\pi E}\tau - \frac{\omega_0}{2\pi}\frac{e\langle V\rangle}{E} - \frac{\omega_0 e}{2\pi E}\langle\frac{dV}{dt}\rangle\tau.$$

With the condition $e\hat{V}\sin\phi_s = e\langle V\rangle$ we find

$$\begin{aligned} \dot{\epsilon} &= \omega_0^2 \frac{he\hat{V}\cos\phi_s}{2\pi E}\tau + \frac{\omega_0 e}{2\pi E}\langle\frac{dV}{dt}\rangle\tau \\ \dot{\tau} &= \eta_c \epsilon\,, \end{aligned}$$

or, combined into a second-order equation,

$$\ddot{\epsilon} + -\left(\frac{\omega_0^2 h\eta_c e\hat{V}\cos\phi_s}{2\pi E} + \frac{\eta_c}{E}\frac{\omega_0}{2\pi}e\langle\frac{dV}{dt}\rangle\right)\epsilon = 0.$$

The solution is an undamped oscillation with the frequency given by

$$\omega_s^2 = \omega_{s0}^2 + \frac{2\omega_0^2\eta_c e}{2\pi E I_0}\sum_{p=1}^{\infty} p|I_p|^2 Z_i(p\omega_0)$$

which can be written as

$$\omega_s^2 = \omega_{s0}^2\left(1 + \frac{2\omega_0^2\eta_c e}{2\pi E\omega_{s0}^2 I_0}\sum_{p=1}^{\infty} p|I_p|^2 Z_i(p\omega_0)\right) = \omega_{s0}^2\left(1 + \frac{2}{h\hat{V}\cos\phi_s I_0}\sum_{p=1}^{\infty} p|I_p|^2 Z_i(p\omega_0)\right).$$

There is a shift of the incoherent synchrotron frequency. For a small effect this shift can be expressed as

$$\frac{\Delta\omega_{si}}{\omega_{s0}} = \frac{1}{I_0 h\hat{V}\cos\phi_s}\sum_{p=1}^{\infty} pI_p^2 Z_i(p\omega_0). \tag{10}$$

4 INTERACTION OF AN OSCILLATING BUNCH WITH A CAVITY

4.1 Spectrum of an oscillating bunch

We consider now a bunch which executes a synchrotron oscillation with frequency $\omega_s = \omega_0 Q_s$. This results in a modulation of its passage time t_k at a cavity in successive turns k as illustrated in Fig. 9

$$t_k = kT_0 + \tau_k,$$

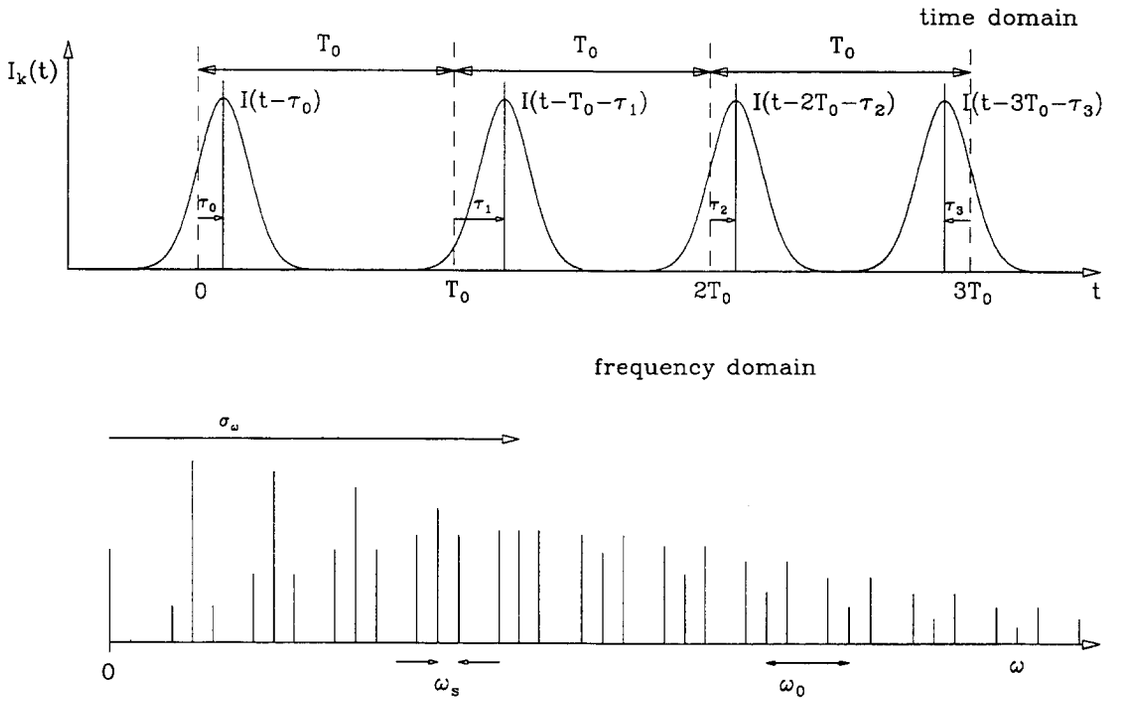


Figure 9: Oscillating bunch in time and frequency domain

with

$$\tau_k = \hat{\tau}\cos(2\pi Q_s k),$$

where k is the revolution number and $\hat{\tau}$ the amplitude of the modulation. The current represented by this oscillating bunch is given in time domain by

$$I_k(t) = \sum_{-\infty}^{\infty} I(t - kT_0 - \tau_k).$$

To get the Fourier transform of this current we use again the shift theorem and the symmetry condition $I(t) = I(-t)$

$$\tilde{I}_k(\omega) = \tilde{I}(\omega) \sum_{k=-\infty}^{\infty} \mathrm{e}^{-j\omega(kT_0+\tau_k)}.$$

We assume now that the oscillation is small $\omega\hat{\tau} \ll 1$ for all frequencies ω contained in the bunch spectrum $\tilde{I}(\omega)$ and approximate

$$\begin{aligned}
\tilde{I}_k(\omega) &= \tilde{I}(\omega) \sum_{k=-\infty}^{\infty} \mathrm{e}^{-j\omega(kT_0)} \left(1 - j\omega\hat{\tau}\cos(2\pi Q_s k)\right) \\
&= \tilde{I}(\omega) \sum_{k=-\infty}^{\infty} \mathrm{e}^{-j\omega kT_0} - j\frac{\omega\hat{\tau}}{2}\tilde{I}(\omega \sum_{k=-\infty}^{\infty} \left(\mathrm{e}^{-jk(\omega T_0 - 2\pi Q_s)} + \mathrm{e}^{-jk(\omega T_0 + 2\pi Q_s)}\right).
\end{aligned}$$

We use again the relations

$$\sum_{k=-\infty}^{\infty} \mathrm{e}^{-jkx} = 2\pi \sum_{p=-\infty}^{\infty} \delta(x - 2\pi p) \quad \text{and} \quad \delta(ax) = \frac{1}{a}\delta(x)$$

and get

$$\tilde{I}_k(\omega) = \omega_0 \tilde{I}(\omega) \left[\sum_{p=-\infty}^{\infty} \delta(\omega - p\omega_0) - j\frac{\omega\hat{\tau}}{2} \sum_{p=-\infty}^{\infty} \left(\delta(\omega - p\omega_0 - \omega_s) + \delta(\omega - p\omega_0 + \omega_s)\right) \right]. \tag{11}$$

This is the Fourier-transformed current of a bunch executing a synchrotron oscillation. It is represented by a line spectrum having spectral lines at harmonics of the revolution frequency $p\omega_0$ due to the stationary bunch motion and side bands caused by the bunch oscillation. This is not astonishing since the stationary motion is periodic which leads to a simple line spectrum shown in an earlier section. The modulation of the passage time with the synchrotron frequency is expected to lead to side bands spaced by ω_s from the revolution harmonics. One would think that the above spectrum could be obtained directly using the theory of phase modulation without the relatively lengthy derivation given above. However there is a subtlety concerning the size of the spectral lines which will be discussed below. We prefer to get this current in time domain and apply the inverse Fourier transform

$$I_k(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{I}_k(\omega) \mathrm{e}^{j\omega t} d\omega \tag{12}$$

to the above expression, so obtaining

$$I_k(t) = \frac{\omega_0}{\sqrt{2\pi}} \sum_{p=-\infty}^{\infty} \Big[\tilde{I}(p\omega_0)\mathrm{e}^{jp\omega_0 t} - j\frac{\omega_0}{\sqrt{2\pi}}\frac{\omega_0\hat{\tau}}{2} \left((p - Q_s)\tilde{I}((p - Q_s)\omega_0)\mathrm{e}^{j(p-Q_s)\omega_0 t} + (p + Q_s)\tilde{I}((p + Q_s)\omega_0)\mathrm{e}^{j(p+Q_s)\omega_0 t} \right) \Big].$$

This is a relatively complicated way to express the current of an oscillating bunch but is, however, of some use for certain calculations. We can simplify it with some approximations. The bunch length is usually much shorter than the circumference of the machine. As a consequence its spectrum is much larger than the revolution frequency ω_0. We can therefore neglect the very small difference in the bunch spectrum between the harmonics of the revolution frequency and its side band

$$\tilde{I}(p\omega_0) \approx \tilde{I}((p + Q_s)\omega_0) \approx \tilde{I}((p - Q_s)\omega_0) = \frac{\sqrt{2\pi}}{2\omega_0} I_p.$$

We use now the current component I_p at the frequency $p\omega_0$ instead of the Fourier transform of $\tilde{I}(p\omega_0)$, and represent the current of an oscillating bunch as a standard Fourier series

$$I_k(t) = \sum_{p=-\infty}^{\infty} I_p \left[\mathrm{e}^{jp\omega_0 t} - j\frac{\omega_0\hat{\tau}}{2} \left((p - Q_s)\mathrm{e}^{j(p-Q_s)\omega_0 t} + (p + Q_s)\mathrm{e}^{j(p+Q_s)\omega_0 t} \right) \right].$$

By combining terms with positive and negative values of the summing parameter p and using the fact that for a symmetric bunch we have $\tilde{I}(\omega) = \tilde{I}(-\omega)$ we can express

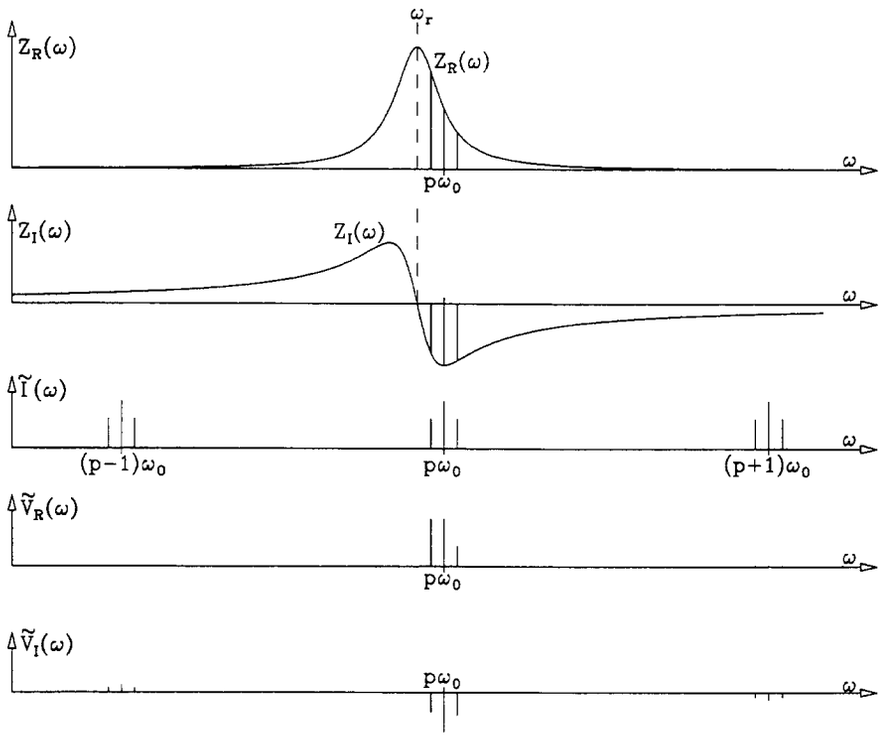


Figure 10: Voltage induced by an oscillating bunch in a narrow band impedance

the current of the oscillating bunch with real functions

$$I_k(t) = I_0 + 2\sum_{p=1}^{\infty} I_p \left[\cos(p\omega_0 t) + \frac{\omega_0\hat{\tau}}{2}\left((p-Q_s)\sin((p-Q_s)\omega_0 t) + (p+Q_s)\sin((p+Q_s)\omega_0 t)\right)\right]. \quad (13)$$

The line spectrum of the current represented by an oscillating bunch is shown in Fig. 9.

We can split the trigonometric function in the expression Eq. (14) for the current

$$\begin{aligned} I_k(t) = I_0 + 2\sum_{p=1}^{\infty} I_p \Big[&\cos(p\omega_0 t) \\ &+\frac{\omega_0\hat{\tau}}{2}\left((p-Q_s)(\sin(p\omega_0 t)\cos(Q_s\omega_0 t) + \cos(p\omega_0 t)\sin(Q_s\omega_0 t))\right. \\ &\left.+(p+Q_s)(\sin(p\omega_0 t)\cos(Q_s\omega_0 t) - \cos(p\omega_0 t)\sin(Q_s\omega_0 t))\right)\Big]. \end{aligned} \quad (14)$$

4.2 Voltage induced by an oscillating bunch

We calculate the voltage induced by the current $I_k(t)$ in an impedance $Z(\omega)$. The Fourier transform of this voltage is given by

$$\tilde{V}_k(\omega) = \tilde{I}_k(\omega)Z(\omega).$$

To get the voltage in time domain we use the expression Eq. (11) for the current and apply the inverse Fourier transform Eq. (12)

$$V_k(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{I}_k(\omega) Z(\omega) \mathrm{e}^{j\omega t}$$

which, with our approximation, can be written as

$$V_k(t) = 2 \sum_{p=1}^{\infty} \Big[I_p Z(p\omega_0) \mathrm{e}^{p\omega_0 t} + \frac{\omega_0 \hat{\tau}}{2} I_p \left((p+Q_s) Z((p+Q_s)\omega_0) \mathrm{e}^{j((p+Q_s)\omega_0 t)} + (p-Q_s) Z((p-Q_s)\omega_0) \mathrm{e}^{j((p-Q_s)\omega_0 t)} \right) \Big].$$

We consider now a narrow-band impedance $\delta\omega < \omega_0$ and calculate the voltage which is induced in it by an oscillating bunch. Due to the small bandwidth we have to consider only one revolution harmonic p and its two side bands to the induced voltage as indicated in Fig. 10.

We start with the resistive part Z_r of the impedance and get for the voltage

$$\begin{aligned} V_{kr}(t) &= 2 I_p Z_r(p\omega_0) \cos(p\omega_0 t) \\ &+ \omega_0 \hat{\tau} I_p \left((p+Q_s) Z_r((p+Q_s)\omega_0) \sin((p+Q_s)\omega_0 t) \right. \\ &\quad \left. + (p-Q_s) Z_r((p-Q_s)\omega_0) \sin((p-Q_s)\omega_0 t) \right). \end{aligned}$$

We split the trigonometric functions into two parts

$$\begin{aligned} V_{kr}(t) &= 2 I_p Z_r(p\omega_0) \cos(p\omega_0 t) \\ &+ \omega_0 \hat{\tau} I_p \left[(p+Q_s) Z_r((p+Q_s)\omega_0) \left(\sin(p\omega_0 t)\cos(\omega_s t) + \cos(p\omega_0 t)\sin(\omega_s t) \right) \right. \\ &\quad \left. (p-Q_s) Z_r((p-Q_s)\omega_0) \left(\sin(p\omega_0 t)\cos(\omega_s t) - \cos(p\omega_0 t)\sin(\omega_s t) \right) \right]. \end{aligned}$$

The synchrotron motion consists of a modulation of the arrival time each revolution k. We approximate it now as a modulation in time

$$\tau_k = \hat{\tau} \cos(2\pi Q_s k) \quad \rightarrow \quad \tau = \hat{\tau} \cos(\omega_s t) \ , \ \dot{\tau} = -\omega_s \hat{\tau} \sin(\omega_s t).$$

Since the synchrotron phase advance $2\pi Q_s$ per turn is very small in all practical cases this is a very good approximation. Using these expressions we get for the voltage induced in the resistive impedance

$$\begin{aligned} V_{kr}(t) &= 2 I_p \Big[Z_r(p\omega_0) \cos(p\omega_0 t) \\ &+ \frac{\omega_0}{2} \left((p+Q_s) Z_r((p+Q_s)\omega_0) \left(\sin(p\omega_0 t)\tau - \cos(p\omega_0 t)\frac{\dot{\tau}}{\omega_s} \right) \right. \\ &\quad \left. + (p-Q_s) Z_r((p-Q_s)\omega_0) \left(\sin(p\omega_0 t)\tau + \cos(p\omega_0 t)\frac{\dot{\tau}}{\omega_s} \right) \right) \Big]. \end{aligned}$$

We can also express the current Eq. (14) of the oscillating bunch with τ and $\dot{\tau}$

$$\begin{aligned} I_k(t) &= I_0 + 2 \sum_{p=1}^{\infty} I_p \Big[\cos(p\omega_0 t) \qquad (15) \\ &+ \frac{\omega_0}{2} \left((p+Q_s) \left(\sin(p\omega_0 t)\tau - \cos(p\omega_0 t)\frac{\dot{\tau}}{\omega_s} \right) \right. \\ &\quad \left. + (p-Q_s) \left(\sin(p\omega_0 t)\tau + \cos(p\omega_0 t)\frac{\dot{\tau}}{\omega_s} \right) \right) \Big]. \qquad (16) \end{aligned}$$

Next we calculate the average voltage seen by the particles in one revolution

$$\langle V_r \rangle = \frac{\int I_k(t) V_k(t) dt}{\int I_k(t) dt} = \frac{1}{I_0 T_0} \int_0^{T_0} I_k(t) V_k(t) dt.$$

In carrying out this integration we approximate for small oscillation amplitudes and take only linear terms in τ and $\dot{\tau}$. We also neglect the change of τ and $\dot{\tau}$ within one turn. This leads to

$$\langle V_r \rangle = \frac{2I_p^2}{I_0} \Big[Z_r(p\omega_0) - \frac{\dot{\tau}\omega_0}{2\omega_s} \left(2Q_s Z_r(p\omega_0) + (p+Q_s) Z_r((p+Q_s)\omega_0) - (p-Q_s) Z_r((p-Q_s)\omega_0)\right) \Big].$$

For the voltage induced in the reactive (imaginary) part of the impedance we have

$$\begin{aligned} V_{ki}(t) &= 2I_p \Big[-Z_i(p\omega_0)\sin(p\omega_0) \\ &+ \frac{\omega_0 \hat{\tau}}{2} \left((p+Q_s) Z_i((p+Q_s)\omega_0) \cos((p+Q_s)\omega_0 t) \right. \\ &\left. + (p-Q_s) Z_i((p-Q_s)\omega_0) \cos((p-Q_s)\omega_0 t) \right) \Big]. \end{aligned}$$

Splitting the trigonometric functions and using τ and $\dot{\tau}$ leads to

$$\begin{aligned} V_{ki}(t) &= 2I_p \Big[-Z_i(p\omega_0)\sin(p\omega_0) \\ &+ \frac{\omega_0 \hat{\tau}}{2} \left((p+Q_s) Z_i((p+Q_s)\omega_0) \left(\cos(p\omega_0 t)\tau + \sin(p\omega_0 t)\frac{\dot{\tau}}{\omega_s} \right) \right. \\ &\left. + (p-Q_s) Z_r((p-Q_s)\omega_0) \left(\cos(p\omega_0 t)\tau - \sin(p\omega_0 t)\frac{\dot{\tau}}{\omega_s} \right) \right) \Big]. \end{aligned}$$

For the average voltage per turn we get within linear approximation in τ and $\dot{\tau}$

$$\langle V_i \rangle = \frac{I_p^2 \omega_0 \tau}{I_0} \left[-2p Z_i(p\omega_0) + (p+Q_s) Z_i((p+Q_s)\omega_0) + (p-Q_s) Z_i((p-Q_s)\omega_0) \right].$$

We derived the voltage induced by an oscillating bunch in a narrow-band impedance. We assumed that the oscillation frequency is small compared to the revolution frequency $Q_s \ll 1$. We will find later when we treat the synchrotron oscillation that only the average voltage seen by the particles in the bunch is of importance for the coherent dipole oscillation. We found that this voltage has two components. The first one, V_r, is induced in the resistive impedance and is proportional to the derivative $\dot{\tau}$ of the bunch excursion. The second component, V_i, is induced in the reactive impedance and is proportional to the bunch excursion τ itself. We can collect the two parts and get the average voltage per turn seen by the particles in the bunch. To make the expressions more compact we introduce

$$Z_{pr}^{+} = Z_r(p\omega_0 + \omega_s) \ , \ Z_{pr}^{-} = Z_r(p\omega_0 - \omega_s) \ , \ Z_{pr}^{0} = Z_r(p\omega_0)$$

$$Z_{pi}^{+} = Z_i(p\omega_0 + \omega_s) \ , \ Z_{pi}^{-} = Z_i(p\omega_0 - \omega_s) \ , \ Z_{pi}^{0} = Z_i(p\omega_0)$$

and get

$$\langle V\rangle = \frac{2I_p^2}{I_0}\left[Z^0_{pr} - \frac{p\omega_0\dot{\tau}}{2\omega_s}\left(Z^+_{pr} - Z^-_{pr} - Q_s(2Z^0_{pr} + Z^+_{pr} + Z^-_{pr})\right)\right.$$
$$\left. -\frac{p\omega_0\tau}{2}\left(-2Z^0_{pr} + Z^+_{pi} + Z^-_{pi} + Q_s(Z^+_{pi} - Z^-_{pi})\right)\right].$$

Within the approximation $Q_s \ll 1$ this gives

$$\langle V\rangle = \frac{2I_p^2}{I_0}\left[Z^0_{pr} + \frac{p\omega_0\dot{\tau}}{2\omega_s}(Z^+_{pr} - Z^-_{pr}) + \frac{p\omega_0\tau}{2}(-2Z^0_{pr} + Z^+_{pi} + Z^-_{pi})\right]. \qquad (17)$$

The part of the voltage shown in the first line of the above equation does not depend on τ or $\dot{\tau}$ and is therefore present for a stationary bunch as we calculated before. Its resistive part leads to an energy loss and the reactive part to a change of the incoherent synchrotron frequency as we will see later.

We took here the case of a narrow-band impedance and in which only one sideband pair of the bunch spectrum leads to an induced voltage. This case is easier to understand in detail and we will use it also for the most simple case of a longitudinal instability. However, a generalization to a broadband impedance is straightforward. Many sidebands of the bunch spectrum induce a voltage in the impedance at the corresponding frequencies $\omega_0(p \pm Q_s)$. The total voltage is obtained by a sum over the parameter p. This leads to the expression

5 ROBINSON INSTABILITY

5.1 Qualitative treatment

The most important effect of the interaction between a longitudinally-oscillating bunch and a cavity is the so-called Robinson instability [1] which is treated here in some detail since it can be generalized to describe all multi-turn instabilities in storage rings. We start with a qualitative treatment by considering a single bunch circulating in a storage ring and exciting a cavity resonance with resonance frequency ω_r and impedance $Z(\omega)$ of which we consider only the resistive part Z_r.

The revolution frequency ω_0 of the circulating bunch depends on its energy deviation ΔE

$$\frac{\Delta\omega_0}{\omega_0} = -\eta_c\frac{\Delta E}{E} \text{ or } \omega_0 = p\omega_0\left(1 - \eta_c\frac{\Delta E}{E}\right).$$

While the bunch is executing a coherent dipole mode oscillation $\epsilon(t) = \hat{\epsilon}\cos(\omega_s t)$ its revolution frequency is modulated. *Above transition* the revolution frequency ω_0 is *small* when the *energy is high* and ω_0 is *large* when the *energy is small.* If the cavity is tuned to a resonant frequency slightly smaller than the RF frequency $\omega_r < p\omega_0$, Fig. 11 left, the bunch sees a higher impedance and *loses more energy* when it has an *energy excess* and it *loses less energy* when it has a *lack of energy.* This leads to a *damping* of the oscillation. If $\omega_r > p\omega_0$ this is reversed, Fig. 11 right, and leads to an *instability.* Below transition energy the dependence of the revolution frequency is reversed which changes the stability criterion.

5.2 Quantitative treatment

We consider a narrow-band cavity with a circulating bunch as before. The bunch executes a synchrotron oscillation which is approximately described as $\tau = \hat{\tau}\cos(\omega_s t)$ and produces sidebands to the revolution frequency harmonics of the bunch. The current of this oscillating bunch is obtained from Eq. (16) in the approximation of $Q_s \ll 1$

$$I_k(t) = I_0 + 2\sum_{p=1}^{\infty} I_p\,[\cos(p\omega_0 t)$$

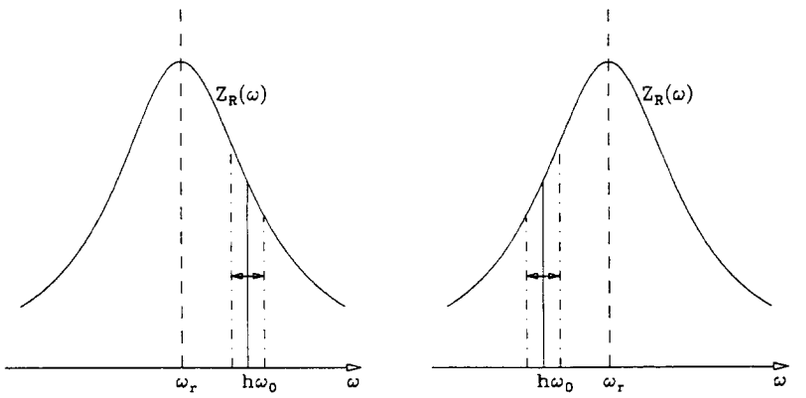


Figure 11: Qualitative treatment of the Robinson instability

$$+ \quad p\omega_0 \left(\sin(p\omega_0 t)\tau - \cos(p\omega_0 t)\frac{\dot{\tau}}{\omega_s}\right) + p\omega_0 \left(\sin(p\omega_0 t)\tau + \cos(p\omega_0 t)\frac{\dot{\tau}}{\omega_s}\right)\Big].$$

Since we assumed a narrow-band impedance we consider only one revolution harmonic p with its side bands and get for the averaged induced voltage Eq. (17) or the energy loss U of the particle in the bunch

$$U = e\langle V\rangle = e\frac{2I_p^2}{I_0}\left[Z_{pr}^0 + \frac{p\omega_0\dot{\tau}}{2\omega_s}(Z_{pr}^+ - Z_{pr}^-) + \frac{p\omega_0\tau}{2}(-2Z_{pr}^0 + Z_{pi}^+ + Z_{pi}^-)\right].$$

We include this induced voltage in the equation Eq. (1) for the energy gain and loss.

$$\dot{\epsilon} = \frac{\omega_0 e\hat{V}\sin\phi_s}{2\pi E} + \frac{\omega_0^2 he\hat{V}\cos\phi_s}{2\pi E}\tau - \frac{\omega_0 e}{2\pi E}\langle V\rangle.$$

Using the equilibrium condition

$$e\hat{V}\sin\phi_s = \frac{2I_p^2 Z_{pr}^0}{I_0}$$

we get

$$\dot{\epsilon} = \frac{\omega_0^2 he\hat{V}\cos\phi_s}{2\pi E} + \frac{I_p^2}{I_0}\frac{ep\omega_0^2}{2\pi E\omega_s}(Z_{pr}^+ - Z_{pr}^-)\dot{\tau} + \frac{I_p^2}{I_0}\frac{ep\omega_0^2}{2\pi E}(-2Z_{pi}^0 + Z_{pi}^+ + Z_{pi}^-)\tau.$$

Using the synchrotron frequency in the absence of an impedance

$$\omega_{s0}^2 = -\omega_0^2\frac{h\eta_c e\hat{V}\cos\phi_s}{2\pi E}$$

and the relation $\dot{\tau} = \eta_c\epsilon$ we get the second-order equation

$$\ddot{\tau} + \frac{pI_p^2\omega_{s0}}{2I_0 h\hat{V}\cos\phi_s}(Z_{pr}^+ - Z_{pr}^-)\dot{\tau} + \omega_{s0}^2\left(1 - \frac{pI_p^2}{2I_0 h\hat{V}\cos\phi_s}(-2Z_{pi}^0 + Z_{pi}^+ + Z_{pi}^-)\right)\tau = 0.$$

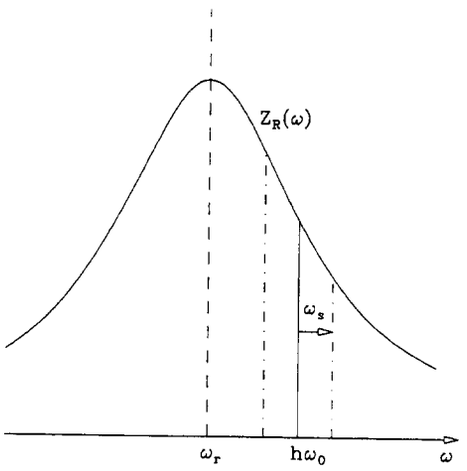


Figure 12: Quantitative treatment of the Robinson instability

This is the oscillation equation with the solution

$$\epsilon = \hat{\epsilon} e^{-\alpha_s t} \cos(\omega_s t + \phi)$$

and the damping or growth rate

$$\alpha_s = \frac{\omega_{s0} p I_p^2 (Z_{pr}^+ - Z_{pr}^-)}{2 I_0 h \hat{V} \cos\phi_s}.$$

The growth rate of the Robinson instability is given by the difference of the resistive impedance at the upper and lower synchrotron sideband, Fig. 12. Above transition energy we have $\cos\phi_s < 0$ and $\alpha_s > 0$, i.e. stability if $Z_{pr}^- > Z_{pr}^+$ as we found already from qualitative arguments.

The RF cavity itself has a narrow-band impedance around $h\omega_0$ which can drive an instability. Since the bunch length is usually much shorter than the RF wavelength we have $I_p = I_h \approx I_0$ so that

$$\alpha_s \approx \frac{\omega_{s0} I_0 (Z_{pr}^+ - Z_{pr}^-)}{2 I_0 \hat{V} \cos\phi_s}.$$

There is also a frequency shift due to the reactive part of the impedance

$$\omega_s^2 = \omega_{s0}^2 \left(1 + \frac{p I_p^2 Z_{pi}^0}{I_0 h \hat{V} \cos\phi_s} - \frac{p I_p^2 (Z_{pi}^+ + Z_{pi}^-)}{I_0 h \hat{V} \cos\phi_s}\right).$$

The second term in the parenthesis only depends on the impedance at the revolution harmonic $p\omega_0$ and not on the one at the sidebands. It is present also in the absence of a coherent motion and produces a change of the incoherent synchrotron frequency which we calculated before Eq. (10)

$$\omega_{si}^2 = \omega_{s0}^2 \left(1 + \frac{2 p I_p^2 Z_{pi}^0}{I_0 h \hat{V} \cos\phi_s}\right).$$

The coherent synchrotron motion produces a further shift compared to ω_{si}

$$\omega_s^2 = \left(\omega_{si}^2 - \frac{pI_p^2(Z_{pi}^+ + Z_{pi}^-)}{I_0 h\hat{V}\cos\phi_s}\omega_{s0}^2\right).$$

For a small effect, the shift of the coherent frequency with respect to the incoherent one is given by

$$\frac{\Delta\omega_s}{\omega_{s0}} \approx -\frac{pI_p^2(Z_{pi}^+ + Z_{pi}^-)}{2I_0 h\hat{V}\cos\phi_s}.$$

5.3 Discussion and generalization of the Robinson instability

We have derived the Robinson instability for the case of a single bunch and a single, narrow band and relatively weak resonance. We will here discuss the way this instability can be extended to cover more general cases.

A more general impedance will cover not just a single revolution harmonic with the two synchrotron oscillation sidebands but many such frequency lines. In this case the voltage induced in the impedance by each such line contained in the spectrum of the bunch current has to be considered. The growth rate will no longer be given by the difference between the impedance at the upper and lower synchrotron sideband but between the sums of the impedance times the spectral power taken at each upper and each lower synchrotron sideband contained in the spectrum of the oscillating bunch.

This can be extended to the case of many bunches [3, 4]. With M equidistant bunches in the machine we have M different modes of oscillation each having a different phase between the oscillations executed by adjacent bunches. The spectrum of each such coupled bunch mode has synchrotron sidebands at distinct revolution harmonics. In calculating the stability of a certain coupled-bunch mode we have to sum over these sidebands.

So far, we considered only dipole oscillations where the bunch makes a rigid oscillation around the nominal phase without changing the form. There are higher modes of oscillation, called bunch-shape oscillations, which can be classified as quadrupole ($m = 2$), sextupole ($m = 3$), octupole ($m = 4$), etc. modes. Each mode has a spectrum with side bands at $m\omega_s$ from the revolution harmonics. Again, to calculate the stability of these modes we have to sum over these sidebands.

We have assumed that the effect of the impedance is relatively weak such that the changes of the synchrotron frequency and the growth rate of the instability are small compared to the synchrotron frequency itself. For very narrow-band cavities with high shunt impedance, e.g. superconducting cavities, this might no longer be true. In this case we have to evaluate the impedance not at the unperturbed sideband ω_{s0} but at the shifted synchrotron frequency ω_s. Furthermore, if we are interested in the growth rate we have to consider the cavity impedance for a growing oscillation which is different as soon as the growth time of the oscillation becomes comparable to the filling time of the cavity. Taking all this into account one arrives at a 4th-order equation for the shifted synchrotron frequency and the growth rate for which a more general stability criterion can be derived, often called the second Robinson instability [1].

In the above paragraph we have considered stability only for the case of infinitesimally small oscillations and we have calculated their growth or damping time. If, however, the oscillation amplitude becomes large, some non-linear effects should be included. The modulation index of the phase oscillation will become large leading to sidebands at twice the synchrotron frequency. They have to be included when summing over the impedance contribution. This can lead to a situation where the beam is unstable for small oscillation amplitudes but becomes stable again at large amplitudes. In practice, such cases have bunches oscillating with finite but more or less constant amplitudes, [5, 6].

References

[1] K.W. Robinson, Stability of beam in radio frequency systems, Cambridge Electron Accel. CEAL-1010 (1964).

[2] A. Hofmann, Physics of beam instabilities, Proc. of a Topical Course 'Frontiers of Particle Beams' held by the Joint US-CERN School on Particle Accel. at South Padre Island, Texas, Oct. 1986, ed. M. Month and S. Turner, Lecture notes in Physics 296. Springer (1988) 99.

[3] F. Sacherer, A longitudinal stability criterion for bunched beams; Proc. of the 1977 Particle Accel. Conf., IEEE Trans. on Nucl. Sci. NS 20-3 (1973) 825.

[4] J.L. Laclare, Bunched beam coherent instabilities; CAS CERN Accelerator School, Advanced Accelerator Physics, 1985; ed. S. Turner. CERN 87-03, p. 264.

[5] F. Pedersen, CERN, Private communication.

[6] S. Krinsky, Saturation of a longitudinal instability due to nonlinearity of the wake field; Proc. of the 1985 Particle Accel. Conf., IEEE Trans. on Nucl. Sci. NS 32-5 (1985) 2320.

WAKE FIELDS AND IMPEDANCE

L. Palumbo [◊*)], *V.G. Vaccaro* [@§)] *and M. Zobov* [◊)]

◊) INFN-LNF Frascati
§) INFN, Sezione di Napoli
*) Università di Roma "La Sapienza"
@) Università di Napoli "Federico II"

Abstract
Knowledge of the electromagnetic interaction between a beam and the surrounding vacuum chamber is necessary in order to optimize the accelerator performance in terms of stored current. Many instability phenomena may occur in the machine because of the fields produced by the beam and acting back on itself as in a feedback device. Basically, these fields produce an extra voltage and energy gain, affecting the longitudinal dynamics, and a transverse momentum kick which deflects the beam. In this paper we describe the main features of this interaction with typical machine components.

1. INTRODUCTION

The so-called "collective effects" are responsible for many phenomena which limit the performance of an accelerator in terms of beam quality and stored current. The beam travelling inside a complicated vacuum chamber, induces electromagnetic fields which may affect the dynamics of the beam itself. An accelerator can be seen therefore as a feedback device, where any longitudinal or transverse perturbation appearing in the beam distribution may be amplified (or damped) by the e.m. forces generated by the perturbation itself.

The e.m. fields induced by the beam are referred to as *wake fields* due to the fact that they are left mainly behind the travelling charge. In the limit case of a charge moving at the light velocity, $\beta = 1$, the fields can only stay behind the charge because of the causality principle.

The study of the longitudinal and transverse beam dynamics requires the knowledge of the forces acting on the beam or, alternatively, the change in momentum caused by these e.m. forces. The *longitudinal wake potential* (*volts*) is the voltage gain of a unit trailing charge due to the fields created by a leading charge. The *transverse wake potential* (*volts*) is the transverse momentum kick experienced by the beam because of the deflecting fields. They are sometimes confused with the *wake functions*, defined as the wake potentials per unit charge (*volt /coulomb*) defining, therefore, a Green's function for the problem.

When we study the beam dynamics in the time domain, as is usually done for linear accelerators, it is convenient to make use of the wake functions or potentials. Conversely the frequency domain analysis is usually adopted for circular accelerators due to the intrinsic periodicity. There we need to compute the frequency Fourier transform of the wake function, which having Ohms units, is called *coupling impedance*.

In this paper we describe the main features of the electromagnetic fields induced in the most typical components installed on the beam pipe of an accelerator. In some examples we make use of numerical codes, reliable tools for the estimate of wake potentials and impedances, particularly useful in the design of the machine components. On this subject we address the readers to Ref. [1] where an exhaustive review on the available computer codes is presented. Methods and techniques for measurements of wake potentials and impedance are described in Ref. [2].

2. LONGITUDINAL WAKE FUNCTION AND LOSS FACTOR

2.1 Longitudinal wake function and loss factor of a point charge

Let us consider a charge q_1 travelling with constant velocity $v = \beta c$ on trajectories parallel to the axis of a vacuum chamber. Let z_1 be the longitudinal position and $\boldsymbol{r}_1$ the transverse vector positions (Fig. 1).

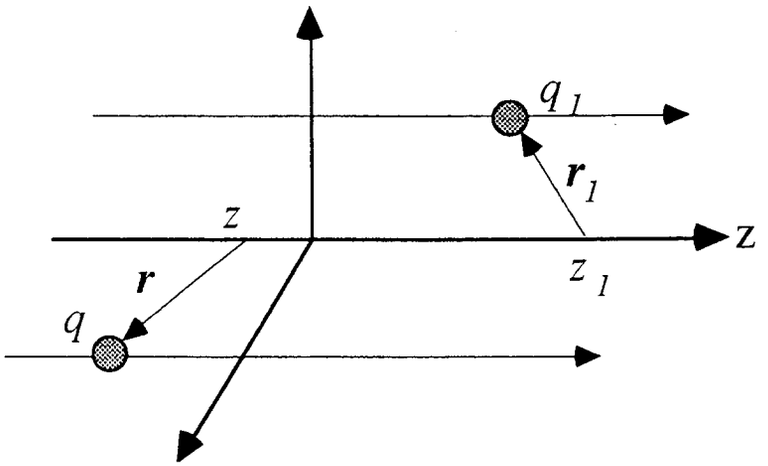


Fig. 1 Relevant coordinates system

The electromagnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ produced by the charge q_1 in the structure can be derived by solving the Maxwell equations satisfying proper boundary conditions. The Lorentz force acting on a charge q at a given position $\boldsymbol{r}$, z:

$$\boldsymbol{F}(\boldsymbol{r},z,\boldsymbol{r}_1,z_1\ ;\ t)=q\left[\boldsymbol{E}(\boldsymbol{r},z,\boldsymbol{r}_1,z_1\ ;\ t)+v\times\boldsymbol{B}(\boldsymbol{r},z,\boldsymbol{r}_1,z_1\ ;\ t)\right] \tag{1}$$

has in general field components along and perpendicular to the trajectory. These e.m. fields affect the dynamics of the charge itself and on any trailing charge as well. Calling τ the time delay of the trailing charge with respect to the leading one, at any instant "t" the leading and trailing charges have longitudinal coordinates $z_1(t) = vt$ and $z(t) = v(t - \tau)$ respectively.

The energy lost by the charge q_1 is computed as the work done by the longitudinal e.m. force along the structure:

$$U_{11}(\boldsymbol{r}_1)=-\int_{-\infty}^{\infty}\boldsymbol{F}(\boldsymbol{r}_1,z_1,\boldsymbol{r}_1,z_1\ ;\ t)\cdot dz \quad ; \quad t=\frac{z_1}{v} \tag{2}$$

The quantity U_{11} accounts for the energy loss in the resistive walls and in the diffracted fields radiated due to the discontinuities of the vacuum pipe. For a point charge, apart from particular cases, it is generally positive (energy loss).

The trailing charge also changes its energy under the effect of the fields produced by the leading one:

$$U_{21}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = -\int_{-\infty}^{\infty} \boldsymbol{F}(\boldsymbol{r},z,\boldsymbol{r}_1,z_1\,;\,t)\cdot dz \quad ; \quad t=\frac{z_1}{v}+\tau \tag{3}$$

where the force is calculated on the charge q, on the same path but with a time delay τ. The quantity U_{21} depending on the time delay τ can be positive (energy loss) or negative (energy gain). As long as we consider charges moving on trajectories parallel to the z-axis, the magnetic field cannot change the particle's energy, the product $\boldsymbol{v}\times\boldsymbol{B}\cdot dz = \boldsymbol{0}$ being identically zero. Accordingly, the energy gain of Eqs. (2) and (3) is computed considering the longitudinal component of the electric field only.

In the above definitions we have considered the integration over an infinite path. Of course infinite structures do not exist in practice, neither in linac nor in accelerator rings. In real machine components we may have fields confined in a limited region (for example resonant fields below the beam-pipe cut off), or propagating into the vacuum chamber. Extension of the integration path over an infinite pipe is certainly allowed in the former case. In the latter, definition (3) gives an estimate of the energy gain, which is a good approximation as long as the field wavelength is short compared to the device length.

A real vacuum chamber is formed by a smooth beam pipe with regular cross section (circular, rectangular or elliptic) and by various devices such as the RF cavity, the kickers, the diagnostic components etc. The exact solution of the Maxwell equation for the whole structure is impossible to obtain, even with the most sophisticated computer codes. Usually, one analyses a component at a time and sums-up the various effects. This procedure may lead to inexact estimates at high frequency where interference effects are not negligible.

It has to be underlined that in Eqs. (2) and (3) we assume the charge velocity unchanged during the motion. One can imagine that an external force keeps constant the charge velocity doing the work computed in Eqs. (2,3). In absence of the external force, this work corresponds to the energy loss (or gain) of the charges, provided that the velocity of the charge does not change significantly. In practice Eqs. (2,3) may be used when the relative change of energy is very small, so that it does not produce an appreciable variation of the relativistic factor β. This is the case, for instance, of ultra-relativistic charges. Otherwise, one has to introduce the equations of the dynamics combined to Maxwell equations.

We define *loss factor* k the energy lost by q_1 per unit charge squared:

$$k(\boldsymbol{r}_1) = \frac{U_{11}(\boldsymbol{r}_1)}{q_1^2} \qquad [\mathrm{V/C}] \tag{4}$$

and *longitudinal wake function* $w_z(\boldsymbol{r}_1,\boldsymbol{r}_2;\tau)$ the energy lost by the trailing charge q per unit of both charges q_1 and q [3,4,5]:

$$w_z(\boldsymbol{r},\boldsymbol{r}_1;\tau) = \frac{U_{21}(\boldsymbol{r},\boldsymbol{r}_1;\tau)}{q_1 q} \qquad [\mathrm{V/C}] \tag{5}$$

The explicit dependence on β although omitted, should be borne in mind. We note that both the wake function and the loss factor have the same units *volt/coulomb*. Sometimes in the literature one finds that the quantity $w(\tau)$ is improperly called wake potential; the wake function is numerically equal to the potential seen by the charge only when one considers unity charges.

In some cases, such as an infinite beam pipe with perfectly conducting or resistive walls, the e.m. force is constant along the integration path; it is therefore useful to introduce the wake function per unit length, *volt/(coulomb meter)*, given by:

$$\frac{dw_z(\boldsymbol{r},\boldsymbol{r}_1,\tau)}{dz} = -\frac{1}{q_1 q}\boldsymbol{F}_z(\boldsymbol{r},z,\boldsymbol{r}_1,z_1\ ;\ t) \quad ; \quad z = z_1 - v\tau \qquad \text{[V/Cm]} \tag{6}$$

which, apart from the sign, is in practice the longitudinal force per unit charge acting on q. In some other cases where we deal with periodic structures, we rather calculate a wake force per unit period length.

It is worth noting that in the most cases of interest we deal with structures having particular symmetric shapes: rectangular, elliptic, circular. Moreover, it is generally verified that during the machine operation the beam can only be slightly displaced from the axis. Accordingly the above quantities can be expanded around the axis keeping only few relevant terms. This multipolar expansion assumes a particular form in case of cylindrical symmetry and ultra relativistic charges, as will be shown in Sec. 2.6. The dominant term produced by a charge on the axis is called *monopole wake* (Fig. 2).

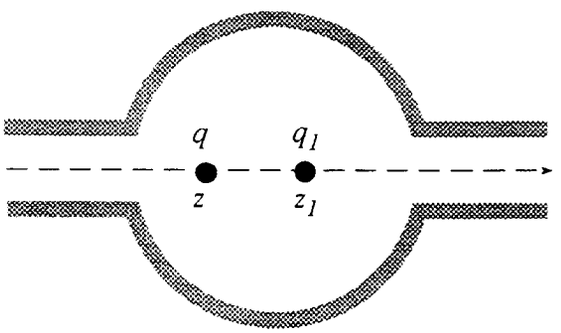


Fig. 2 Leading and trailing charges on the axis of a cavity with cylindrical symmetry

2.2 Beam loading theorem for a point charge

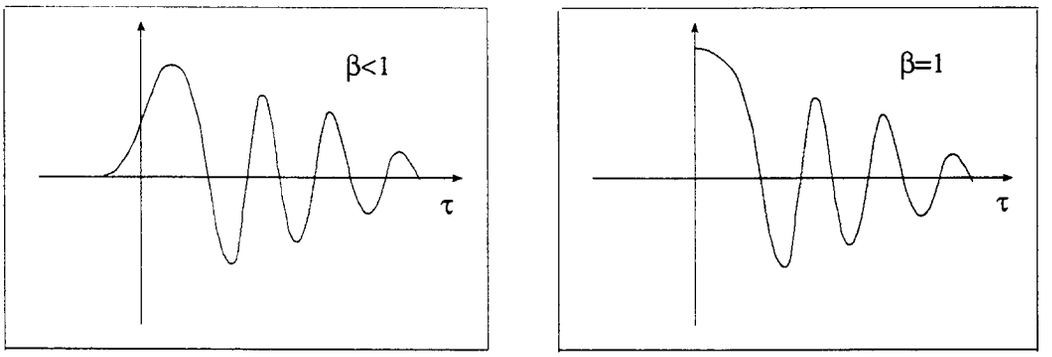
From the above definitions we easily derive that, when the charges travel on the same trajectory, the loss factor is given by the wake function in the limit of zero distance between q_1 and q. Omitting the radial dependence, one obtains : $k = w_z(0)$. This is generally true as long as $\beta < 1$, however, in the relevant case $\beta = 1$ it has been proved that [3]:

$$k = \frac{w_z(\tau \to 0^+)}{2} \tag{7}$$

This property, referred to as the fundamental theorem of the beam loading [3], is a consequence of the causality principle. In fact, due to the finite propagation velocity of the induced fields and to the motion of the source charge, the wake function is not symmetric with respect to the leading charge (Fig. 3a). In the limit case of a charge with light velocity it exists only in the region $\tau > 0$ (Fig. 3b), showing a discontinuity at the origin.

To prove the theorem, let us consider the wake function produced by a point charge as the sum of an even and odd function of τ (Fig. 4):

$$w_z(\tau) = w_z^e(\tau) + w_z^o(\tau) \tag{8}$$



(a) (b)

Fig. 3 Example of wake functions for a) $\beta < 1$, and b) $\beta = 1$

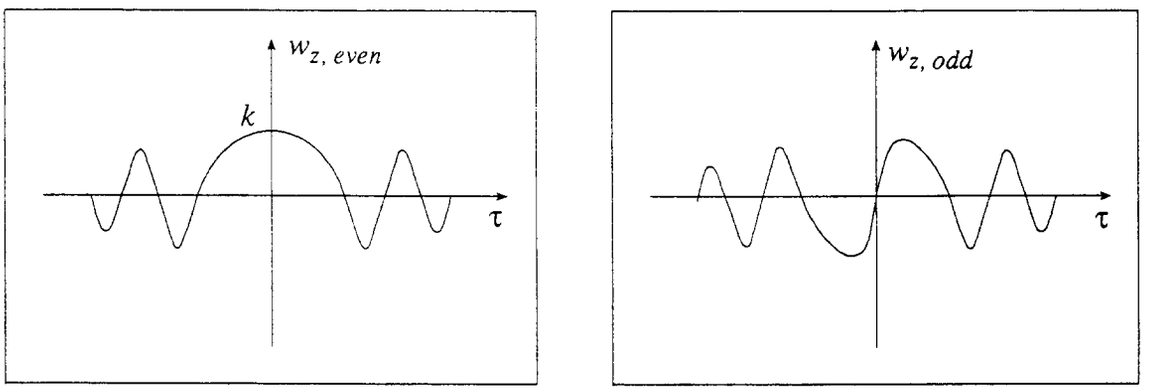


Fig. 4 Even and odd part of the wake of Fig. 3a ($\beta < 1$)

It is apparent that only $w_z^e(\tau)$ may change the energy of the point charge, $w_z^o(\tau)$ being zero at $\tau = 0$. Therefore we can say that the loss factor of a point charge is given by:

$$k = w_z^e(\tau = 0) \tag{9}$$

For $\beta = 1$, we have that $w_z(\tau) = 0$ for $\tau < 0$, because of the causality principle. In this region the wake vanishes if:

$$w_z^e(\tau) = -w_z^o(\tau) \tag{10}$$

However we have, for $\tau > 0$

$$w_z^e(\tau) = w_z^o(\tau) \tag{11}$$

$$w_z(\tau) = 2w_z^e(\tau) = 2w_z^o(\tau) \tag{12}$$

Therefore from Eq. (9) we obtain:

$$k = w_z^e(\tau \to 0) = \frac{w_z(\tau \to 0^+)}{2} \tag{13}$$

We call the reader's attention to the fact that in general, as long as $\beta < 1$, i.e. in all the realistic cases, the wake is a continuous function of τ. Therefore, it is more than reasonable to wonder about the meaning of Eq. (13) that applies only in the unrealistic case $\beta = 1$. It is easy to see that although the wake is a continuous function for any realistic value of β, its shape approaches more and more the discontinuous curve of Fig. 3 when $\beta \to 1$. In other words, one could not, in principle, exchange the limits $\tau \to 0$ and $\beta \to 1$.

2.2.1 Example: Point-charge wake for a single resonating mode HOM.

As will be shown in Sec. 7.4, a point charge q_1 passing through a resonant cavity excites all the resonating modes. In the limit case $\beta = 1$, each mode is schematized by the electric RLC parallel circuit driven by a point-charge current $i_b(\tau) = q_1 \delta(\tau)$:

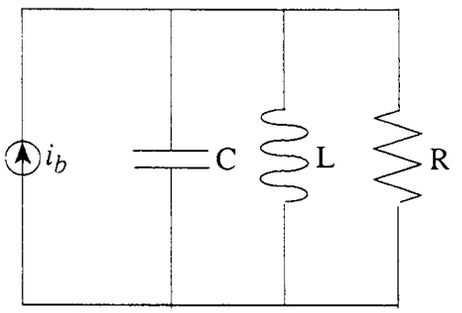


Fig. 5 Scheme of a RLC parallel circuit driven by the current $i_b(\tau)$

At the time $\tau = 0^+$ we observe that the capacitor is charged with a voltage:

$$V(0^+) = \frac{q_1}{C} \equiv V_o \text{ and } \dot{V}(0^+) = \frac{V(0^+)}{RC} \tag{14}$$

For $t > 0$ the system will undergo free oscillations. In particular the voltage $V(\tau)$ will be a solution of differential equation of the circuit:

$$\ddot{V}(\tau) + 2\Gamma\dot{V}(\tau) + \omega_r^2 V(\tau) = 0 \tag{15}$$

where:

$$2\Gamma = \frac{1}{RC}, \quad \text{and} \quad \omega_r^2 = \frac{1}{LC} \tag{16}$$

Solving Eq. (15) with the initial conditions (14) and according to the definition (5), one gets:

$$w_z(\tau) = \frac{V(\tau)}{q_1} = \frac{e^{-\Gamma\tau}}{C}\left[\cos(\overline{\omega}_r\tau) - \frac{\Gamma}{\overline{\omega}_r}\sin(\overline{\omega}_r\tau)\right]H(\tau) \tag{17}$$

where $H(\tau)$ is the Heaviside function and

$$\overline{\omega}_r^2 = \omega_r^2 - \Gamma^2 \tag{18}$$

Using the merit factor of the circuit defined by:

$$Q = \frac{\omega_r}{2\Gamma} \tag{19}$$

$$w_o = \frac{1}{C} = \frac{R\omega_r}{Q}$$

$$\frac{\Gamma}{\overline{\omega}_r} = \frac{1}{\sqrt{4Q^2 - 1}}$$

$$\overline{\omega}_r = \omega_r \sqrt{1 - \frac{1}{4Q^2}} \tag{20}$$

A qualitative behaviour of the wake function is shown in Fig. 6.

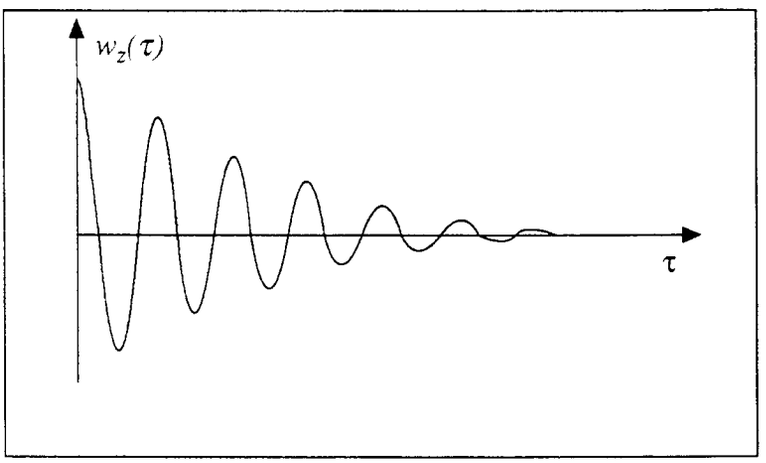


Fig. 6 Wake function of a resonant mode

The loss factor, according to the definition (4), can be computed as the energy lost by the unit charge after its passage through the cavity. Applying the energy conservation law, we can obtain the energy lost by the charge q_1 as the e.m. energy initially stored in the capacitor. We get:

$$k = \frac{1}{2C} = \frac{\mathrm{w}(\tau \to 0)}{2} \tag{21}$$

which satisfies the beam loading theorem.

In terms of the merit factor Q we get:

$$k = \frac{R\omega_r}{2Q} \tag{22}$$

2.3 Longitudinal wake function and loss factor of a bunch

The wake function defined in Eq. (5), being generated by a point charge, is a Green function and allows one to compute the wake produced by any bunch distribution. Let us consider now a bunch of particles moving on a trajectory parallel to the axis, at a distance r_1, with

a longitudinal time distribution function $i_b(\tau)$ such that:

$$q_1 = \int_{-\infty}^{+\infty} i_b(\tau)\, d\tau \tag{23}$$

The wake function produced by the bunch distribution at a point with time delay τ is simply given by the convolution of the Green function over the bunch distribution. We remember that, in practice, the convolution integral is obtained by applying the superposition principle. We split the distribution into an infinite number of infinitesimal slices and sum up their wake contributions at the point τ. According to the definitions given so far, the energy lost by a trailing charge q because of the wake produced by the slice at τ' is:

$$dU(\boldsymbol{r}, \tau - \tau') = q\, i_b(\tau') w_z(\boldsymbol{r}, \tau - \tau') d\tau' \tag{24}$$

Summing up all the effects we get the wake function of a bunch distribution as:

$$W_z(\boldsymbol{r}, \tau) = \frac{U(\boldsymbol{r}, \tau)}{q_1 q} = \frac{1}{q_1} \int_{-\infty}^{\infty} i_b(\tau') w_z(\boldsymbol{r}, \tau - \tau') d\tau' \tag{25}$$

For a bunch travelling with velocity "c", because of the causality, the above folding integral has the observation point "τ" as uppermost limit.

Once the bunch wake function $W_z(\boldsymbol{r}, \tau)$ is known, it is straightforward to derive the loss factor of the charge distribution by applying again the superposition principle, we get:

$$K(\boldsymbol{r}) = \frac{U(\boldsymbol{r})}{q_1^2} = \frac{1}{q_1} \int_{-\infty}^{\infty} W_z(\boldsymbol{r}, \tau)\, i_b(\tau)\, d\tau \tag{26}$$

which depends on the transverse displacement of the bunch.

2.3.1 Example: rectangular bunch distribution exciting a single HOM.

Let us consider a bunch distribution with a simple rectangular shape on the axis at $r = 0$.

$$i_b(\tau) = \frac{q_1}{2T}\left[H(\tau + T) - H(\tau - T)\right] \tag{27}$$

and compute the wake function of such a charge distribution assuming that it excites a single HOM in a RF cavity. Further, let us assume that the factor Q is so high that, in the range of interest, the impulsive wake function can be approximated by:

$$w_z(\tau) = w_o \cos(\omega_r \tau) H(\tau) \tag{28}$$

By using the folding integral (25) we get two expressions of the bunch wake for τ inside and outside the distribution. Inside the charge distribution, i.e. for $-T < \tau < T$, we get:

$$W_z(\tau) = \frac{w_o}{2} \frac{\sin[\omega_r(\tau+T)]}{\omega_r T} H(\tau+T) \tag{29}$$

It is worth noting that in the limit $T \to 0$, the rectangular distribution becomes an impulsive function $i_b(\tau) = q_1 \delta(\tau)$ and the bunch wake $W_z(\tau) \to w_z(\tau)$. In particular it is interesting to see that:

$$\lim_{T \to 0} W_z(0) = \frac{w_o}{2} \tag{30}$$

i.e. looking at the centre of the bunch, one finds that the wake function approaches with continuity the limit value (7).

The bunch loss factor is obtained from Eq. (26) which gives:

$$K = \frac{w_o}{2} \left(\frac{\sin(\omega_r T)}{\omega_r T} \right)^2 \tag{31}$$

The "point charge" loss factor is derived from the above expression in the limit $T \to 0$:

$$k = \lim_{T \to 0} K = \frac{w_0}{2} \tag{32}$$

Therefore, when we consider any bunch distribution, the somewhat "artificial" arguments presented in Sec. 2.2 are unnecessary, since the loss factor can be computed straightforwardly from the bunch wake which turns out to be continuous, even in the "point charge" limit.

Finally we find externally to the distribution, i.e. for $\tau \geq T$:

$$W_z(\tau) = w_o \frac{\sin(\omega_r T)\cos(\omega_r \tau)}{\omega_r T} H(\tau - T) \tag{33}$$

It is interesting to note that outside the distribution, the limit for $T \to 0$ and $\tau \to 0$ of Eq. (33) gives w_o.

2.4 Loss factor and Poynting Vector

The bunch wake function has been defined as the energy loss by a bunch crossing a given structure. We already said that the non-consistency due to the constant velocity of the bunch can be avoided assuming an external force acting on the bunch (for instance related to an electric external potential). Since the kinetic energy of the bunch is constant (constant velocity) the work done by the external force has to be equal to the energy loss, according to the energy conservation law. However, it is well known that, any electromagnetic energy loss can be computed as the flux of the Poynting vector over a closed surface surrounding the sources of the fields.

The Poynting theorem states that the electromagnetic energy U_{em} stored in a volume V limited by the surface S can change because of ohmic losses and electromagnetic radiation:

$$\frac{\partial U_{em}}{\partial t} = -\int_S \boldsymbol{P}\cdot\hat{n}\,dS + \int_V (\boldsymbol{E}\cdot\boldsymbol{J})\,dV \tag{34}$$

where $\hat{n}$ is the unity normal to the surface S, $\boldsymbol{J}$ is the current density, $\boldsymbol{E}$ the electric field and $\boldsymbol{P}$ is the Poynting vector defined as:

$$\boldsymbol{P} = \frac{1}{\mu}\,\boldsymbol{E}\times\boldsymbol{B} \tag{35}$$

Let us consider now a single charge moving on the axis of a given structure. The current density is given by:

$$\boldsymbol{J}(\boldsymbol{r},z,t) = q_1 v \frac{\delta(r)}{2\pi r}\,\delta(z - vt) \tag{36}$$

We choose as surface S a cylinder of infinitesimal radius around the charge trajectory. Integrating Eq. (34) with respect to the time from $-\infty$ to $+\infty$, and noting that in the volume V $U_{em}(t=-\infty) = U_{em}(t=\infty)$, we get:

$$\int_{-\infty}^{\infty} dt \int_V (\boldsymbol{E}\cdot\boldsymbol{J})\,dV = \int_{-\infty}^{\infty} dt \int_S \boldsymbol{P}\cdot\hat{n}\,dS \tag{37}$$

Making use of (2),(4) and (36) we get for the loss factor:

$$k = \frac{-1}{q_1}\int_{-\infty}^{\infty} E_z\left(z,t=\frac{z}{v}\right)dz = \frac{-1}{q_1}\int_{-\infty}^{\infty} dt \int_S \boldsymbol{P}\cdot\hat{n}\,dS \tag{38}$$

2.5 The synchronous fields

When a bunch crosses the various elements installed in the beam pipe, it excites secondary fields because of induction effects and diffraction phenomena. Some of these fields are localized around the bunch, as for example the space-charge or the resistive-wall fields, others are localized in resonant structures like the RF cavity, and others, at high frequency can propagate within the beam pipe. All these fields interact with the circulating beam.

We want to show that this interaction is such that only the field components synchronous with the charges can change the charges' energy. In order to prove this statement it is convenient to express the longitudinal electric field in terms of waves propagating in the z-direction:

$$E_z(z,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\kappa\, \tilde{E}_z(\omega,\,\kappa) e^{j(\omega t - \kappa z)} \tag{39}$$

where we have omitted the explicit dependence on $(\boldsymbol{r},\boldsymbol{r}_1,z_1)$.

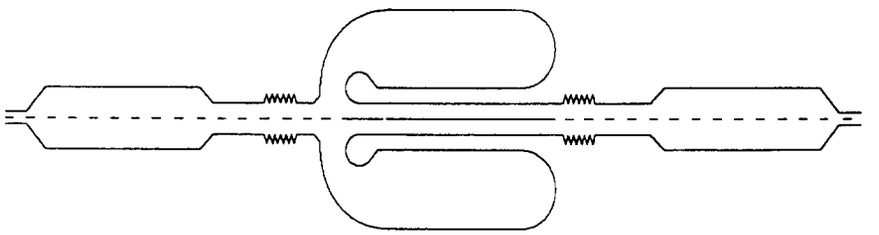


Fig. 7 DAΦNE Accumulator vacuum chamber (RF cavity, kicker tanks, bellows)

The longitudinal electric field is given by a superposition of longitudinal waves having any phase velocity, since ω and κ can vary from $-\infty$ to ∞. Among these waves only those having the same phase velocity of the charge can contribute to the energy gain and therefore to the wake function. In fact let us put the field expression (39) into the wake function definition (5), we get:

$$w_z(\tau)=\frac{-1}{q_1}\int_{-\infty}^{\infty}E_z\left(z,t=\frac{z}{v}+\tau\right)dz=\frac{-1}{2\pi q_1}\int_{-\infty}^{\infty}d\omega\, e^{j\omega\tau}\int_{-\infty}^{\infty}d\kappa\,\tilde{E}_z(\omega,\ \kappa)\int_{-\infty}^{\infty}e^{-jz(\kappa-\kappa_o)}dz \qquad (40)$$

with $\kappa_o=\frac{\omega}{v}$. We recognize the impulsive function:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-jz(\kappa-\kappa_o)}dz=\delta(\kappa-\kappa_o) \qquad (41)$$

that allows us to get the following simple expression:

$$w_z(\tau)=\frac{-1}{q_1}\int_{-\infty}^{\infty}\tilde{E}_z(\kappa=\kappa_o,\omega)\,e^{j\omega\tau}d\omega \qquad (42)$$

wherein it is apparent that only those components of the fields propagating with the same phase velocity as the charge can produce a "surfing" effect. All the others in average do not contribute.

The result found above deserves a further investigation. In fact we wonder what would happen if, instead of an infinite structure one would consider a pipe with finite length, say L. It is easy to see that integration between -L/2 and +L/2 does not give an impulsive function, but:

$$\frac{1}{2\pi}\int_{-L/2}^{L/2}e^{-jz(\kappa-\kappa_o)}dz=\frac{L}{2\pi}\frac{\sin\left[(\kappa-\kappa_o)\frac{L}{2}\right]}{(\kappa-\kappa_o)\frac{L}{2}} \qquad (43)$$

which becomes again an impulsive function when $L\to\infty$. For a finite length L, the "sinc" function (sin(x)/x) has a maximum at $\kappa_o=\kappa$, and the first zero at $\kappa_o=\kappa\pm 2\pi/L$.

For long wavelength, the fields do not propagate being stored within a given device (e.g. the cavity HOMs). The actual integration path is therefore confined to a limited region, the fields being evanescently zero above the pipe cut-off. On the other hand, for short wavelengths, fields propagate into the beam pipe. There is a contribution of those harmonics that do not perfectly average to zero their effect on the beam. However, according to Eq. (43) at high frequencies this contribution is small, so that we can consider an infinite pipe instead of a finite one, simplifying the calculation of the wake.

2.6 Expansion of the longitudinal wake in cylindrical symmetry

So far we have considered the case of general boundaries, assuming the two charges moving on any trajectory parallel to the axis. We have already mentioned that in general there is no restriction on the transverse position of both charges. For simplicity we now consider that the trajectories are parallel to the axis of a structure with cylindrical symmetry as shown in Fig. 2. Let $(r_1, \phi_1 = 0, z_1)$ be the coordinates of the leading charge and (r, ϕ, z) those of the trailing one. The density charge q_1 can be represented as a superposition of multipole moments in cylindrical coordinates:

$$\rho_1 = q_1 \frac{\delta(r - r_1)}{r_1} \delta(\phi) \delta(z - z_1) \tag{44}$$

with $z_1 = \beta c \tau$. Exploiting the azimuthal periodicity we can write:

$$\rho_1 = \frac{q_1}{2\pi} \frac{\delta(r - r_1)}{r_1} \delta(z - z_1) \sum_{m=0}^{\infty} \alpha_m \cos(m\phi) \tag{45}$$

with:

$$\alpha_m = \begin{cases} 1, & m = 0 \\ 2, & m \neq 0 \end{cases} \tag{46}$$

According to the above expression the charge can be thought of as a superposition of charged rings with angular dependence $\cos(m\phi)$. It easy to see for instance that the monopolar term with $m = 0$ describes a charged ring of radius r_1 with uniform density. In cylindrical coordinates the e.m. fields created by the distribution (45) can be derived as the sum of multipole terms as well, showing therefore the same angular dependence. For each term we can compute the effect of the longitudinal force. The resulting wake function will show the following form:

$$\begin{aligned} w_z(\boldsymbol{r}, \boldsymbol{r}_1; \tau) &= \sum_{m=0}^{\infty} w_{z,m}(\boldsymbol{r}, \boldsymbol{r}_1; \tau) \\ w_{z,m}(\boldsymbol{r}, \boldsymbol{r}_1; \tau) &= \overline{w}_{z,m}(r, r_1; \tau) \cos(m\phi) \end{aligned} \tag{47}$$

2.7 Radial expansion of the wake function in the limit $\gamma \to \infty$

The e.m. fields produced by the travelling charge in a vacuum chamber are derived from the Maxwell equation imposing the boundary conditions at the pipe walls:

$$\nabla \times \boldsymbol{B} = \mu_o \boldsymbol{J} + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} \tag{48}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{49}$$

$$\nabla \cdot \boldsymbol{E} = -\frac{\rho}{\varepsilon_o} \tag{50}$$

$$\boldsymbol{J} = \rho \boldsymbol{v} \tag{51}$$

The longitudinal electric field can be thought of as produced by the current sources (the bunch) and by the currents induced at the walls. Considering only the induced fields, it can be shown that the Fourier component synchronous with the charges is a solution of the following equation [6,7]:

$$\nabla_\perp^2 \tilde{E}_z - \left(\frac{\omega}{\beta c \gamma}\right)^2 \tilde{E}_z = 0 \tag{52}$$

In the limit $\gamma \to \infty$, i.e. in the case of ultra relativistic charges, we have:

$$\nabla_\perp^2 \tilde{E}_z = 0 \tag{53}$$

Solved in cylindrical symmetry, the above equation gives the following radial dependence of the wake function [4,6,7]:

$$\overline{w}_{z,m}(r, r_1; \tau) = r^m r_1^m \, \overline{\overline{w}}_{z,m}(\tau) \tag{54}$$

The monopole term $m = 0$, does not depend on the radial position of the charges. This result, applied only to ultra-relativistic charges, allows one to simplify the evaluation of the wake function by choosing a suitable integration path. Numerical codes [1,25,57,58] computing the longitudinal monopole wake function of charges with $\beta = 1$, in structure with cylindrical symmetry, perform the integration along trajectories at the radius of the beam pipe. Since the longitudinal electric field vanishes on the pipe surface, the integration is limited to a shorter path.

We want to underline that the expansion (54) concerns only the secondary fields induced by the beam. The primary fields produce the so-called space-charge wake effects that show a different radial dependence, (Sec. 6.3).

2.8 Wake function in accelerator rings

In the case of circular machines, the longitudinal position of the charge is given by the coordinate θ. We compute the wake function by averaging the azimuthal electric field over a revolution period T_o:

$$w(\tau) = -2\pi R \left\langle E_\theta(\theta, t + \tau) \right\rangle_{T_o} \tag{55}$$

Due to the intrinsic periodicity of the e.m. problem, we can expand the longitudinal electric field of a single charge as:

$$E_\theta(\theta,t)=\int_{-\infty}^{\infty} d\omega\, e^{j\omega t}\sum_{n=-\infty}^{\infty}\tilde{E}_z(n,\omega)e^{-jn\theta} \tag{56}$$

which substituted in (55) gives:

$$w(\tau)=-R\int_{-\pi}^{\pi} d\theta\int_{-\infty}^{\infty} d\omega\, e^{j\omega\tau}\sum_{n=-\infty}^{\infty}\tilde{E}_\theta(n,\omega)e^{-j\theta(n-\omega/\omega_o)} \tag{57}$$

The charge itself can be thought of as a train of charges with a beam current:

$$i_b(\tau)=q_1\sum_{k=-\infty}^{\infty}\delta(\tau-kT_o)$$

Making use of (25) and (55), we have:

$$W(\tau)=\frac{1}{q_1}\int_{-\infty}^{\infty} i_b(\tau')w(\tau-\tau')d\tau'=\sum_{k=-\infty}^{\infty} w\left(\tau-\frac{2\pi k}{\omega_o}\right) \tag{58}$$

$$W(\tau)=-R\int_{-\pi}^{\pi} d\theta\int_{-\infty}^{\infty} d\omega\sum_{k=-\infty}^{\infty}e^{j\omega(\tau-2\pi k/\omega_o)}\sum_{n=-\infty}^{\infty}\tilde{E}_\theta(n,\omega)e^{-j\theta(n-\omega/\omega_o)}$$

which, after some mathematics, becomes:

$$W(\tau)=-\frac{2\pi R}{q_1}\sum_{n=-\infty}^{\infty}\tilde{E}_\theta(n,n\omega_o)e^{jn\omega_o\tau} \tag{59}$$

3. LONGITUDINAL COUPLING IMPEDANCE

3.1 Definitions and properties

In the frequency domain we compute the spectrum of the point charge wake function as:

$$\int_{-\infty}^{\infty} w_z(\mathbf{r},\mathbf{r}_1;\tau)\,e^{-j\omega\tau}d\tau\equiv Z(\mathbf{r},\mathbf{r}_1;\omega) \tag{60}$$

which being measured in *Ohms* is called *Coupling Impedance*. Historically, the coupling impedance concept was introduced in the early studies of the instabilities arising in the ISR at CERN [8].

The wake function is derived from the impedance by inverting the Fourier integral:

$$w_z(\boldsymbol{r},\boldsymbol{r}_1;\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}Z(\boldsymbol{r},\boldsymbol{r}_1;\omega)\,e^{j\omega\tau}d\omega \tag{61}$$

In the following we shall omit, for simplicity, the radial dependence. Comparison with Eq. (42) shows that:

$$Z(\omega)=-\frac{2\pi}{q_1}\tilde{E}_z\left(\kappa=\kappa_o,\omega\right) \tag{62}$$

The coupling impedance is a complex quantity:

$$Z(\omega)=Z_r(\omega)+jZ_i(\omega) \tag{63}$$

with $Z_r(\omega)$ and $Z_i(\omega)$ even and odd function of ω respectively. It is easy to prove this property of the impedance remembering that the wake potential $w(\tau)$ is a real function of τ. In fact expanding the exponential in the integral of Eq. (61) we have:

$$\begin{aligned} w_z(\tau)=&\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\,Z_r(\omega)\cos(\omega\tau)-Z_i(\omega)\sin(\omega\tau)\right]d\omega \\ &+\frac{j}{2\pi}\int_{-\infty}^{\infty}\left[\,Z_r(\omega)\sin(\omega\tau)+Z_i(\omega)\cos(\omega\tau)\right]d\omega \end{aligned} \tag{64}$$

where the imaginary part vanishes if:

$$\begin{aligned} Z_r(\omega)&=Z_r(-\omega) \\ Z_i(\omega)&=-Z_i(-\omega) \end{aligned} \tag{65}$$

From Eqs. (8,60,64), we recognize that $Z_r(\omega)$ and $-Z_i(\omega)$ are the Fourier transform of $w_z^e(\tau)$ and $w_z^o(\tau)$ respectively:

$$\begin{aligned} w_z^e(\tau)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}Z_r(\omega)\cos(\omega\tau)d\omega \\ w_z^o(\tau)&=\frac{-1}{2\pi}\int_{-\infty}^{\infty}Z_i(\omega)\sin(\omega\tau)d\omega \end{aligned} \tag{66}$$

Furthermore, in the particular case of $\beta=1$, the wake function has to vanish for $\tau<0$ where $w_z^e(\tau)=-w_z^o(\tau)$. In terms of impedances Eq. (10) becomes:

$$\int_{-\infty}^{\infty}Z_r(\omega)\cos(\omega\tau)d\omega=\int_{-\infty}^{\infty}Z_i(\omega)\sin(\omega\tau)d\omega \tag{67}$$

which expresses a general relationship between the real and imaginary part of the impedance. It can be shown that the above relation is equivalent to the Hilbert transform relating the real and imaginary part of a network impedance. In other words the coupling impedance defined by

Eq. (60) behaves like a usual circuit impedance only when the causality principle applies, namely in the limit case of charges travelling with the velocity of light.

Recalling the relation (7) between loss factor and wake potential for a point charge, we get:

$$k = \frac{w_z(\tau \to 0^+)}{2} = \frac{1}{\pi}\int_0^\infty Z_r(\omega)d\omega \tag{68}$$

where we recognize that the real part of the impedance is the power spectrum of the energy loss of a unit point charge. In general, the complex impedance can be thought of as the complex power spectrum related to the energy loss.

3.1.1 Example: Impedance of a single HOM in a RF cavity.

Using the wake function expression (17) derived for a single HOM, from the definition (60) we have:

$$Z(\omega) = \frac{\omega_r R}{Q}\int_{-\infty}^{\infty}\left[\cos(\overline{\omega}_r\tau) - \frac{\Gamma}{\overline{\omega}_r}\sin(\overline{\omega}_r\tau)\right]e^{-(j\omega+\Gamma)\tau}d\tau \tag{69}$$

$$Z(\omega) = \frac{R}{1 + jQ\left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega}\right)} \tag{70}$$

It is easy to verify that the above impedance satisfies the properties (65) and (67).

3.2 Bunch losses and wake function from the impedance

Consider a bunch radially displaced and with a charge distribution $i_b(\tau)$, whose Fourier spectrum is $I(\omega)$. The total bunch wake function $W_z(r;\tau)$ and loss factor can be expressed in terms of $Z(\omega)$ by transforming the integrals (25) and (26), obtaining:

$$W_z(r;\tau) = \frac{1}{2\pi q_1}\int_{-\infty}^{\infty} Z(r;\omega)I(\omega)\ e^{j\omega\tau}d\omega \tag{71}$$

$$K(r) = \frac{1}{\pi q_1^2}\int_0^\infty Z_r(r;\omega)|I(\omega)|^2 d\omega \tag{72}$$

As an example for a bunch with Gaussian distribution:

$$I(\omega) = q_1\, e^{-\frac{(w\sigma_t)^2}{2}} \tag{73}$$

the loss factor is given by:

$$K(r) = \frac{1}{\pi}\int_0^\infty Z_r(r;\omega)\, e^{-(\omega\sigma_\tau)^2} d\omega \tag{74}$$

It is apparent that the loss factor is, in general, a function of the r.m.s. length of the bunch distribution. It is interesting to note that there exists a general relation, useful in the measurements, between the frequency dependence of the impedance and the dependence of the loss factor on the bunch length. For a Gaussian bunch:

$$Z_r(\omega) \propto \omega^a \Leftrightarrow K \propto \sigma_\tau^{-(a+1)} \tag{75}$$

3.3 Multipole longitudinal impedance for cylindrical symmetry

In Sec. 2.6 we have seen that the wake function created by a charge on a trajectory parallel to the axis of a device with cylindrical symmetry, can be expanded into a sum of multipolar terms. The wake expansion used in the impedance definition allows one to express also the impedance as a multipole expansion:

$$Z(r,r_1,\phi;\omega) = \sum_{m=0}^{\infty} Z_m(r,r_1,\phi;\omega) = \sum_{m=0}^{\infty} \bar{Z}_m(r,r_1;\omega)\cos(m\phi) \tag{76}$$

For ultra relativistic charges the radial dependence of the wake function is known, according to Eq. (54), we have:

$$\bar{Z}_m(r,r_1;\omega) = r^m r_1^m \bar{\bar{Z}}_m(\omega) \tag{77}$$

where $\bar{\bar{Z}}_m(\omega)$ has dimensions Ω / m^{2m}.

4 TRANSVERSE WAKE FUNCTION

4.1 Transverse wake function and loss factor of a point charge

Let us consider now the leading charge transversely displaced with respect to the axis as shown in Fig. 8. The charge excites in the structure electromagnetic fields which can be expanded in their multipole components (dipole, quadrupole, sextupole etc.) in the transverse plane. For small transverse displacements the dipole term is of course dominant.

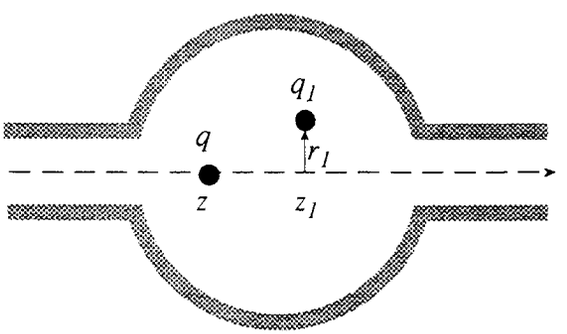


Fig. 8 The leading charge is transversely displaced

The trailing charge q experiences a Lorentz force which has longitudinal and transverse components. Therefore, it is subject to a transverse momentum kick given by:

$$M_{2l}(\boldsymbol{r},\boldsymbol{r}_l;\tau)=\int_{-\infty}^{\infty}\boldsymbol{F}_\perp(\boldsymbol{r},z,\boldsymbol{r}_l,z_l\ ;\ t)\,dz\ ,\quad t=\frac{z_l}{v}+\tau \tag{78}$$

the integration, as for the longitudinal case is assumed over an infinite distance. The above momentum kick, measured in *Newton meter* [Nm], depends on the pipe shape and on the transverse position of both charges. In general, the transverse kick is not parallel to the displacement of the leading charge. In fact a horizontal displacement can lead to both vertical and horizontal kicks, and the same generally happens for a vertical displacement. Only in the case of cylindrical symmetry, are the two transverse directions decoupled for a beam on the axis. The transverse kick per unit of both charges, measured in *volt/coulomb* [V/C], defines the transverse wake function:

$$w_\perp(\boldsymbol{r},\boldsymbol{r}_l;\tau)=\frac{M_{2l}(\boldsymbol{r},\boldsymbol{r}_l;\tau)}{q_l q}\qquad [\text{V/C}] \tag{79}$$

Analogously to the longitudinal case it is useful to define the dipole transverse loss factor as the amplitude of the transverse momentum kick given to the charge by its own wake per unit charge:

$$k_\perp(\boldsymbol{r}_1)=\frac{M_{11}(\boldsymbol{r}_1)}{q_1^2}\qquad [\text{V/C}] \tag{80}$$

Usually the dipole component of the transverse kick is the dominant term for ultra-relativistic charges. This term is proportional to the displacement of the charge q_1. For this particular case we define the *transverse dipole wake function* as the transverse wake per unit of transverse displacement:

$$w'_\perp(\boldsymbol{r},\boldsymbol{r}_l;\tau)=\frac{w_\perp(\boldsymbol{r},\boldsymbol{r}_l;\tau)}{r_l}\qquad [\text{V/Cm}] \tag{80}$$

and a transverse loss factor:

$$k'_\perp(\boldsymbol{r}_1)=\frac{M_{11}(\boldsymbol{r}_1)}{q_1^2 r_1}\qquad [\text{V/Cm}] \tag{81}$$

4.2 Transverse wake function and loss factor of a bunch

The transverse wake potential produced by a continuous bunch distribution, transversely displaced by $\boldsymbol{r}$, can be obtained by applying the superposition principle; we get:

$$W_\perp(\boldsymbol{r};\tau)=\frac{1}{q_1}\int_{-\infty}^{\infty}w_\perp(\boldsymbol{r};\tau-\tau')\,i_b(\tau')\,d\tau'\qquad [\text{V/C}] \tag{82}$$

The bunch transverse loss factor [V/C] is:

$$K_\perp(\boldsymbol{r})=\frac{1}{q_1}\int_{-\infty}^{\infty}W_\perp(\boldsymbol{r};\tau)\,i_b(\tau)d\tau \tag{83}$$

and the transverse wake and loss factor per unit displacement are:

$$W_{\perp}'(r;\tau) = \frac{W_{\perp}(r;\tau)}{r} \tag{84}$$

$$K_{\perp}'(r) = \frac{K_{\perp}(r)}{r} \tag{85}$$

measured in *volt/(coulomb meter)*.

4.3 Relationship between longitudinal and transverse wake functions

Let us consider, for simplicity, a charge which moving with constant velocity v along the z-axis, through an e.m. field. It will experience an e.m. force with components:

$$\begin{aligned} F_z &= q\,E_z \\ F_r &= q\left(E_r - vB_\phi\right) \\ F_\phi &= q\left(E_\phi + vB_r\right) \end{aligned} \tag{86}$$

The Maxwell equations give:

$$\begin{aligned} \frac{\partial E_z}{\partial r} &= \frac{\partial E_r}{\partial z} + \frac{\partial B_\phi}{\partial t} \\ \frac{1}{r}\frac{\partial E_z}{\partial \phi} &= \frac{\partial E_\phi}{\partial z} - \frac{\partial B_r}{\partial t} \end{aligned} \tag{87}$$

from the above relationships, in a moving frame $\zeta = z - vt$, we derive:

$$\nabla_{\perp} F_z = \frac{\partial F_{\perp}}{\partial \zeta} \tag{88}$$

The moving frame has as origin of the axis the position of the leading charge ($z = vt \rightarrow \zeta = 0$), while on the trailing charge we have $\zeta = -v\tau$. Since we are considering the force on the trailing charge, derivation with respect to ζ can be substituted with derivation with respect to τ:

$$-\frac{1}{v}\frac{\partial}{\partial \tau} w_{\perp}(r, r_1; \tau) = \nabla_{\perp, r} w_z(r, r_1; \tau) \qquad \text{[V/Cm]} \tag{89}$$

The transverse operator $\nabla_{\perp}$ applies on r, the transverse coordinates of the trailing charge. The above relation is often referred to as the "Panofsky-Wenzel" theorem [9].

If the leading charge is slightly displaced from the axis, we can expand the rhs of (89) retaining only the first-order term [10]:

$$w_z(r, r_1; \tau) \approx w_z(r, 0; \tau) + \left[\nabla_{\perp, r_1} w_z(r, r_1; \tau)\right]_{r_1 = 0} \cdot r_1 + 0(r_1)^2 \tag{90}$$

where $\nabla_{\perp,r_1}$ is the gradient operator acting on the transverse coordinates $\boldsymbol{r}_1$ of the leading charge. From (89) we get:

$$-\frac{1}{v}\frac{\partial}{\partial\tau}\boldsymbol{w}_{\perp}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = \nabla_{\perp,r}\left\{w_z(\boldsymbol{r},0;\tau)+\left[\nabla_{\perp,r_1}w_z(\boldsymbol{r},\boldsymbol{r}_1;\tau)\right]_{r_1=0}\cdot\boldsymbol{r}_1\right\} \tag{91}$$

The first term in the brackets is a "monopole" contribution to the transverse impedance which disappears for a particular symmetric geometry (circular, rectangular, elliptic). The latter is the dipole transverse impedance, which, in the linear approximation, is obtained from the longitudinal wake expression $w_z(\boldsymbol{r},\boldsymbol{r}_1;\tau)$by applying twice the transverse gradient operator to $\boldsymbol{r}_1$ and $\boldsymbol{r}$:

$$-\frac{1}{v}\frac{\partial}{\partial\tau}\boldsymbol{w}_{\perp}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = \nabla_{\perp,r}\left[\nabla_{\perp,r_1}w_z(\boldsymbol{r},\boldsymbol{r}_1;\tau)\right]_{r_1=0}\cdot\boldsymbol{r}_1 \tag{92}$$

For instance, in Cartesian and cylindrical coordinates, the transverse operator becomes:

$$\nabla_{\perp,r}\nabla_{\perp,r_1} = \begin{vmatrix} \dfrac{\partial^2}{\partial x\partial x_1} & \dfrac{\partial^2}{\partial x\partial y_1} \\ \dfrac{\partial^2}{\partial y\partial x_1} & \dfrac{\partial^2}{\partial y\partial y_1} \end{vmatrix} \quad \nabla_{\perp,r}\nabla_{\perp,r_1} = \begin{vmatrix} \dfrac{\partial^2}{\partial r\partial r_1} & \dfrac{1}{r_1}\dfrac{\partial^2}{\partial r\partial\phi_1} \\ \dfrac{1}{r}\dfrac{\partial^2}{\partial\phi\partial r_1} & \dfrac{1}{rr_1}\dfrac{\partial^2}{\partial\phi\partial\phi_1} \end{vmatrix} \tag{93}$$

It is worth noting that in general, after the application of the matrix (93) to the vector $\boldsymbol{r}_1$, the transverse dipole wake is not necessarily directed along the offset of the driving charge.

4.4 Mode expansion in cylindrical symmetry

As for the longitudinal case, the transverse wake can be expresses as a superposition of multipole terms [11]:

$$\boldsymbol{w}_{\perp}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = \sum_{m=0}^{\infty}\boldsymbol{w}_{\perp,m}(\boldsymbol{r},\boldsymbol{r}_1;\tau) \tag{94}$$

For $\gamma\to\infty$, making use of the expressions (47), (54) and (89), we get:

$$\frac{\partial}{\partial\tau}\boldsymbol{w}_{\perp,m}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = -\mathrm{c}\ \mathrm{m}\ \overline{\overline{w}}_{z,m}(\tau)\ \mathrm{r}^{m-1}\mathrm{r}_1^{m}\left\{\cos(m\phi)\hat{\boldsymbol{r}} - \sin(m\phi)\hat{\boldsymbol{\phi}}\right\} \tag{95}$$

The transverse dipole term $m = 1$ is proportional to the transverse displacement of the leading charge while it does not depend on the transverse position of the trailing one. It is easy to show that in cylindrical coordinates the dipole transverse force is directed along the offset of the leading charge:

$$\frac{\partial}{\partial\tau}\boldsymbol{w}_{\perp,1}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = -\mathrm{c}\ \overline{\overline{w}}_{z,1}(\tau)\ \boldsymbol{r}_1 \tag{96}$$

where, we remember that, $\overline{\overline{w}}_{z,1}$ is the amplitude of the dipole longitudinal wake measured in V/(C m^2). The same result is obtained by applying Eq. (92).

5. TRANSVERSE COUPLING IMPEDANCE

5.1 Definitions and properties

The Fourier transform of the transverse wake function in the frequency domain times the imaginary unity defines the transverse coupling impedance:

$$j\int_{-\infty}^{\infty} w_{\perp}(\boldsymbol{r},\boldsymbol{r}_2;\tau)\, e^{-j\omega\tau} d\tau \equiv \boldsymbol{Z}_{\perp}(\boldsymbol{r},\boldsymbol{r}_2;\omega) \quad [\Omega] \tag{97}$$

Historically, the imaginary constant was introduced in order to make the transverse impedance play the same role as the longitudinal one in the beam stability theory. Since the transverse dynamics is dominated by the dipole transverse wake, we can define the transverse dipole impedance normalized to r_1 as:

$$\boldsymbol{Z}'_{\perp}(\boldsymbol{r}_1,\boldsymbol{r}_2;\omega) = \frac{\boldsymbol{Z}_{\perp}(\boldsymbol{r}_1,\boldsymbol{r}_2;\omega)}{r_1} \quad [\Omega/\mathrm{m}] \tag{98}$$

it has *ohm/meter* units. Conversely, the transverse wake is obtained from the inverse Fourier transform of the transverse impedance:

$$w_{\perp}(\boldsymbol{r},\boldsymbol{r}_2;\tau) = \frac{\mathrm{j}}{2\pi}\int_{-\infty}^{\infty} \boldsymbol{Z}_{\perp}(\boldsymbol{r},\boldsymbol{r}_2;\omega)\, e^{j\omega\tau} d\omega \tag{99}$$

5.2 Relationship between longitudinal and transverse impedances

The Fourier transform of (89) gives the dipole transverse impedance on terms of the longitudinal one:

$$\boldsymbol{Z}_{\perp}(\boldsymbol{r},\boldsymbol{r}_1;\omega) = \frac{c}{\omega}\nabla_{\perp} Z(\boldsymbol{r},\boldsymbol{r}_1;\omega) \quad [\Omega] \tag{100}$$

The transverse dipole impedance for an arbitrary shape, according to Eq. (92) is:

$$\boldsymbol{Z}_{\perp}(\boldsymbol{r},\boldsymbol{r}_1;\omega) = \frac{c}{\omega}\nabla_{\perp,\boldsymbol{r}}\left[\nabla_{\perp,\boldsymbol{r}_1} Z(\boldsymbol{r},\boldsymbol{r}_1;\omega)\right]_{\boldsymbol{r}_1=0}\cdot\boldsymbol{r}_1 \tag{101}$$

In cylindrical symmetry, applying (96) or (92) to Eqs. (76) we get:

$$\boldsymbol{Z}_{\perp,1}(\boldsymbol{r},\boldsymbol{r}_1;\omega) = \frac{c}{\omega}\bar{\bar{Z}}_1(\omega)\,\boldsymbol{r}_1 \quad [\Omega] \tag{102}$$

6. UNIFORM BOUNDARIES

6.1 General properties

In this section we start the analysis of the wake fields and impedances for some relevant cases: charge in the free space and in a beam pipe with uniform cross section. Fields and

potentials for these cases have a common feature: they travel together with the charge. In other words, the field map does not change during the charge flight, as long as the trajectory is parallel to the pipe axis.

Considering a charge with velocity $v = \beta c\,\hat{z}$ we may write:

$$\boldsymbol{E} = -\text{grad V} + \beta^2 \frac{\partial V}{\partial z} z = -\frac{1}{\gamma^2}\frac{\partial V}{\partial z} z - \text{grad}_\perp V \tag{103}$$

where the scalar potential $V(r, \phi, z - vt)$ is the solution of the equation:

$$\nabla_\perp^2 \text{V} + \frac{1}{\gamma^2}\frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\varepsilon} \tag{104}$$

satisfying the boundary conditions. The Laplacian operator $\nabla_\perp^2$ is applied to the transverse coordinates, ρ is the charge density. The longitudinal wake potential per unit length is given by:

$$\frac{\partial \text{w}(\text{r}, \phi, \tau)}{\partial \text{z}} = -\frac{1}{\text{q}} E_z(r, \phi, z - vt)\Big|_{t=\frac{z}{v}+\tau} \tag{105}$$

One can see from Eq. (104) that in the ultra relativistic limit $\gamma \to \infty$, fields can be derived in the static approximation.

6.2 Relativistic charge in free space

A point charge moving with constant velocity $v\hat{z}$ in free space generates fields which are solutions of the Maxwell equations (48, 51). The fields can be derived applying the Lorentz transform to the static field created by the charge in the rest frame. Because of the symmetry, we have the fields [12]:

$$\boldsymbol{E}(\boldsymbol{r}, z, t) = \frac{q}{4\pi\varepsilon_o} \frac{\gamma[\boldsymbol{r} + (z - vt)\hat{z}]}{\left[\gamma^2 (z - vt)^2 + r^2\right]^{\frac{3}{2}}} \tag{106}$$

$$\boldsymbol{B}(\boldsymbol{r}, z, t) = \frac{\beta}{c}\hat{\boldsymbol{n}} \times \boldsymbol{E}(\boldsymbol{r}, z, t) \tag{107}$$

where the $\hat{\boldsymbol{n}}$ vector is directed from the charge to the observation point. The magnetic field has only the azimuthal component B_ϕ. It is well known that at high energies, because of the relativistic contraction, the fields are mainly confined inside a region with an opening angle $1/\gamma$ and perpendicular to the trajectory. The longitudinal field E_z vanishes as $1/\gamma^2$, while E_r and B_ϕ are proportional to γ.

$$E_z\left(r = 0, t = \frac{z}{v} + \tau\right) = \frac{-q}{4\pi\varepsilon_o (\beta\gamma c\tau)^2} \tag{108}$$

$$E_r\left(r, t=\frac{z}{v}\right)=\frac{q}{4\pi\varepsilon_o}\frac{\gamma}{r^2} \tag{109}$$

$$B_\phi\left(r, t=\frac{z}{v}\right)=\frac{qZ_o}{4\pi}\frac{\gamma}{r^2} \tag{110}$$

Because of the field's confinement within an angular region of the order of $1/\gamma$, at a given distance r from the charge, the fields can be thought of as generated by a relativistic charge distribution with line density λ. In the stationary approximation, applying the Gauss law at a cylindrical surface of radius r, we find an effective charge density $\lambda = q\gamma / r$. The singularities at $\tau = 0$ and $r = 0$ can be removed by considering a charge with longitudinal and radial distribution.

According to the definition (6), since a test charge on the axis would experience a repulsive force independently of its position, the longitudinal wake function per unit length is an odd function of τ. The corresponding impedance is purely imaginary. Because of the lack of interest, we do not derive the explicit expressions of the wake and impedance. However, it is interesting to compute the amount of e.m. energy stored in a region outside a tube of radius b:

$$U(r \geq b)=\frac{3\pi}{16}\left(\frac{r_o}{b}\right)\gamma m_o c^2 \tag{111}$$

where r_o is the classic radius and m_o is the rest mass of an electron. The e.m. energy is proportional to the kinetic energy of the charge.

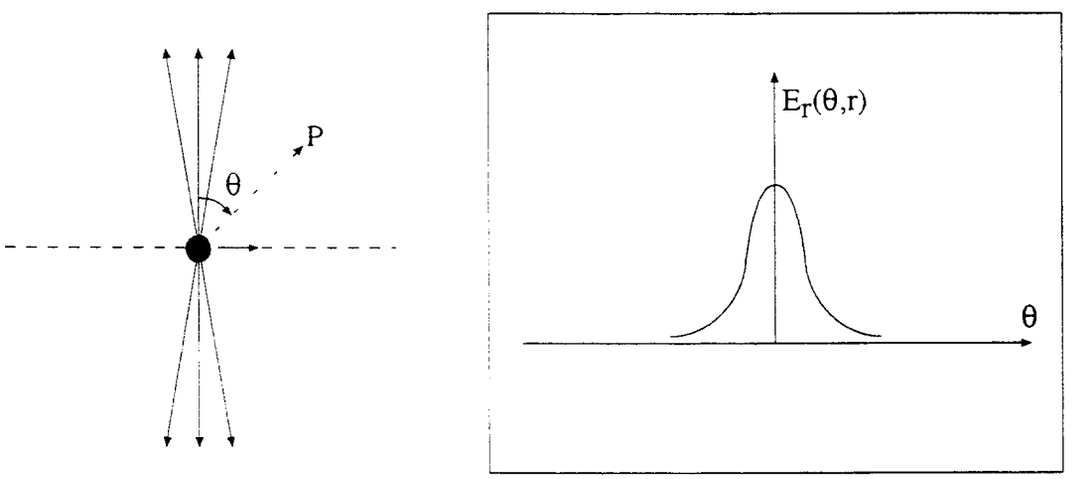


Fig. 9 Electric field lines of an ultra-relativistic charge in free space (qualitative behaviour)

6.3 Cylindrical pipe with perfectly conducting walls

The fields produced by a point charge travelling inside a perfectly conducting cylindrical pipe are found from the scalar potential $V(r, \phi, z - vt)$ solution of the Maxwell equation (48,51), with homogeneous boundary conditions at the pipe wall $r = b$ (Fig. 10).

Using the density charge (44), in cylindrical coordinates we can express the scalar potential as sum of multipole terms [10]:

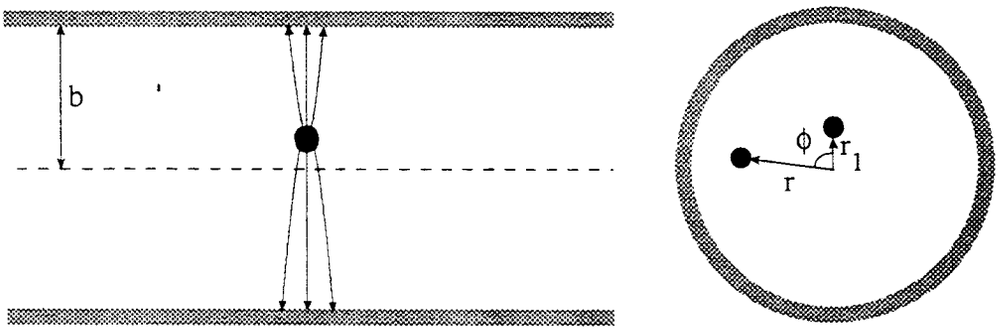


Fig. 10 Point charge inside a perfectly conducting cylindrical pipe

$$V(r,\phi,z-vt)=\frac{1}{2\pi}\sum_{m=0}^{\infty}\cos(m\phi)\int_{-\infty}^{\infty}\tilde{V}_m(r,r_1,\kappa)e^{-j\kappa(z-vt)}d\kappa \tag{112}$$

where we made use of the Fourier transform from the z-space to the wave number domain κ. Each Fourier component $\tilde{V}_m(r,r_1,\kappa)$ is obtained by solving the differential equation (104) and imposing the boundary conditions at the pipe walls [10]. We get:

$$\tilde{V}_m(r,r_1,\kappa)=\frac{q\alpha_m}{2\pi\varepsilon_o}\begin{cases}K_m(\xi r)I_m(\xi r_1)-\dfrac{I_m(\xi r_1)}{I_m(\xi b)}K_m(\xi b)I_m(\xi r); & r\geq r_1\\ K_m(\xi r_1)I_m(\xi r)-\dfrac{I_m(\xi r_1)}{I_m(\xi b)}K_m(\xi b)I_m(\xi r); & r\leq r_1\end{cases} \tag{113}$$

where $\xi=\kappa/\beta\gamma$, and I_m, K_m are the modified Bessel functions.

The longitudinal coupling impedance per unit length, using (103) and (105), is given by:

$$\frac{\partial\bar{Z}_m(r,r_1,\omega)}{\partial z}=\frac{-j\omega}{q(c\beta\gamma)^2}\tilde{V}_m\left(r,r_1,\kappa=\frac{\omega}{\beta c}\right) \tag{114}$$

6.3.1 Monopole longitudinal impedance $m=0$, $r<r_1$

$$\frac{\partial\bar{Z}_{m=0}}{\partial z}=\frac{-j\omega Z_o}{2\pi c(\beta\gamma)^2}\left[K_o(\xi r_1)-\frac{I_o(\xi r_1)}{I_o(\xi b)}K_o(\xi b)\right]I_o(\xi r) \tag{115}$$

Where $Z_o=\sqrt{\mu_o/\varepsilon_o}=1/c\varepsilon_o$ is the impedance of the vacuum. The behaviour of the above term, purely imaginary, as function of ξb is shown in Fig. 11. We note that in the region $\xi b<<1$, the impedance per unit length grows linearly, and does not depend on the radial position of the trailing charge:

$$\frac{\partial\bar{Z}_{m=0}}{\partial z}=\frac{-j\omega Z_o}{2\pi c(\beta\gamma)^2}\ln\left(\frac{r_1}{b}\right) \tag{116}$$

For $\xi b>>1$, the impedance shows an exponential roll off.

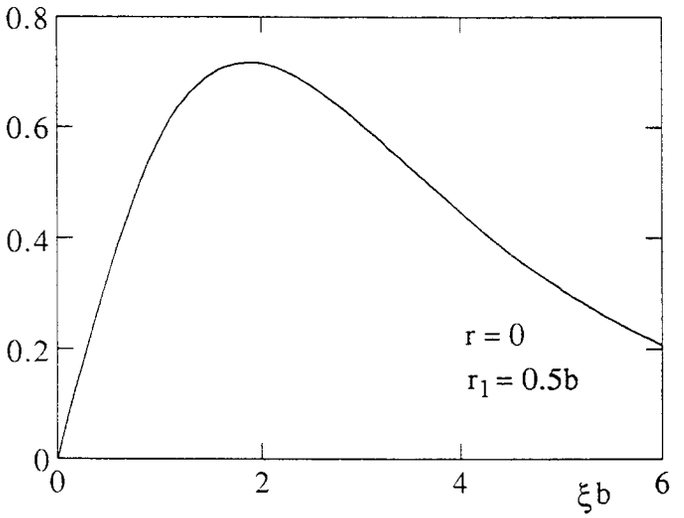


Fig. 11 Monopole space charge impedance versus ξb

It is apparent that the impedance (116) does not satisfy the radial dependence (77) found in the high energy limit. In fact this term is called Space Charge Impedance, and kept distinct from the forces related to the secondary fields induced at the pipe walls. In the literature, it is usually presented as the space charge monopole term due to a disk of radius a, centered on the pipe axis. Integrating the impedance expression (116) over the charge distribution $0 \le r_1 \le a$, for $\xi b << 1$, we get:

$$\frac{\partial \overline{Z}_{m=0}}{\partial z} = \frac{-j\omega Z_o}{4\pi c(\beta\gamma)^2}\left[1 + 2\ln\left(\frac{b}{a}\right)\right] \tag{117}$$

Using the above expression in the inverse Fourier transform, we get in the limit $\gamma \to \infty$ the wake function per unit length:

$$\frac{\partial \overline{w}_{m=0}}{\partial z} = \frac{1}{4\pi\varepsilon_o\gamma^2}\left[1 + 2\ln\left(\frac{b}{a}\right)\right]\frac{\partial}{\partial z}\delta(z - vt) \tag{118}$$

6.3.2 Simple physical approach for $\gamma \to \infty$

We have seen in Sec. 6.1 that at high energies one can solve Maxwell's equations in the static approximation. Accordingly we derive the fields produced by a charged cylinder of radius a, with longitudinal distribution $\lambda(z - vt)$, moving with light velocity inside a perfectly conducting cylinder (Fig. 12).

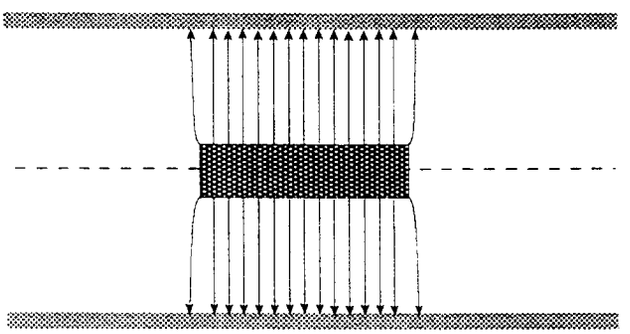


Fig. 12 On axis cylindrical bunch of radius "a"

For ultra relativistic charges, according to (103) and (104), the fields are derived by applying the Gauss and Ampere laws; we get :

$$E_r(r,z-vt)=\begin{cases}\dfrac{l(z-vt)}{2pe_o}\dfrac{r}{a^2}; & r\le a\\ \dfrac{l(z-vt)}{2pe_o}\dfrac{1}{r}; & r\ge a\end{cases} \tag{119}$$

$$B_f(r,z-vt)=\begin{cases}\dfrac{l(z-vt)Z_o}{2p}\dfrac{r}{a^2}; & r\le a\\ \dfrac{l(z-vt)Z_o}{2p}\dfrac{1}{r}; & r\ge a\end{cases} \tag{120}$$

The scalar potential is obtained integrating the radial field (119) from the disk centre ($r = 0$) to the pipe radius ($r = b$); we get:

$$V(r,z-vt)=\frac{\lambda(z-vt)}{4\pi\varepsilon_o}\left\{1-\left(\frac{r}{a}\right)^2+2\ln\left(\frac{b}{a}\right)\right\} \tag{121}$$

And the wake function per unit length at $r = 0$, $t=\frac{z}{v}+t$ becomes:

$$\frac{\partial w_{m=0}}{\partial z}=\frac{1}{q\gamma^2}\frac{\partial V}{\partial z}=\frac{1}{4\pi\varepsilon_o\gamma^2 q}\left[1+2\ln\left(\frac{b}{a}\right)\right]\frac{\partial\lambda(z-vt)}{\partial z} \tag{122}$$

which reproduces Eq. (118) for a point charge with density $\lambda(z-vt)=q\delta(z-vt)$.

6.3.3 Dipole longitudinal impedance $m = 1$, $r < r_1$

The dipole impedance per unit length is:

$$\frac{\partial Z_{m=1}}{\partial z}=\frac{-j\omega Z_o}{2\pi c(\beta\gamma)^2}\left[K_1(\xi r_1)-\frac{I_1(\xi r_1)}{I_1(\xi b)}K_1(\xi b)\right]I_1(\xi r)\cos(\phi) \tag{123}$$

In the limit $\xi b << 1$ the dipole impedance is proportional to the transverse displacement of the charge:

$$\frac{\partial Z_{m=1}}{\partial z}=\frac{-j\omega Z_o}{4\pi c(bg)^2}\left[\frac{1}{r_1^2}-\frac{1}{b^2}\right]rr_1\cos(\phi) \tag{124}$$

6.3.4 Dipole transverse impedance $\xi b << 1$

Applying the relationship (102) between dipole transverse and longitudinal impedances, and noting that:

$$\nabla_\perp[r\cos(\phi)]=\hat{r}\cos(\phi)-\hat{\phi}\sin(\phi)\equiv\hat{r}_1 \tag{125}$$

we get the transverse dipole impedance per unit length and per unit transverse displacement:

$$\frac{\partial Z'_{\perp,1}(\omega)}{\partial z} \equiv \frac{1}{r_1}\frac{dZ_{\perp,1}(\omega)}{dz} = \frac{-jZ_o}{2\pi(\beta\gamma)^2}\left[\frac{1}{r_1^2}-\frac{1}{b^2}\right]\hat{r}_1 \quad [\Omega/\mathrm{m}^2] \tag{126}$$

The same result could be obtained applying Eq. (102), recognizing in (124) the dipole term $\overline{\overline{Z}}_{m=1}$ introduced in Eq. (76):

$$\frac{\partial \overline{\overline{Z}}_{m=1}}{\partial z} = \frac{-j\omega Z_o}{4\pi c(\beta\gamma)^2}\left[\frac{1}{r_1^2}-\frac{1}{b^2}\right] \tag{127}$$

Notice that according to the standard symbols, also the dipole term can be obtained in terms of the radius of a cylindrical beam by putting $r_1 = a$ in Eqs. (124,126,127).

6.4 Elliptic pipe with perfectly conducting walls

The impedance expression Eqs. (116), (117), (126) and (127) have been extended to the case of an elliptic pipe [13] in the ultra-relativistic limit. An equivalent radius b_{eq} is introduced for both longitudinal and transverse cases as function of the elliptic parameter:

$$\overline{q} = \frac{h-b}{h+b} \tag{128}$$

where h and b are the pipe half-width and half-height respectively. The longitudinal equivalent radius normalized to b is reported in Fig. 13. We see that when $h >> b$ the curve approaches the parallel plates case with $b_{eq} \equiv 4b/\pi$. In Fig. 14 the transverse equivalent radius is reported as a function of $\overline{q}$ for both horizontal and vertical oscillations.

6.5 Pipe with lossy walls

When we consider a pipe with resistive wall of infinite thickness, the Maxwell equations have to be solved both in the pipe space and in the material with finite conductivity σ where the fourth Maxwell equation becomes:

$$\nabla \times \boldsymbol{B} = \varepsilon\mu\frac{\partial \boldsymbol{E}}{\partial t} + \mu\sigma\boldsymbol{E} + \mu\rho\boldsymbol{v} \tag{129}$$

Continuity of tangent magnetic field and normal electric field at the wall surface allows us to derive the e.m. fields components. The problem of a cylindrical pipe has been solved in the ultra-relativistic limit [14] and for any value of the parameter ξb [13]. Extension to the elliptic pipe is found in Ref. [15] in the ultra-relativistic limit. More recently the impedance for an arbitrary cross section has been developed [16,17] in the same approximation. Here we present the results for the most relevant cases: circular, rectangular and elliptic pipes.

The longitudinal impedance has the general expression:

$$\frac{\partial \overline{Z}_{m=0}}{\partial z} = \frac{1+j}{2\pi b}\sqrt{\frac{\omega Z_o}{2c\sigma}}\,F \tag{130}$$

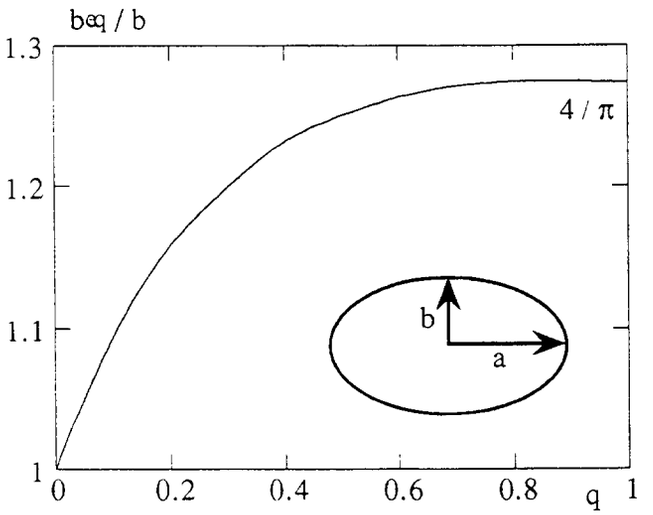


Fig 13 Normalized equivalent radius for an elliptic pipe

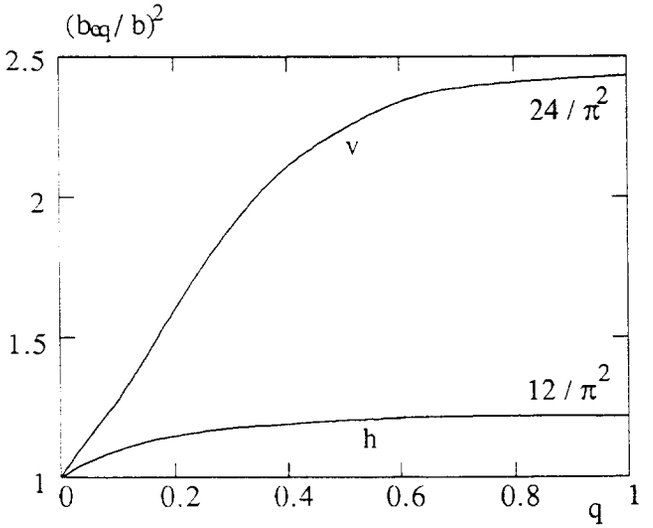


Fig. 13 Normalized $(b_{eq}/b)^2$ for horizontal (h) and vertical (v) oscillations

where F is a form factor depending on the pipe cross section, and b is the half-height of the pipe cross section (b is the radius in the circular case). The inverse Fourier transform of Eq. (130) gives the wake function:

$$\frac{\partial w_{z,m=0}}{\partial z} = -F\frac{1}{4\pi b}\sqrt{\frac{Z_o}{\pi c\sigma}}\,\tau^{-3/2} \tag{131}$$

The above expression, being derived from a static approximation, fails at distances very close to the charge. At very short distances the wake changes sign as shown in Fig. 15.

The transverse dipole resistive wall impedance is:

$$\frac{\partial Z'_{\perp,1}}{\partial z} \equiv \frac{1}{r_1}\frac{\partial Z_{\perp,1}}{\partial z} = \mathrm{F}_{\perp}\frac{1+j}{2\pi b^3}Z_o\delta\,\hat{r}_1 \qquad [\Omega/\mathrm{m}^2] \tag{132}$$

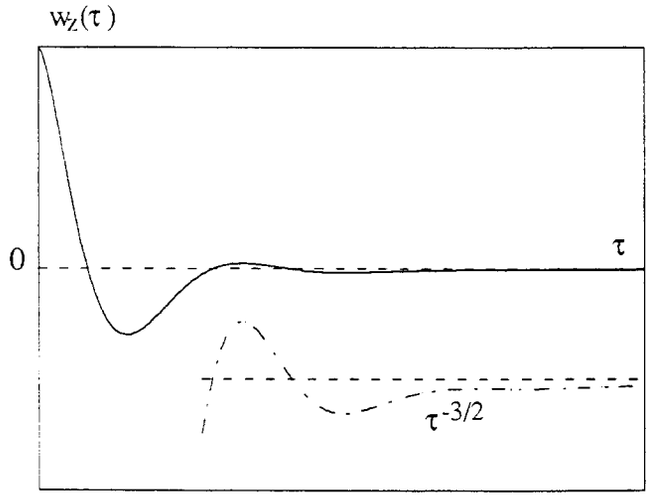


Fig. 15 Qualitative behaviour of the longitudinal wake of a lossy pipe

where $F_\perp$ is the transverse form factor for vertical and horizontal oscillations. The transverse wake is:

$$\frac{dw'_{\perp,l}(\tau)}{dz} = \frac{1}{r_1}\frac{dw_{\perp,l}(\tau)}{dz} = F_\perp \frac{1}{\pi b^3}\sqrt{\frac{cZ_o}{\pi\sigma}}\,\tau^{-1/2} \quad [\mathrm{V/m^2C}] \tag{133}$$

6.5.1 Circular beam pipe

For a circular pipe, $F = 1$ and $F_\perp = 1$, the longitudinal monopole impedance is:

$$\frac{\partial \bar{Z}_{m=0}}{\partial z} = \frac{1+j}{2\pi b}\sqrt{\frac{\omega Z_o}{2c\sigma}} \tag{134}$$

and for the transverse dipole impedance is:

$$\frac{\partial Z_{\perp,l}}{\partial z} = \frac{1+j}{2\pi b^3} Z_o \delta \tag{135}$$

Simple physical approach:

In the simple case of cylindrical symmetry, according to Sec. 2.4 and 3.1, the longitudinal impedance can be computed as the complex power spectrum related to the energy flowing into the lossy walls. For materials with a high conductivity, the fields inside the pipe are almost the same as in the perfect conductor case (perturbative approach). In the frequency domain we have:

$$\begin{aligned} \tilde{E}_r(\omega) &= \frac{qZ_o}{2\pi r}\,\mathrm{e}^{-jkz} \\ \tilde{H}_\phi(\omega) &= \frac{q}{2\pi r}\,\mathrm{e}^{-jkz} \end{aligned} \tag{136}$$

The continuity conditions at the boundary $r = b$ requires that the magnetic field $\tilde{H}_\phi$ component inside the material surface is the same as outside. Inside the wall the field is sustained by a surface current flowing into the z-direction. The electric field $\tilde{E}_z$ is related to $\tilde{H}_\phi$ by the Leontovich condition:

$$\tilde{E}_z(\omega) = Z_c \tilde{H}_\phi(\omega) \tag{137}$$

where:

$$Z_c = \sqrt{\frac{j\omega\mu_o}{\sigma}} \tag{138}$$

is the intrinsic impedance of the lossy material. The flux of the Poynting vector at the pipe wall gives:

$$\frac{\partial \overline{Z}_{m=0}}{\partial z} = 2\pi b Z_c \left|\tilde{H}_\phi\right|^2 = \frac{1+j}{2\pi b}\sqrt{\frac{\omega Z_o}{2c\sigma}} \tag{139}$$

6.5.2 Rectangular cross section.

For a rectangular beam pipe with half-width h and half-height b, putting $\lambda = b/h$, the form factor F for the longitudinal impedance is :

$$F(\lambda) = \pi\left[\sum_{\substack{n=1,\\ odd}}^{\infty} \frac{1}{\cosh^2\left(\frac{n\pi}{2\lambda}\right)} + \lambda \sum_{\substack{n=1,\\ odd}}^{\infty} \frac{1}{\cosh^2\left(\frac{n\pi\lambda}{2}\right)}\right] \tag{140}$$

The form factor for the dipole transverse impedance in the x-direction:

$$F_x(\lambda) = \frac{\pi^3}{8}\left[\sum_{\substack{n=1,\\ odd}}^{\infty} \frac{n^2}{\sinh^2\left(\frac{n\pi}{2\lambda}\right)} + \lambda^3 \sum_{\substack{n=2,\\ even}}^{\infty} \frac{n^2}{\cosh^2\left(\frac{n\pi\lambda}{2}\right)}\right] \tag{141}$$

The function $F_y(\lambda)$ is simply obtained from Eq. (141) by moving the factor λ^3 to the first sum in the brackets:

$$F_y(\lambda) = \frac{\pi^3}{8}\left[\lambda^3 \sum_{\substack{n=1,\\ odd}}^{\infty} \frac{n^2}{\sinh^2\left(\frac{n\pi}{2\lambda}\right)} + \sum_{\substack{n=2,\\ even}}^{\infty} \frac{n^2}{\cosh^2\left(\frac{n\pi\lambda}{2}\right)}\right] \tag{142}$$

The behaviour of $F(\lambda)$, $F_x(\lambda)$ and $F_y(\lambda)$ for the rectangular pipe is presented in Fig. 16 as a function of the parameter $\overline{q}$.

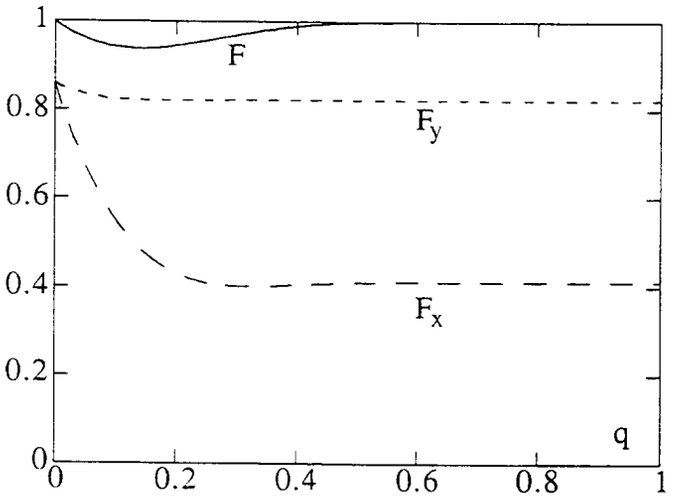


Fig. 16 Form factors for a rectangular pipe

In the limit case of a pair of parallel plates $\lambda \to 0$, we have:

$$F_o(0) = 1, \quad F_x(0) = \frac{\pi^2}{24}, \quad F_y(0) = \frac{\pi^2}{12}$$

6.5.3 Elliptical beam pipe

For a beam pipe with an elliptical cross-section, major axis $2a$ and minor axis $2b$, the form factor is given as a function of the elliptic parameter u_o related to the parameter $\bar{q}$ by:

$$\bar{q} = e^{-2u_o}$$

We get:

$$F(u_o) = \frac{\sinh(u_o)}{2\pi} \int_0^{\infty} \frac{G(u_o, \alpha) d\alpha}{\sqrt{\sinh^2(u_o) + \sin^2(\alpha)}} \tag{143}$$

where

$$G(u_o, \alpha) = 2 \sum_{m=-1}^{\infty} (-1)^m \frac{\cos(2m\alpha)}{\cosh(2mu_o)} \tag{144}$$

The transverse form factor in the x and y directions is:

$$F_{x,y}(u_o) = \frac{\sinh^3(u_o)}{4\pi} \int_0^{\infty} \frac{G_{x,y}^2(u_o, \alpha) d\alpha}{\sqrt{\sinh^2(u_o) + \sin^2(\alpha)}} \tag{145}$$

with

$$G_x(u_o, \alpha) = 2 \sum_{m=0}^{\infty} (-1)^m (2m+1) \frac{\cos[(2m+1)\alpha]}{\cosh[(2m+1)u_o]} \tag{146}$$

$$G_y(u_o,\alpha) = 2\sum_{m=0}^{\infty}(-1)^m(2m+1)\,\frac{\sin[(2m+1)\alpha]}{\sinh[(2m+1)u_o]} \tag{147}$$

A graph of the numerical values of $F(u_o)$, $F_{x,y}(u_o)$ for the elliptical pipe is presented in Fig. 17 as a function of the elliptic parameter $\bar{q}$.

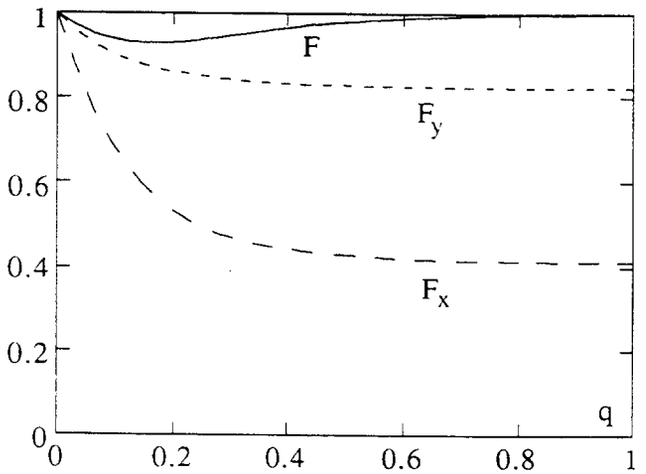


Fig. 17 $F(u_o)$,and $F_{x,y}(u_o)$ as function of $\bar{q}$

7. NON-UNIFORM BOUNDARIES

7.1 General properties

The uniform boundary cases described in the previous section allows one to estimate the effect of smooth pieces of the vacuum chamber on the beam dynamics. Usually we mainly worry about the resistive-wall impedance, which produces a shift of the transverse tunes and drives the head-tail and multibunch instabilities. The pipe is, however, interrupted by many devices installed on the machine, RF cavities, diagnostics, wigglers, cross-section jumps etc.

Unlike the uniform boundary case, the discontinuities in the vacuum chamber are sources of radiated fields which do not travel with the charge. We observe several consequences: excitation of resonant HOMs in resonant structures, new configuration of the self field (after a jump in the cross section), propagation of e.m. fields at frequencies above the cut-off of the beam pipe [18].

As an example, in Fig. 18 we show the case of a relativistic point charge crossing a hole in an infinite perfectly conducting plane.

The diffraction is caused by the primary fields which, impinging on the hole edge, produce secondary scattered fields propagating at the light velocity. The distance for the radiated fields to catch up the charge itself is $\bar{z} = \gamma b$. A test charge travelling a distance $\beta c\tau$ ($\beta \approx 1$) behind will be reached by the same fields at $\bar{z} \approx (b^2 - c^2\tau^2)/(2c\tau)$.

Another basic feature of the diffraction effects concerns the frequency bandwidth of the power spectrum. Despite the point-like nature of the charge, the primary fields exciting the edge have an effective size $\sigma_{eff} = b/\gamma$. The diffraction excitation has a power spectrum extending up to a "radiation cut off frequency" $\omega^{rad}_{cut-off} \approx c\gamma/b$ above which there is an exponential roll off.

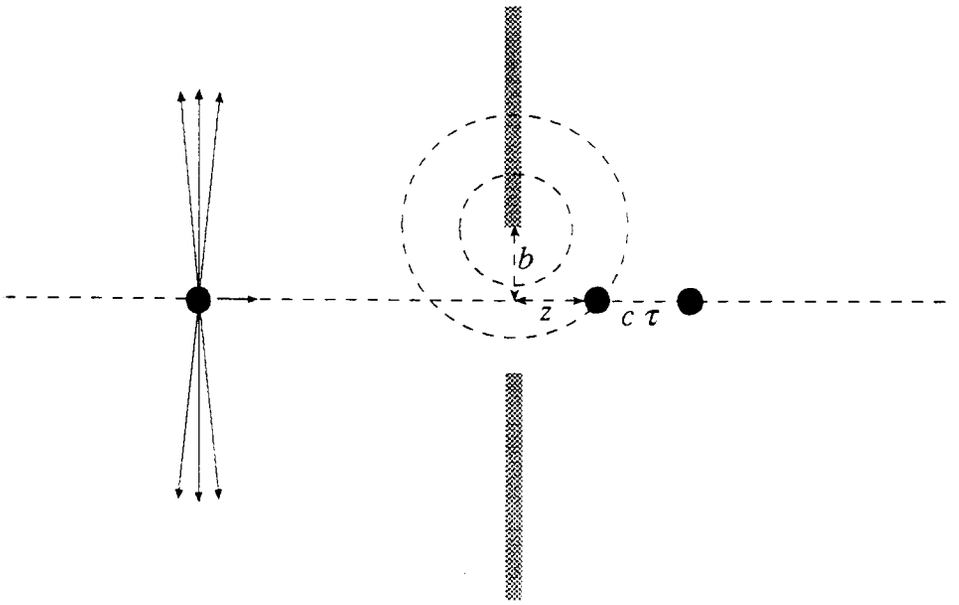


Fig. 18 Relativistic charge passing through a hole of radius b

The geometry of Fig. 18 has been extensively studied [19]. An ultra-relativistic charge passing through the hole loses the energy:

$$U_{11} = 2U(r \geq b) = \frac{3\pi}{8}\left(\frac{r_o}{b}\right)\gamma m_o c^2 \tag{148}$$

It is interesting to note that the energy loss is twice as much as given by Eq. (111), i.e. the amount of energy stored outside the tube of radius b. This feature has been found also for other geometries (such as the step discontinuity): an ultra-relativistic point charge deposits the same amount of energy in rebuilding the self field as in the radiated fields. This result, in general, can be explained as the typical phenomenon occurring in the charge or discharge of a capacitor.

At low frequencies the longitudinal impedance is [20]:

$$Z(\omega) \approx \frac{Z_o}{2\pi\beta}\left[\log\left(\frac{1+\beta}{1-\beta}\right) - \beta + i\pi\right] \tag{149}$$

7.2 A step transition

Let us consider an abrupt change in the cross section of a circular beam pipe from a radius b to radius d (Fig. 19). When the charge crosses the vacuum-chamber discontinuity secondary fields are scattered at the sharp edges. The total fields, "primary" plus "secondary" diffracted fields, are such as to restore the boundary condition at the pipe walls. This problem has been treated by several authors with numerical and analytical techniques [21,22]. An exact analytical solution has been found for a discontinuity made of two coaxial circular pipes for which both longitudinal and transverse dipole impedances have been derived [23,24]. Here we will report the main relevant results and features.

We will distinguish two cases: a particle exiting into a beam pipe of a bigger radius, "step-out" case, and a particle entering a narrowing pipe, "step-in" case. Theoretical results show that the impedance is mostly resistive in the step-out case with a big contribution at high

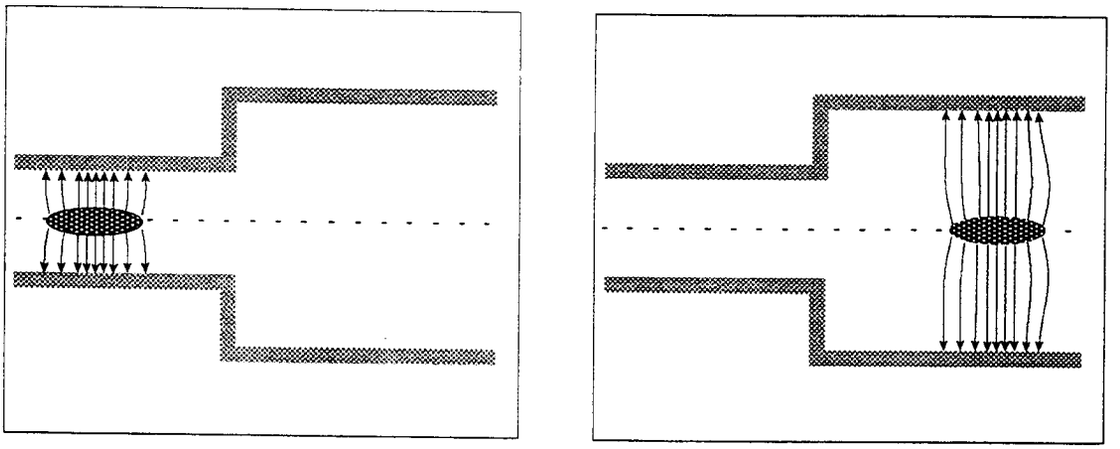


Fig. 19 Step discontinuity in the beam pipe

frequencies above cut off, while in the step-in case the impedance is low, vanishing at high frequencies. In Fig. 20 we show examples of impedances for the step-in and step-out problems, as derived with the computer code ABCI [25].

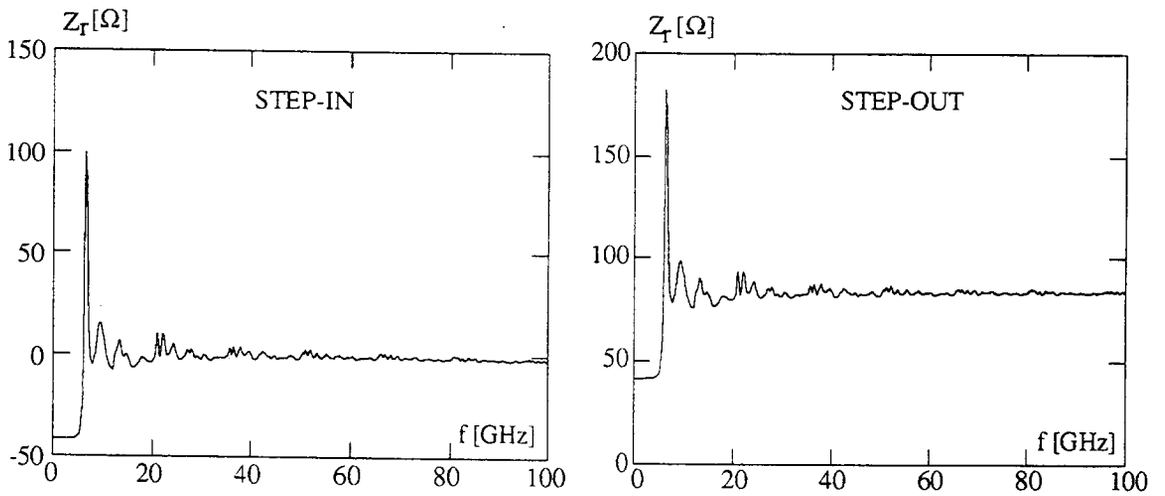


Fig. 20 Longitudinal impedances of a step-in and step-out discontinuity

The impedances have a resonant behaviour just after the beam pipe cut-off and reaches a constant asymptotic value at high frequencies. The asymptotic behaviour of the real part of the impedance is:

$$Z^{out}_{m=0} \sim \frac{Z_o}{\pi} \ln\left(\frac{d}{b}\right) \quad and \quad Z^{in}_{m=0} \sim 0 \tag{150}$$

Such different results are explained by recognizing two main effects contributing to the energy loss. In the step out case, when the charge crosses the discontinuity, the self field restoring the boundary conditions has to fill the extra space $b < r < d$ between the two pipes, while diffracted fields propagate into the pipes. Both these effects lead to an energy loss that can be put as:

$$q^2 k^{out} = U(b < r < d) + E_{rad} \tag{151}$$

where E_{rad} is the energy radiated at the edges and $U(b < r < d)$ is the energy necessary to fill the region $b < r < d$.

In the step-in case, the radiated energy is reflected back with respect to the particle motion without changing its kinetic energy.

$$q^2 k^{out} = -U(b < r < d) + E_{rad} \tag{152}$$

For a point charge, since the radiated energy is taken out of the energy "missing" in the smaller radius pipe: $E_{rad} \approx U(b < r < d)$, we have:

$$\begin{aligned} q^2 k^{out} &\sim 2U(b < r < d) \\ q^2 k^{in} &\sim 0 \end{aligned} \tag{153}$$

We remember that for a real bunch both $U(b < r < d)$ and E_{rad} depend on the bunch length. In particular, if the bunch spectrum does not cover significantly the frequency region above the pipe cut-off, there is no radiation.

In Fig. 21 we show the dependence of the loss factor on the bunch-length for the step-in and step-out cases, for a step with $b = 2$ cm and $d = 4$ cm. In this case we find that a long bunch loses energy in the step-out, but would regain the same amount of energy in a symmetric step-in. Therefore, if a long bunch crosses a pipe enlargement formed by a step-out and step-in sequence with same radii, the total energy loss is almost zero; in this case the wake function inside the bunch is an odd function and the impedance practically inductive.

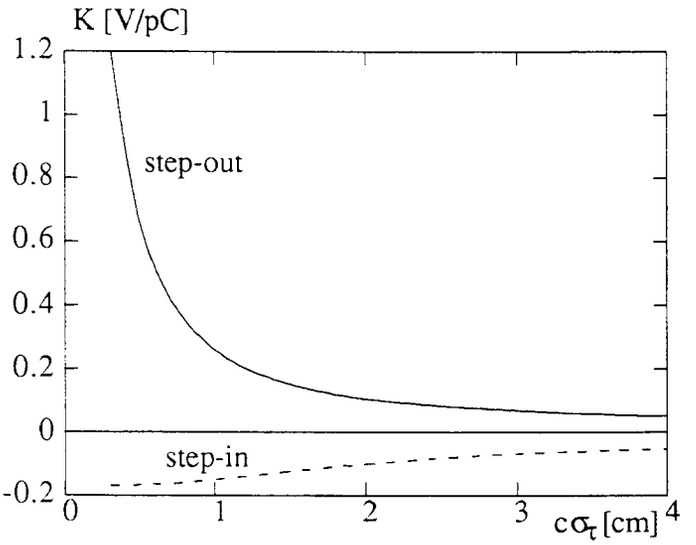


Fig. 21 Example of step-in and step-out loss factor versus bunch length

The dipole longitudinal and transverse impedance for a discontinuous coaxial pipe has been derived in [24]. The model allows for an exact solution of the e.m. problem, and furnishes simple expressions of the impedance at high frequencies that can be used also for real step transitions:

$$Z_{m=1}^{out} \sim \frac{Z_o}{2\pi^2}\left[\frac{1}{b^2}-\frac{1}{d^2}\right] r r_1 \cos(\phi) \qquad [\Omega] \tag{154}$$

$$Z_{m=1}^{in} \sim 0$$

The high frequency transverse dipole impedance for the step out case is:

$$\mathbf{Z}'_{\perp,1} \equiv \frac{1}{r_1}\mathbf{Z}_{\perp,1} = \frac{cZ_o}{2\pi^2\omega}\left[\frac{1}{b^2}-\frac{1}{d^2}\right]\hat{\boldsymbol{r}}_1 \qquad [\Omega/\mathrm{m}] \tag{155}$$

Simple physical approach.

To complete this section we find it worth while to show that the asymptotic expressions (150) can be derived in with a simple physical approach. We compute the energy $U(b<r<d)$:

$$U(b<r<d) = \varepsilon_o \int_V E_r^2 dV \tag{156}$$

For ultra-relativistic charges in a cylindrical pipe, the fields are almost the same as in the free space:

$$E_r \sim \frac{q}{2\pi\varepsilon_o r}\left(\frac{\gamma}{b}\right) \tag{157}$$

Integration (156) gives:

$$\begin{aligned} U(b<r<d) &\sim \frac{q^2 Z_o}{2\pi}\ln\left(\frac{d}{b}\right)\left(\frac{c\gamma}{b}\right) \\ k^{out} &= 2\frac{\Delta W}{q^2} \sim \frac{Z_o}{\pi}\ln\left(\frac{d}{b}\right)\left(\frac{c\gamma}{b}\right) \end{aligned} \tag{158}$$

Remembering that $\sigma_{eff} = (b/\gamma)$ is the effective charge size and that as far as k^{out} is inversely proportional to the size, we can expect that $Z_r(\omega)$ is a constant function of frequency (see the discussion in 3.2). Now, applying the definition (72) and considering the spectrum of a bunch with rectangular distribution, we get for a constant impedance:

$$K = \frac{1}{\pi}\int_0^\infty \mathrm{Z_r}(\omega)\frac{\sin^2(\omega\sigma_{eff}/2c)}{(\omega\sigma_{eff}/2c)^2}d\omega = \frac{\mathrm{Z_r}(\omega)}{\sigma_{eff}/c}. \tag{159}$$

From the comparison (158) and (159) the expected result follows.

7.3 Taper

If one uses long gradual tapers instead of the abrupt step transitions the total energy loss may be drastically reduced. Indeed, the infinitely long taper reduces the radiated energy E_{rad} to zero. For a point charge we have:

$$k_{taper}^{out} \sim \frac{U(b<r<d)}{q^2} = \frac{1}{2} k_{step}^{out}$$
$$k_{taper}^{in} \sim -\frac{U(b<r<d)}{q^2} = -\frac{1}{2} k_{step}^{out} \tag{160}$$

It means that in the limit of long tapers the loss factor of a taper-out reaches half the value of the loss factor for a step-out. There may be even an energy gain for the taper-in case. It must be considered, however, that in a vacuum chamber of a circular accelerator there are taper-in and taper-out transitions. As it can be easily seen, long symmetric tapers reduce total losses practically to zero:

$$k = k_{taper}^{out} + k_{taper}^{in} \sim 0 \tag{161}$$

Certainly, we can not use infinitely-long tapers in an accelerator design. In [18] it is shown that for a short bunch of rms length σ the dependence of the longitudinal loss factor of a one-sided taper on its angle can be approximated by the formula:

$$K = \frac{Z_o c}{2\sigma\pi^{3/2}} \left[1 - \frac{\tilde{\eta}_1}{2} \right] \ln \frac{d}{b} \tag{162}$$

where

$$\tilde{\eta}_1 = \min\left\{ 1, \ \frac{g\sigma}{(d-b)^2} \right\} \tag{163}$$

For a symmetric taper, $\tilde{\eta}_1 / 2$ is replaced by $\tilde{\eta}_1$. So, the condition:

$$\frac{g\sigma}{(d-b)^2} > 1 \tag{164}$$

can be considered as an approximate criterion to choose a reasonable taper length. We should say here that the formula (162) is valid for short bunches when the main contribution to the losses comes from the high-frequency impedance and the diffraction model [18,26] can be applied. Because of that we advise the reader to use numerical codes in order to check the criterion (164) for any particular case. As an example, we show in Fig. 22 the loss factor versus the taper length g for a tapered-out structure passing from b = 2 cm to d = 4 cm. One can see that the loss factors for a bunch of 0.5 cm and 1.0 cm computed by the code ABCI, approach an asymptotic value corresponding to the case of no radiation, at a value of g such that $\tilde{\eta}_1 = 1$. For the 3 cm bunch length the whole bunch spectrum lies below the beam pipe cut-off and no radiation occurs.

7.4 Single-cell cavity

Cross-section variations in an accelerator vacuum chamber can create resonant cavities. Part of the fields excited in the cavities is entrapped reflecting back and forth generating the resonant modes. Above cut-off the amplitudes of the resonance drops because of the energy leakage into the vacuum chamber, the resonances overlap leading to the smooth, "broad-band" impedance.

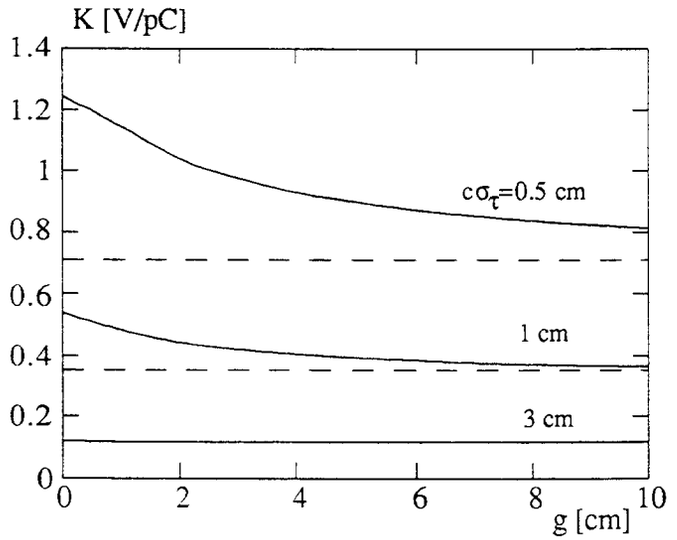


Fig. 22 Loss factor of a tapered discontinuity

A typical cavity impedance is shown in Fig. 23. At the frequencies below ω_c a real high-Q cavity has many sharp resonances. In a RF cavity the fundamental one is used to supply energy to the beam; all the others are "parasitic" modes (higher-order modes - HOM) which subtract energy from the beam. Above cut-off the resonances are broadened.

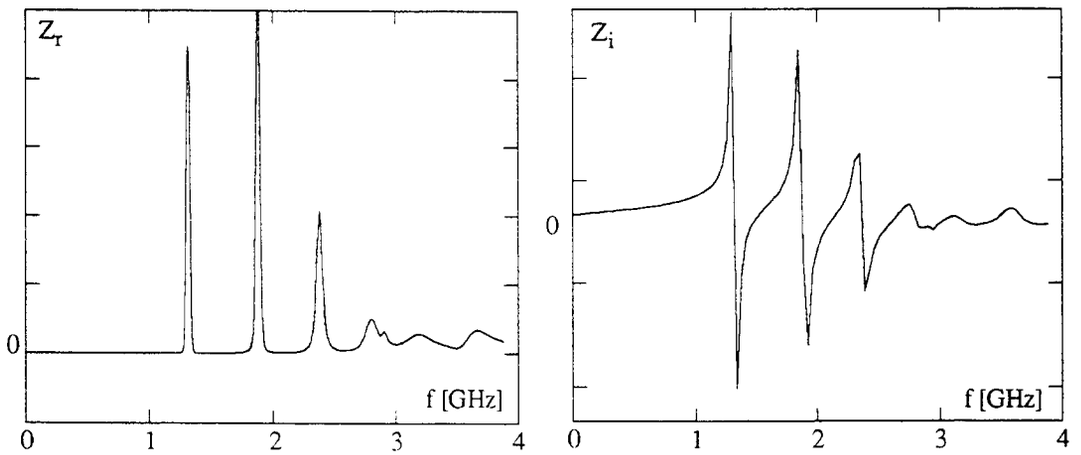


Fig. 23 Typical impedance spectrum for a cavity with attached tubes

7.4.1 Monopole HOM (longitudinal)

In Sec. 2.2.1 we have found the wake potential of a single HOM. Following the results (47) of Sec. 2.8 for cylindrical symmetry, we can write the longitudinal wake of a monopole HOM ($m = 0$) as:

$$w_{z,o}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = 2k_o(r,r_1)e^{-\Gamma_o\tau}\left[\cos(\bar{\omega}_o\tau) - \frac{\Gamma_o}{\bar{\omega}_o}\sin(\bar{\omega}_o\tau)\right]H(\tau) \tag{165}$$

$$\text{with}\quad \Gamma_o = \frac{\omega_o}{2Q_o}\ ,\ \bar{\omega}_o^2 = \omega_o^2 - \Gamma_o^2 \tag{166}$$

In Sec. 2.9 we have also found that in the ultra-relativistic limit the monopole longitudinal wake does not depend on the radial displacement of both leading and trailing charges (54). This result is conveniently exploited in the numerical codes where the loss factor $k_o(r,r_1)$ is computed at the pipe radius, thus limiting the calculation of the energy loss over a definite and limited path:

$$k_o(r,r_1) \equiv k_o(b) = \frac{\omega_o R_o}{2Q_o} = \frac{|V_o(b)|^2}{2U_o} \tag{167}$$

where $V_o(b)$ is the voltage gain computed at $r = b$ and U_o is the average energy stored in the HOM.

Applying the Fourier transform to (165) we get the longitudinal impedance of a monopole resonant HOM:

$$Z(\omega) = \frac{R_o}{1 + jQ_o\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)} \tag{168}$$

It is interesting to note that the shunt impedance is also defined as:

$$R_o = \frac{|V_o(b)|^2}{P_{od}} T^2 \tag{169}$$

where P_{od} is the power dissipated at the cavity wall or in any damping device (loops, waveguides etc.), and T is the transit-time factor defined as the ratio between the accelerating voltage seen by a travelling charge and the voltage at the gap:

$$T = \frac{1}{\int_{gap} E_z \, dz} \left| \int_{gap} E_z \, e^{jkz} dz \right| \tag{170}$$

which takes into account the time evolution of the fields during the cavity crossing. The transit-time factor approaches unity at low frequencies (wavelength much bigger than the gap).

In the low frequency limit $\omega \rightarrow 0$ the impedance is purely inductive. In case of n HOMs we have:

$$Z(\omega) = j\omega \sum_n \left[\frac{R_n}{Q_n \omega_n} \right] = j\omega L \tag{171}$$

7.4.2 Dipole HOM (longitudinal)

The wake potential of a single dipole ($m = 1$) HOM for cylindrical symmetry is given by:

$$w_{z,1}(\boldsymbol{r}, \boldsymbol{r}_1; \tau) = 2\cos(\phi) k_1(r,r_1) e^{-\Gamma_1 \tau} \left[\cos(\overline{\omega}_1 \tau) - \frac{\Gamma_1}{\overline{\omega}_1} \sin(\overline{\omega}_1 \tau) \right] H(\tau) \tag{172}$$

$$\text{with} \quad \Gamma_1 = \frac{\omega_1}{2Q_1} \;,\; \overline{\omega}_1^2 = \omega_1^2 - \Gamma_1^2 \tag{173}$$

The dipole wake potential in the ultra-relativistic limit is proportional to the transverse displacements of both charges. Therefore, we may scale the loss factor computed at the pipe radius as:

$$k_1(r,r_1) \equiv k_1(b)\frac{rr_1}{b^2} \tag{174}$$

$$k_1(b) = \frac{\omega_1 R_1}{2Q_1} = \frac{|V_1(b)|^2}{2U_1} \tag{175}$$

$$Z_1(\boldsymbol{r},\boldsymbol{r}_1;\omega) = \frac{R_1}{1+jQ_1\left(\dfrac{\omega}{\omega_1}-\dfrac{\omega_1}{\omega}\right)}\frac{rr_1}{b^2}\cos(\phi) \tag{176}$$

7.4.3 Dipole HOM (transverse)

The general relationship between transverse and longitudinal wake functions makes it possible to obtain the transverse dipole wake function (96):

$$w_{\perp,1}(\boldsymbol{r},\boldsymbol{r}_1;\tau) = \boldsymbol{r}_1\frac{2c}{\overline{\omega}_1 b^2}k_1(b)e^{-\Gamma_1\tau}\sin(\overline{\omega}_1\tau)H(\tau) \tag{177}$$

The transverse impedance (102) is:

$$Z_{\perp,1}(\boldsymbol{r},\boldsymbol{r}_1;\omega) = \left(\frac{c}{\omega}\right)\frac{\dfrac{R_1}{b^2}}{1+jQ_1\left(\dfrac{\omega}{\omega_1}-\dfrac{\omega_1}{\omega}\right)}\boldsymbol{r}_1 \tag{178}$$

sometimes expressed as:

$$Z_{\perp,1}(\boldsymbol{r},\boldsymbol{r}_1;\omega) = \left(\frac{\overline{\omega}_1}{\omega}\right)\frac{R'_{\perp,1}}{1+jQ_1\left(\dfrac{\omega}{\omega_1}-\dfrac{\omega_1}{\omega}\right)}\boldsymbol{r}_1 \tag{179}$$

with

$$R'_{\perp,1} = \frac{cR_1}{\overline{\omega}_1 b^2} = \frac{\overline{\omega}_1}{c}R_\perp^{URMEL} \tag{180}$$

The quantities R_l and $R'_{\perp,l}$ are derivable from the output data of numerical codes like URMEL.

7.4.2 High-frequency case

The high-frequency impedance is mainly due to the interaction of the charges with the fields diffracted at the cavity-pipe edge. When the bunch length is much smaller than the beam pipe radius, the high frequency contribution to energy loss and impedance can be dominant. Numerical calculations require significant time to find the wake potentials of very short bunches, therefore an analytical analysis of the high-frequency behaviour of the impedance (short-range wake potential) becomes necessary. Methods of diffraction theory are used to calculate the impedance at the high frequencies, $\omega >> c / b$. For a pill-box cavity of length g radius d, Fig. 24, and with side pipes of radius b the diffraction model gives [27,28,29]:

$$Z(\omega) = (1-j)\frac{Z_o}{2\pi b}\sqrt{\frac{gc}{\omega\pi}} \tag{181}$$

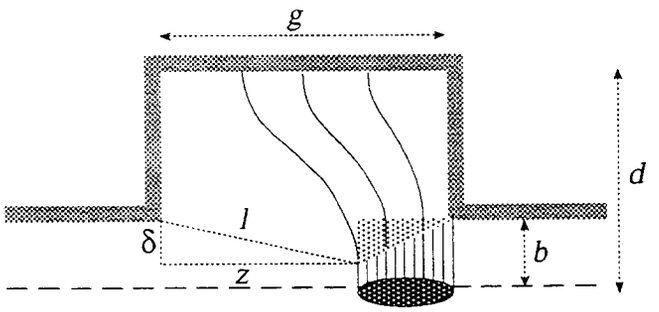


Fig. 24 Relevant parameters for a pill box with attached tubes

The formula is valid for the region of parameters where $g << kb^2$ (a 'cavity regime'). For the region of parameters $g >> kd^2$ the diffraction model of [26] gives the same impedance as found for a "step":

$$Z(\omega) = \frac{Z_o}{\pi}\ln\frac{d}{b} \tag{182}$$

In the transition region of parameters $kb^2 << g << kd^2$

$$Z(\omega) = \frac{Z_o}{2\pi}\ln\frac{gc}{\omega b^2} \tag{183}$$

The transition from the cavity to the step regime is explained by the transition from the Fresnel diffraction for the cavity to the Fraunhofer diffraction for the step. It was shown that the transition from one regime to another occurs when:

$$\eta = \frac{2g\sigma}{(d-b)^2} \sim 1 \tag{184}$$

The cavity regime is reached when the parameter $\eta << 1$. In the opposite case, $\eta >> 1$, the regime of a step is fulfilled.

For a short Gaussian bunch, for which $\sigma << b$, the high frequency tail of the impedance mainly contributes to the energy loss. In this case the longitudinal and transverse loss factors are given by:

$$K = \frac{Z_o c}{4\pi^2 b}\Gamma\left(\frac{1}{4}\right)\sqrt{\frac{g}{\pi\sigma}}, \qquad \Gamma\left(\frac{1}{4}\right) = 3.6256. \tag{185}$$

$$K_\perp = \frac{Z_o c}{4\pi}\frac{\sqrt{\pi g\sigma}}{b^3}$$

Simple physical approach

The high-frequency behaviour of the impedance and loss factor is qualitatively explained by the features of the fields diffracted at the cavity edges [30]. Consider a bunch of charge q and rms size σ passing a pill-box cavity (Fig. 24). The field excited by the head of the bunch and diffracted by the left edge, touches the bunch tail if:

$$\frac{l}{c} = \frac{z+\sigma}{\beta c} \qquad with \qquad l = \sqrt{\delta^2 + z^2} \tag{186}$$

When the bunch leaves the cavity, for $z = g$, the self-field has been perturbed in the shadow region with $\delta \sim \sqrt{2\sigma g}$. The bunch has to rebuild the field in this region in order to restore the boundary condition in the rhs beam pipe. The e.m. energy removed in the shadowed region of Fig. 24 is:

$$U \sim \frac{q^2}{4\pi\varepsilon_o\sigma}\int_{b-\delta}^{b}\frac{dr}{r} \tag{187}$$

if $\delta << b$:

$$U \sim \frac{q^2}{4\pi\varepsilon_o\sigma}\frac{\delta}{b} = \frac{q^2}{4\pi\varepsilon_o\sigma}\frac{\sqrt{2\sigma g}}{b} \tag{188}$$

thus giving:

$$K = \frac{U}{q^2} = \frac{1}{4\pi\varepsilon_o b}\sqrt{\frac{2g}{\sigma}} \tag{189}$$

Note that the condition $\delta << b$ implies also that $\sigma << b$ (g is comparable with b) and most of the bunch spectrum lies well above the beam pipe cut-off frequency. As far as the loss factor is inversely proportional to the bunch length the impedance scales as $1/\sqrt{\omega}$. For a point charge the effective bunch length is b/γ giving the scaling for the loss factor $k \sim \sqrt{\gamma}$.

7.5 Periodic RF structure

An array of periodic cavities shows an impedance spectrum similar to the single-cell cavity below the cut-off frequencies. Many sharp resonant modes appear in the impedance spectrum corresponding to the normal modes of the single cavities but also to the coupled modes between cells, as in a system of coupled oscillators. Above cut-off, however, the broad impedance decays at high frequencies with the asymptotic law $\omega^{-3/2}$ [31,32]. A qualitative picture of the diffraction phenomenon at the basis of such behaviour is given in [30]. Basically, the energy refilling becomes much less in a periodic structure, since the self-field has no time to be rebuild from one cell to the next. The transition from the single-cell regime to the

periodic-structure regime has also been studied [33]. For a structure of M rf cells of period d and iris radius b, we have the transition from the single-cell regime to the periodic-structure regime at the frequency:

$$\omega_t \approx \frac{cMd}{b^2} \tag{190}$$

above which the impedance follows the single-cell asymptotic behaviour $\omega^{-1/2}$. It is worth noting that in the periodic structure, the asymptotic decay $\omega^{-3/2}$ makes the loss factor independent from the bunch length and from the energy. Similar expressions have been found for the case of a charge passing through an array of holes in infinite planes, reproducing the same diffraction phenomenon [34,35].

7.6 Small discontinuities

In a real vacuum chamber there are many small discontinuities such as shallow cavities, tapers, masks, bellows, etc. In spite of the small size, their overall contribution to the inductive impedance, up to rather high frequencies, cannot be neglected. These elements give the main contribution to the longitudinal inductive impedance responsible for the potential-well bunch-lengthening process.

At low frequencies a satisfactory estimate of the inductive impedance $Z(\omega) = j\omega L$ can be obtained by applying Faraday's law in the static approximation. However, it has been shown that one can get more accurate results taking into account a correction factor coming from distortion of the electric field at the chamber discontinuity [36]. Here we do not discuss the method used by authors, but reproduce helpful expression for the impedance of some typical discontinuities shown in Fig. 25.

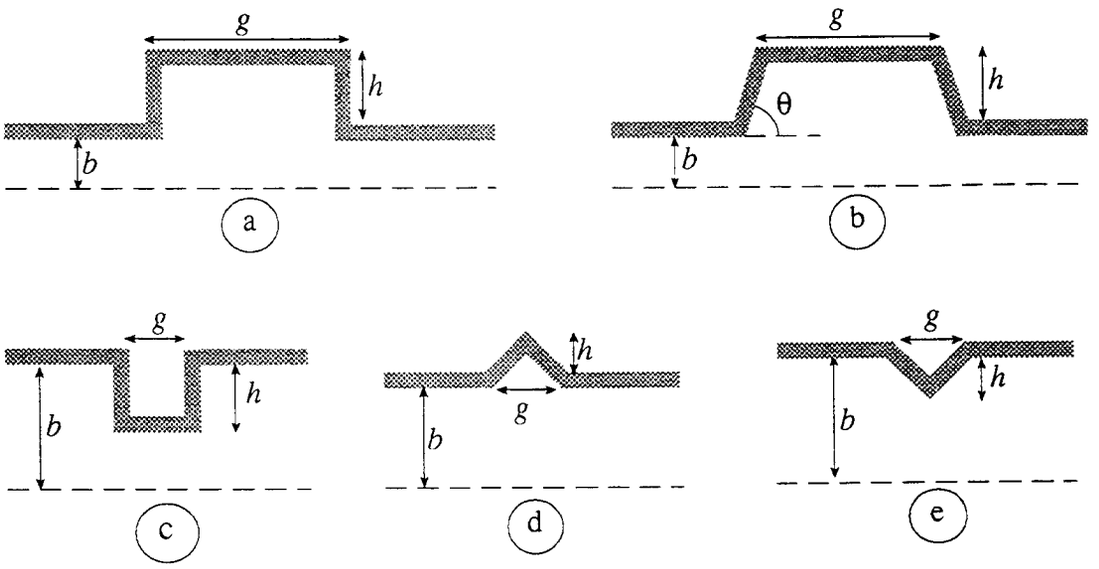


Fig. 25 Examples of vacuum chamber discontinuities

7.5.1 Shallow cavity (Fig. 25a)

The low-frequency impedance of the small, short pill-box ($g < h$) is given by:

$$Z(\omega) = j\omega \frac{Z_o}{2\pi bc}\left(gh - \frac{g^2}{2\pi}\right) \tag{191}$$

For the opposite extreme, $g >> h$, but still $g < b$:

$$Z(\omega) = j\omega \frac{Z_o h^2}{2\pi^2 bc}\left(2\ \ln\left(\frac{2\pi g}{h}\right) + 1\right) \tag{192}$$

When the length g of the pill-box is greater than the pipe radius b, the pill-box can be considered as composed of two independent steps each giving the following contribution to the impedance:

$$Z(\omega) = j\omega \frac{Z_o h^2}{4\pi^2 bc}\left(2\ln\left(\frac{2\pi b}{h}\right) + 1\right) \tag{193}$$

The transverse impedance of the shallow cavity, if the assumption $(g,h) < b$ is valid, is given by :

$$Z_\perp = jZ_o \frac{g}{\pi b^2}\left(\frac{d^2 - b^2}{d^2 + b^2}\right) \tag{194}$$

where $d = b + h$.

7.5.2 Shallow trapezoid (Fig. 25b)

For a long shallow trapezoid ($g >> h$) the impedance of a single sloping step (taper) with the slope angle $\theta = \pi v$ has the form:

$$Z(\omega) = j\omega \frac{Z_o h^2}{2\pi^2 bc}\left[\ln \pi v\left(\frac{b}{h} - 2\cot \pi v\right) + \frac{3}{2} - \gamma - \psi(v) - \frac{\pi}{2}\cot \pi v - \frac{1}{2v}\right] \tag{195}$$

where $\gamma = 0.5772...$ is Euler's constant, $\psi(v)$ is the "psi" function and the transition is assumed to be short compared to the chamber radius, i.e. its length $l = h\cot \pi v << b$.

7.5.3 Shallow iris (Fig. 25c)

Two extreme cases can be considered for the iris geometry. When $g >> h$ the impedance of the iris coincides with that of a shallow cavity. For the case of a thin (or deep) iris $g << h$ the expression for the longitudinal impedance takes the form:

$$Z(\omega) = j\omega \frac{Z_o}{4bc}\left\{h^2 + \frac{gh}{\pi}\left[2\ln(8\pi g / h) - 3\right]\right\} \tag{196}$$

6.5.4 Discontinuities of a triangular shape (Fig. 25d,e)

A short discontinuity of a triangular-shaped cross section with height h and base g ($g << h$) has the following low-frequency impedance:

$$Z(\omega) = j\omega \frac{Z_o}{4\pi bc}\left(gh - \frac{g^2}{\pi}\right) \tag{197}$$

The impedance of a triangular iris is given by:

$$Z(\omega) = j\omega \frac{Z_o}{4bc}\left[h^2 + \frac{2gh}{\pi}(1 - \ln 2)\right] \tag{198}$$

For the case of shallow triangular perturbations, $h << g < b$, both the enlargement and iris have the same inductive impedance, independent of g:

$$Z(\omega) = j\omega \frac{2Z_o h^2 \ln 2}{\pi^2 bc} \tag{199}$$

7.7 Elements of beam diagnostics

Each accelerator has numerous diagnostic elements such as button pick-ups, strip-line beam position monitors, etc.. Impedance calculation of such elements is a rather difficult task. Being inserted into a vacuum chamber a diagnostic element breaks the vacuum chamber symmetry and one has to analyze the interaction of a bunch with a complicated 3-dimensional structure.

Numerical solution of the problem is also not a simple task. Usually, sizes of the diagnostics element are small. This demands a very fine mesh, and so large computation time. Moreover, the image currents induced in the elements flow into the external circuits. The correct simulation of this external load appears to need additional analytical efforts or the use recently developed sophisticated numerical codes.

The method which allows an analytical treatment of the impedance at low frequencies, $\omega << \omega_c$, is based on the electrical approach:

- the 3D geometry is substituted by an equivalent circuit consisting of concentrated radio-technical elements and transmission lines;
- methods of electric circuits and theory of transmission lines are used to find currents and voltages in the circuit elements;
- in the case of a matched load, the coupling impedance is defined by considering the power lost by a bunch to be equal to that dissipated in the load. Some other consideration can be taken into account to relate the beam-coupling impedance with the currents and voltages in the circuit elements.

Here we will closely follow the treatment of [39] in order to illustrate the method. Let us consider a strip-line pick-up of the length l covering the azimuthal angle ϕ (see Fig. 26).

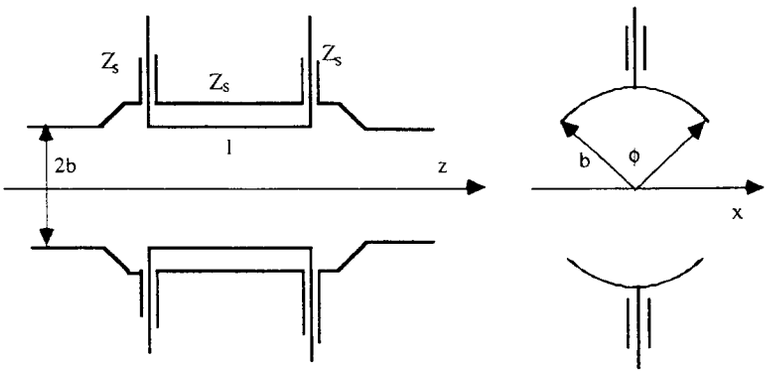


Fig. 26 Strip line pick-up

The strip plate together with the vacuum-chamber wall create a transmission line of the characteristic impedance Z_s. If the transmission line is terminated by a matching resistance Z_s at each end, one can draw the equivalent scheme Fig. 27 where $i_\phi(t)$ is the fraction of the image current intercepted by the strip plate:

$$i_\phi(t) = -\left(\frac{\phi}{2\pi}\right) i_b(t)$$

with $i_b(t)$ the beam current.

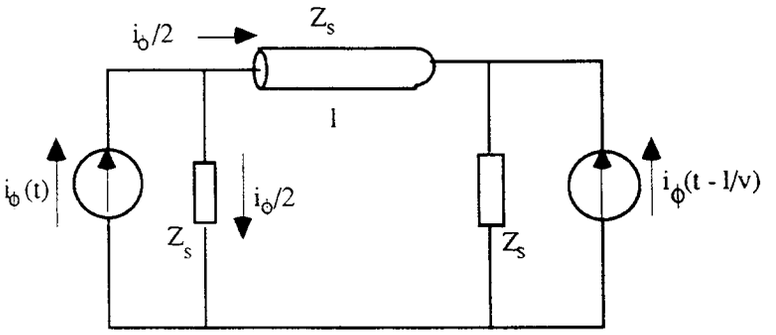


Fig. 27 Equivalent circuit for a strip line

In the absence of a dielectric the signal velocity in the transmission line is equal to that of light. Moreover, as can be seen from the equivalent scheme, the strip-line monitor possesses directional property for the relativistic particle with $\beta \sim c$. In fact the signal appears only across the upstream port (with respect to the beam velocity):

$$V(t) = \frac{Z_s}{2}\left(\frac{\phi}{2\pi}\right)\left[i_b(t) - i_b\left(t - \frac{2l}{c}\right)\right] \tag{200}$$

In turn, only a fraction $\phi / 2\pi$ of the total image current "sees" the potential difference, since the average potential seen by a beam is:

$$V_b(t) = V(t)\left(\frac{\phi}{2\pi}\right) \tag{201}$$

Applying the definition of the impedance as a response for the sinusoidal current perturbation, we have:

$$Z(\omega) = Z_s\left(\frac{\phi}{2\pi}\right)^2\left[\sin^2(\kappa l) + j\sin(\kappa l)\cos(\kappa l)\right] \tag{202}$$

The same result for the real impedance has been derived by considering the real power dissipated in the upstream termination as the power lost by a bunch [38]. The imaginary part was found by Hilbert transform.

The transverse impedance of a pair of symmetric strip-lines matched at both ends has been found in [39]. In the direction perpendicular to the strip-lines:

$$Z_\perp(\omega) = \frac{c}{b^2}\left(\frac{4}{\phi}\right)^2\left(\sin^2\frac{\phi}{2}\right)\left[\frac{Z(\omega)}{\omega}\right] \tag{203}$$

where $Z(\omega)$ is the longitudinal impedance of that pair of strip-lines. A particle shifted with respect to the system axis in the direction parallel to the strips does not experience any transverse kick and the transverse impedance in the direction $Z_\perp(\omega) = 0$.

The same method gives for a strip-line forming a transmission line with the beam pipe with characteristic impedance Z_s and terminated at the center by Z_s the following longitudinal impedance [39]:

$$Z(\omega) = Z_s\left(\frac{\phi}{2\pi}\right)^2(1 - j\kappa l)\left(1 - e^{-j\kappa l}\right) \tag{204}$$

where l is again the strip length and ϕ is the azimuthal angle characterizing the fraction of the image current intercepted by the strip plate.

The longitudinal coupling impedance of a small button-like pickup of a radius r terminated through a coaxial cable of the characteristic impedance Z_s by a resistance $R = Z_s$ is [40]:

$$Z(\omega) = \left(\frac{\omega_1}{\omega_2}\right)^2 \frac{\omega/\omega_1}{\left[1 + (\omega/\omega_1)^2\right]}\left[\frac{\omega}{\omega_1} + j\right] \tag{205}$$

with

$$\omega_1 = \frac{1}{RC} \quad \text{and} \quad \omega_2 = \frac{2bc}{r^2}$$

where b is the beam pipe radius, C is the capacitance between the button and the beam-pipe wall which is roughly estimated as: $C \sim \varepsilon_0 \pi r^2 / \delta$ for $\delta << b$ being the gap between the wall and the button.

Many examples can be found in the literature on the subject. However, we should say that the high-frequency behaviour of the diagnostics elements has not been satisfactorily investigated yet. One of the major problems in further investigations are the high-frequency resonances that can be excited in the structures formed by a diagnostics element and beam pipe walls. Some of these resonances can be associated with standing waves, which do not dissipate their power in the external terminations [41]. Obviously, the high-frequency resonances will give some additional inductive contribution to the low-frequency impedance.

7.8 Holes and slots in the vacuum chamber

In order to reduce the coupling impedance due to pumping volumes, shielding screens with a number of holes or slots are used. The number of such holes and slots may be large in an accelerator and their contribution to the impedance have to be estimated. It is obvious that because of the absence of axial symmetry a numerical solution of the problem is essentially three dimensional. This implies very time-consuming computations even in the case of a simplified model.

The method which allows analytical calculation of the impedance at low frequencies is based on the Bethe theory of diffraction by a small hole [42]. According to the theory, the small hole is excited by the incident electromagnetic waves created by a given current perturbation. Then, the diffracted fields can be obtained by replacing the hole by effective surface "magnetic" currents, which are necessary to satisfy the boundary conditions on the hole. The coupling impedance is found by integrating the fields along the beam trajectory. At low frequencies, $\omega << c / b$, in the case of a small hole of radius $h << b$ the impedance can be calculated in terms of hole polarizability [43,44]. For a circular hole it gives:

$$Z(\omega) = j\omega \frac{Z_o}{6\pi^2} \frac{h^3}{cb^2} \tag{206}$$

As far as a single hole introduces the axial asymmetry, the transverse impedance depends on the angle between the beam-offset and the direction to the hole:

$$\boldsymbol{Z}_\perp(\omega) = jZ_o \frac{2h^3}{3\pi^2 b^4} \boldsymbol{a}_r \cos\theta \tag{207}$$

where $\boldsymbol{a}_r$ is the unit vector to the hole and θ is the azimuthal angle between the direction and the beam-offset.

It is worth noting that, for the number of holes $M \geq 3$ uniformly spaced in one cross section, the restoration of the axial symmetry occurs in a sense that the transverse kick is in the direction of the beam transverse displacement, and the resulting impedance does not depend on the azimuthal positions of the holes:

$$\boldsymbol{Z}_\perp(\omega) = jZ_o \frac{h^3}{3\pi^2 b^4} M \boldsymbol{r}_1 \tag{208}$$

with $\boldsymbol{r}_1$ being the unit vector in the direction of the beam displacement. The coupling impedance of the circular hole falls with wall thickness t, reaching 56% of the value (206) for $t/h > 2$ [44].

The real part of the hole impedance, responsible for the energy losses, is usually neglected because it is much smaller than the imaginary impedance. It is given by [43]:

$$Z_r(\omega) = \frac{5Z_o}{54\pi^3}\left(\frac{\omega h}{c}\right)^4 \frac{h^2}{b^2} \tag{209}$$

The same method gives simple analytical expressions for the impedance of a small longitudinal, elongated, elliptical slot of width w and length l such as $w << l << b$:

$$Z(\omega) = j\frac{Z_o}{96\pi}\frac{\omega}{c}\frac{w^4}{b^2 l}\left(\ln\frac{4l}{w} - 1\right) \tag{210}$$

$$\mathbf{Z}_\perp(\omega) = j\frac{Z_o}{24\pi}\frac{w^4}{b^4 l}\left(\ln\frac{4l}{w} - 1\right)\boldsymbol{a}_r \tag{211}$$

Analytical results are available also for the small transverse narrow elliptical slot with $w << l << b$, cutting the wall image currents [45]:

$$Z(\omega) = j\frac{Z_o}{96\pi}\frac{\omega}{c}\frac{l^3}{b^2\left(\ln\left(\frac{4l}{w}\right) - 1\right)} \tag{212}$$

$$\mathbf{Z}_\perp(\omega) = jZ_o\frac{l^3}{24\pi b^4\left(\ln\left(\frac{4l}{w}\right) - 1\right)}\boldsymbol{a}_r\cos\theta \tag{213}$$

More discussion on the small-slot impedance calculations and estimates can be found in [46].

When the sizes of slots and holes are comparable or larger than the beam-pipe radius the static solution to the problem is no longer valid. In that case analytical solutions are known for two particular cases of a long narrow slot and a narrow gap.

It is clear from the physical point of view that the narrow ($w << b$) long ($l >> b$) slot distorts the fields inside a vacuum chamber only by the slot ends. Because of that the impedance should not depend on the length of such a slot. Indeed, according to the perturbation method [47] the impedance of a narrow, long (but finite) rectangular slot is a constant for $\omega l / c >> 1$:

$$Z(\omega) = j\frac{Z_o \pi w^2}{4(2\pi)^3 b^2} \tag{214}$$

The impedance for the case of narrow transverse gap of width $g << b$ separating two pieces of the infinite beam pipe has been derived in Ref. [48]. In the case $\omega g / c << 1$ the longitudinal impedance of the gap is capacitive:

$$Z(\omega) = -j\frac{Z_o c}{2\omega b \ln(16b/g)} \tag{215}$$

We should note here that, for a beam pipe of a general cross section, in order to reduce the coupling impedance it is preferable to distribute holes in places where fields created by a

bunch are minimal, for example, in corners [45]. In particular, for a rectangular beam pipe of width a and height b the longitudinal impedance of a hole situated on the lateral side ($x_h = \pm a/2$) and displaced by y_h from the horizontal plane of symmetry ($|y_h| < b/2$) is proportional to:

$$Z(\omega) \sim \left(\sum_{m=0}^{\infty} \frac{\cos(2m+1)\pi y_h / b}{\cosh(2m+1)\pi x_h / b} \right)^2 \tag{216}$$

As can be easily seen, the closer to the corner ($|y_h| \to b/2$) the smaller impedance we have. This is certainly valid for small holes.

Discussions on possible coherent effects of many holes can be found in [49,50]. We just mention that the coherence can be destroyed using non-uniform hole spacing.

8. BROAD-BAND IMPEDANCE MODELS

The impedance of an accelerator is usually a very complicated function of frequency with many sharp peaks. This complexity of the impedance makes an analytical treatment almost impossible. However, in the study of the single-bunch dynamics, one finds that the wake potentials over the bunch length is of main interest. This implies that, in the frequency domain, the bunch can not resolve the details of the sharp resonances and it rather experiences an average effect (the peaks are smeared out). In order to demonstrate this we calculate the wake potentials created by a bunch passing through a cavity, by means of the ABCI code.

The longitudinal impedance is found by Fourier transform of the wake potentials. As is clearly seen in Figs. 28 and 29, if one is interested in the shorter-range part of the wake the resonant impedance peaks are transformed into the smoother and broader impedance.

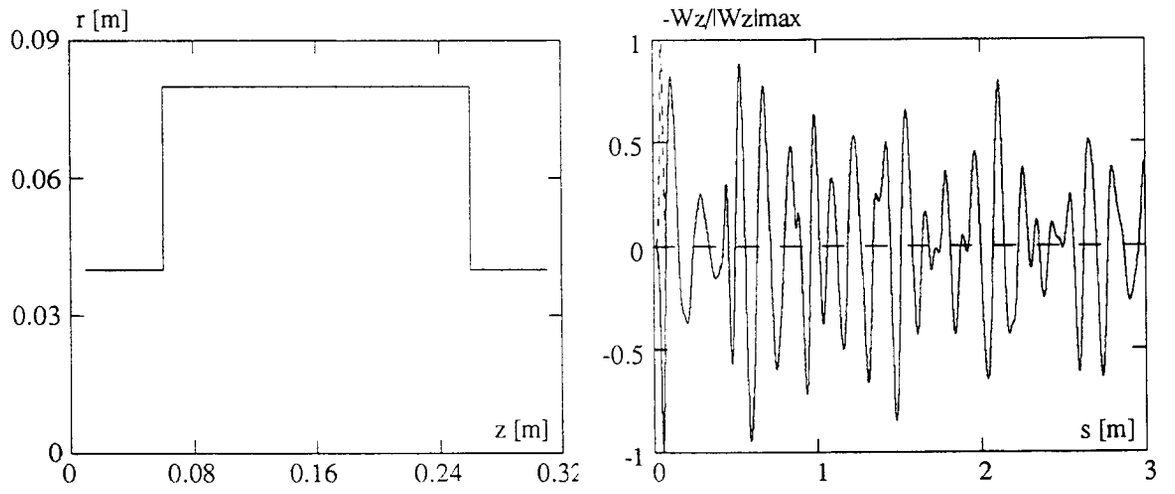


Fig. 28 Pill-box cavity with tubes and wake function for $c\sigma_\tau = 3$ cm

Therefore, the actual impedance can be replaced by some "broad-band-model impedance", which usually is characterized by a small number of parameters, allowing analytical evaluation of stability limits or growth rates of single bunch instabilities.

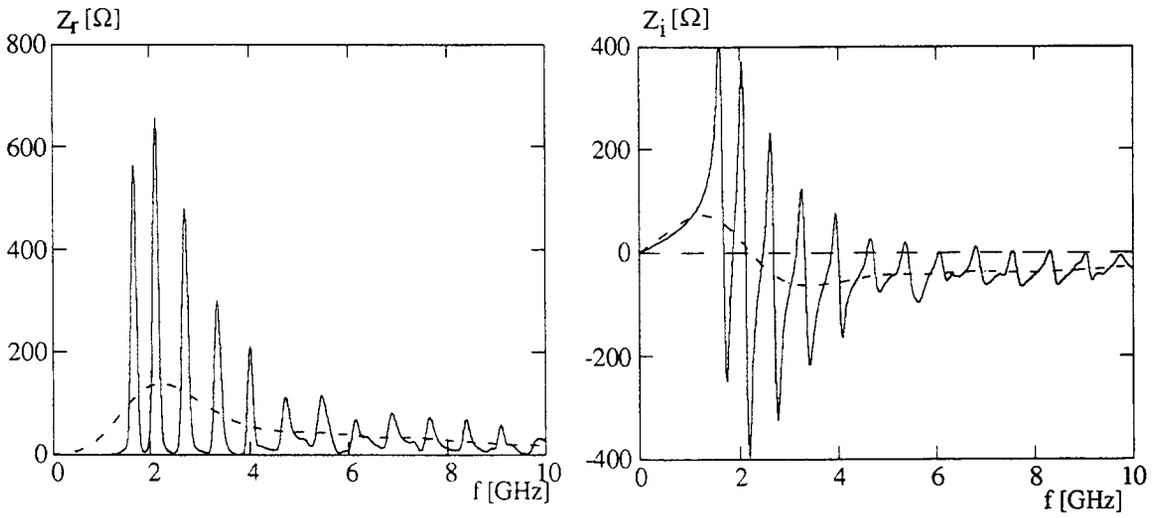


Fig. 29 Broad-band impedance: FFT of the wake function over 3 m (solid line); FFT of the wake function over 30 cm (dashed line).

8.1 Broad-band-resonator model

The reduced frequency resolution has been the main justification for using the broad-band-resonator model [51], historically being introduced first.

The actual impedance is replaced by that of a single-mode cavity with a low quality factor Q:

$$Z(\omega) = \frac{R_s}{1 + jQ\left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega}\right)} \tag{217}$$

So, only three parameters, the shunt impedance R_s, angular resonant frequency ω_r, quality factor Q, are necessary to describe the impedance frequency behaviour.

For long bunches a cavity shunt impedance R_s is estimated by averaging the resistive part of the parasitic resonance (which can be measured by means of the perturbation method, or estimated by computer codes) to give the same energy loss as the whole cavity. Each resonance contributes with only half the area of its spectrum.

The quality factor is usually taken as $Q = 1$, while the resonant frequency $\omega_r = \omega_{\text{cutoff}} = 2.4(c/b)$, i. e. frequency cut-off of the cavity iris of the radius b. These choices are somewhat arbitrary and are satisfactory only for the case of long bunches, when the whole bunch spectrum lies within the beam-pipe cut-off. As an example, let us consider again the cavity of Fig. 28. We find that the effective impedance calculated as the Fourier transform of the wake over 30 cm distance behind the bunch head is approximated reasonably well by the broad-band-resonator impedance with $R_s = 138\ \Omega$, $Q = 1$ and $f_r = 2.2$ GHz (which is close to the $f_{cut\text{-}off} = 2.5$ GHz of the considered cavity) (Fig. 30).

The parameters of the broad-band resonator can be found by measurements, for example, by measuring bunch lengthening in an accelerator. At the project stage these three parameters can be evaluated by comparing the loss factor dependence on the bunch length K or wake potentials along the bunch $W_z(\tau)$, calculated numerically, with those corresponding to the broad-band resonator.

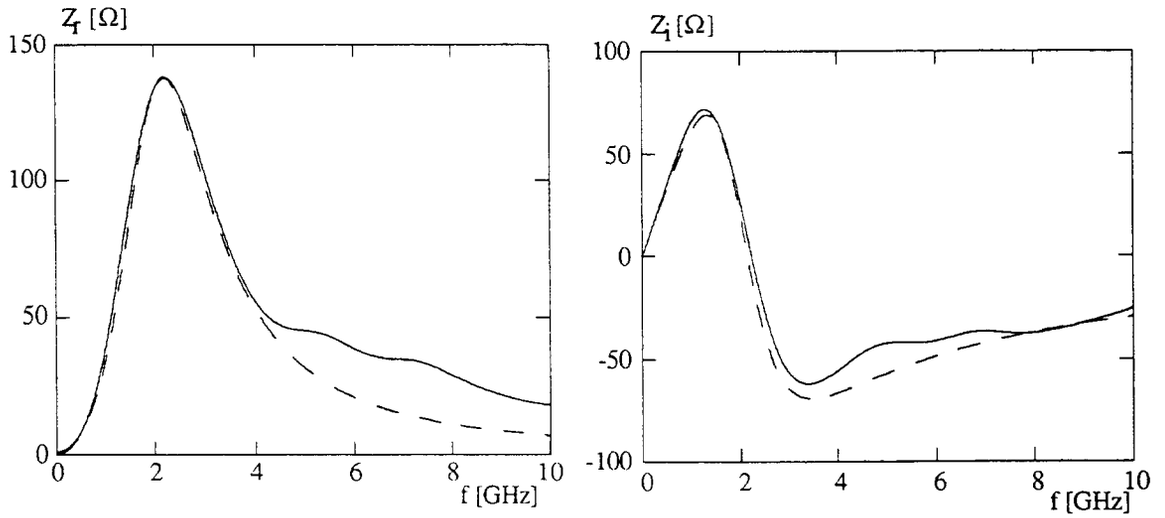


Fig. 30 FFT of the wake function over 30 cm (solid). Broad-band-resonator impedance (dashed)

For the example in Fig. 30 we got $Q = 1$, $R = 138\Omega$ and $f_r = 2.2$ GHz. However, the broad-band resonator model has some disadvantages which can become important for the extremely short or very long bunches.

Asymptotic frequency dependence of the broad-band-resonator impedance is ω^{-2}, while analytical study shows that the real part of the impedance decreases with frequency as $\omega^{-1/2}$ for a cavity with attached tubes and as $\omega^{-3/2}$ for an array of periodic cavities. This leads to incorrect energy loss estimate in the model, especially for very short bunches when the main contribution to the total energy loss is given by the high-frequency tail of the longitudinal impedance.

For very long bunches, according to the broad-band-resonator model, the loss factor decreases with the third power of the bunch length, while numerical calculations with the time domain codes shows that the loss factor of realistic vacuum-chamber elements drops much more quickly, in an exponential fashion, with increasing σ for long bunches with Gaussian distribution.

8.2 Hofmann-Zotter impedance model

Two pairs of the broad-band-impedance models were proposed by A. Hofmann and B. Zotter [52] to overcome the limitations of the broad band resonator model in the high-frequency and low-frequency regions. Both can be adjusted to have either $\omega^{-1/2}$ or $\omega^{-3/2}$ asymptotic behaviour of the real part of the impedance at high frequencies. In the low-frequency region, for the first pair of the impedance models the real part increases as ω^2 and, for the special choice of the model parameters, as ω^4 providing faster decrease of the loss factor with bunch length (σ^{-5}) than in the broad-band-resonator model. The second improved impedance model yields even more rapid decrease of the loss factor with bunch length, having an exponential character for very long bunches.

8.2.1 The first improved impedance model

The expression for the impedance model 1a) is:

$$Z(x) = R\left[\frac{1}{u}\sqrt{\frac{2}{u+1}} - \frac{\alpha^2}{x^2+\alpha^2}\right] + jR\left[\frac{1}{u}\sqrt{\frac{2}{u-1}} + \frac{\alpha^2}{x^2+\alpha^2} - \frac{2}{x}\right] \tag{218}$$

where:

$$x = \frac{\omega}{\omega_1}, \quad u = \sqrt{x^2+1}, \quad \alpha = \frac{\omega_2}{\omega_1} \tag{219}$$

The frequencies ω_1, ω_2 and the shunt impedance R are chosen to fit the impedance of a particular structure. For very high frequencies, the asymptotic behaviour of the impedance is given by:

$$Z_r(x) \sim x^{-3/2}, \quad Z_i(x) \sim \frac{2-\alpha}{x} \tag{220}$$

At low frequencies:

$$Z_r(x) \sim \left(\frac{1}{\alpha^2} - \frac{5}{8}\right)x^2, \quad Z_i(x) \sim \left(\frac{1}{\alpha} - \frac{3}{4}\right)x \tag{221}$$

For a special choice $\alpha = \sqrt{8}/5$ the real part of the impedance increases as ω^4 providing a fast decrease of the loss factor with bunch length $\sim \sigma^{-5}$. In order to provide the high frequency-asymptotic behaviour $\omega^{1/2}$ a different model was proposed:

$$Z(x) = R\left[\frac{\sqrt{u^2+1}}{u\sqrt{2}} - \frac{\alpha^2}{x^2+\alpha^2}\right] - jR\left[\frac{\sqrt{u-1}}{u\sqrt{2}} - \frac{\alpha x}{x^2+\alpha^2}\right] \tag{222}$$

with:

$$x = \frac{\omega_1}{\omega}, \quad u = \sqrt{x^2+1}, \quad \alpha = \frac{\omega_2}{\omega_1} \tag{223}$$

At low frequencies the impedance has the following behaviour:

$$Z_r(x) \sim \left(\frac{1}{\alpha^2} - \frac{3}{8}\right)x^2, \quad Z_i(x) \sim \left(\frac{1}{\alpha} - \frac{1}{2}\right)x \tag{224}$$

8.2.2 The second improved impedance model

While the first impedance model describes well high frequency behaviour, it still can not supply an exponential fall-off of the loss factor with bunch length for long Gaussian bunches. In order to obtain this the real part of the impedance should vanish completely below a certain "cut-off" frequency (as in a realistic structure where the impedance is zero below the lowest mode). The second model impedances are chosen to satisfy this condition.

In one of these models, 2a, the real impedance is given by the expression:

$$Z_r(x) = R\frac{\sqrt{|x|-1}}{x^2}; \quad |x| > 1$$
$$Z_r(x) = 0; \quad |x| < 1 \tag{225}$$

It has a maximum at $x = \omega/\omega_1 = 4/3$ and becomes proportional to $\omega^{-3/2}$ at high frequencies. The Hilbert transform gives the impedance imaginary part:

$$Z_i(x) = \frac{R}{x^2}\left[\sqrt{1+x} - \sqrt{1-x} - x\right]; \quad |x| < 1$$
$$Z_i(x) = \frac{R}{x^2}\left[\sqrt{1+x} - x\right]; \quad |x| > 1 \tag{226}$$

A similar impedance model (2b) has an asymptotic decrease of the real part proportional to the inverse-square-root frequency at high frequencies. The real part is given by:

$$Z_r(x) = R\frac{\sqrt{|x|-1}}{|x|}; \quad |x| > 1$$
$$Z_r(x) = 0; \quad |x| < 1 \tag{227}$$

which has a maximum at $x = 2$. The expression for the imaginary part is:

$$Z_i(x) = \frac{R}{x}\left[2 - \sqrt{1+x} - \sqrt{1-x}\right]; \quad |x| < 1$$
$$Z_i(x) = \frac{R}{x}\left[2 - \sqrt{1+x}\right]; \quad |x| > 1 \tag{228}$$

The loss factor for both models, 2a and 2b, can be easily obtained by numerical integration. It drops exponentially with bunch length for the two models for long Gaussian bunches.

8.3 Heifets-Bane impedance model

Recently, a new broad-band impedance model was proposed by S. Heifets [53] as the further development of K. Bane's approach used in his analysis of the impedance of the SLC damping ring [54,55]. The longitudinal impedance is described phenomenologically by expansion over $\sqrt{\omega}$:

$$Z(\omega) = j\omega L + R + (1 + j\,\mathrm{sign}(\omega))\sqrt{|\omega|}B + \frac{1 - j\,\mathrm{sign}(\omega)}{\sqrt{|\omega|}}\tilde{Z}_c + \ldots \tag{229}$$

This model has been used to estimate the broad-band impedance for the SLAC B-factory and for the DAΦNE main rings (Φ-factory) [56]. Such an impedance model has two attractive features:

- First, the different terms of the expansion have a clear physical interpretation, describing correctly particular impedance-generating elements.
- Second, expressions for wake-fields and loss factors can be easily found analytically, simplifying the fitting procedure to extract the model parameters.

The first term of the expansion represents a low frequency inductive impedance. This impedance is typical for tapers, shielded bellows and vacuum ports, small discontinuities such as slots, shallow cavities in flanges, shallow collimators and so on. Often, these elements give the main contribution to the impedance, leading to the excess bunch lengthening as in the case

of SLC damping rings. For the small discontinuities the impedance remains inductive up to rather high frequencies.

The wake function for a Gaussian bunch corresponding to the *inductive* term is given by:

$$W_z(\tau) = -\frac{L\tau}{\sqrt{2\pi}\sigma_\tau^3}\exp\left\{-\frac{1}{2}\left(\frac{\tau}{\sigma_\tau}\right)^2\right\} \tag{230}$$

It has a minimum (maximum) at $\tau = -\sigma_\tau(+\sigma_\tau)$:

$$W_{\max} = -W_{\min} = \frac{L}{\sqrt{2\pi e}\sigma_\tau^2} \tag{231}$$

The plot of $-W_z(\tau)/|W_z|_{\max}$ which is suitable to compare with TBCI [57] or ABCI [25] code results for azimuthally symmetric structures or MAFIA [58] for 3D structures is shown in Fig. 31.

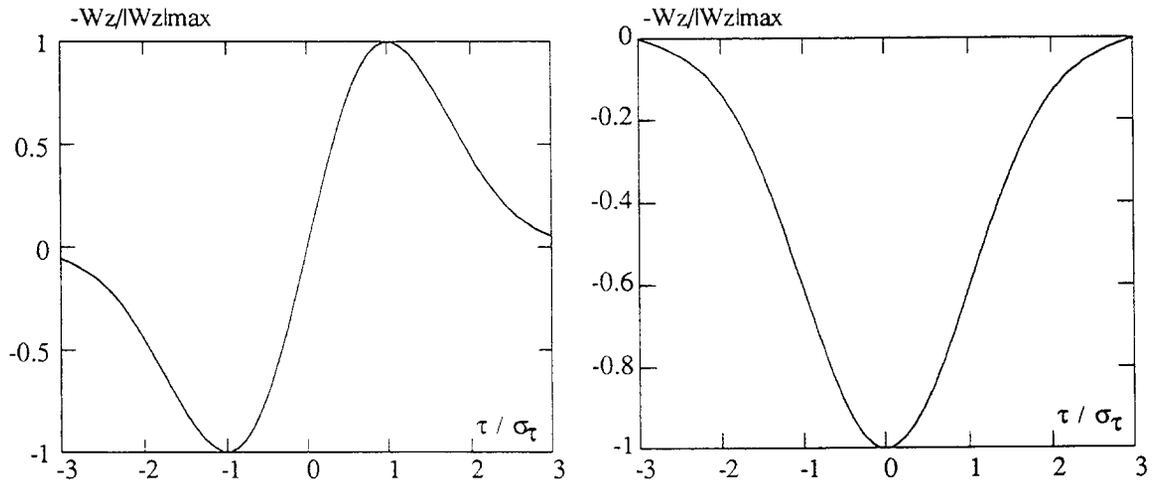


Fig. 31 Normalized inductive wake functions: inductive and resistive

Equation (231) makes it straightforward to extract the inductance L from numerical results. The loss factor for the inductive impedance is zero. An example of a shallow cavity and corresponding wake are shown in Fig. 32.

The wake-field for the *resistive* term $Z(\omega) = R$ is:

$$W_z(\tau) = -\frac{R}{\sqrt{2\pi}\sigma_\tau}\exp\left\{-\frac{1}{2}\left(\frac{\tau}{\sigma_\tau}\right)^2\right\} \tag{232}$$

The ratio $-W_z(\tau)/|W_z|_{\max}$ is shown in Fig. 33. In storage rings such a wake is expected for deep cavities for bunches with length comparable with the beam-pipe radius. Figure 33 reproduces an example of a wake for the SLC damping ring cavity [55].

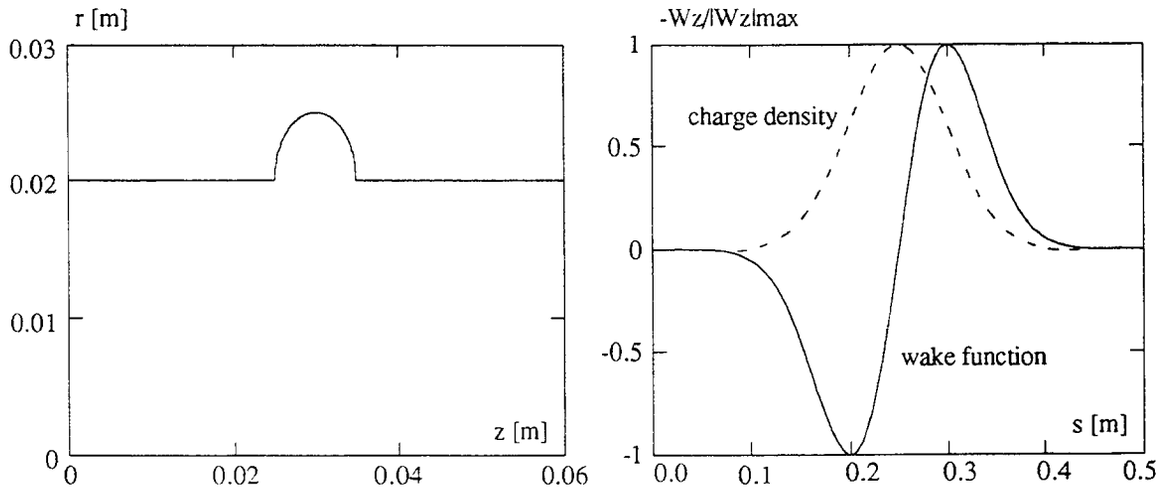


Fig. 32 Shallow cavity and wake function for $c\sigma_\tau = 5$ cm

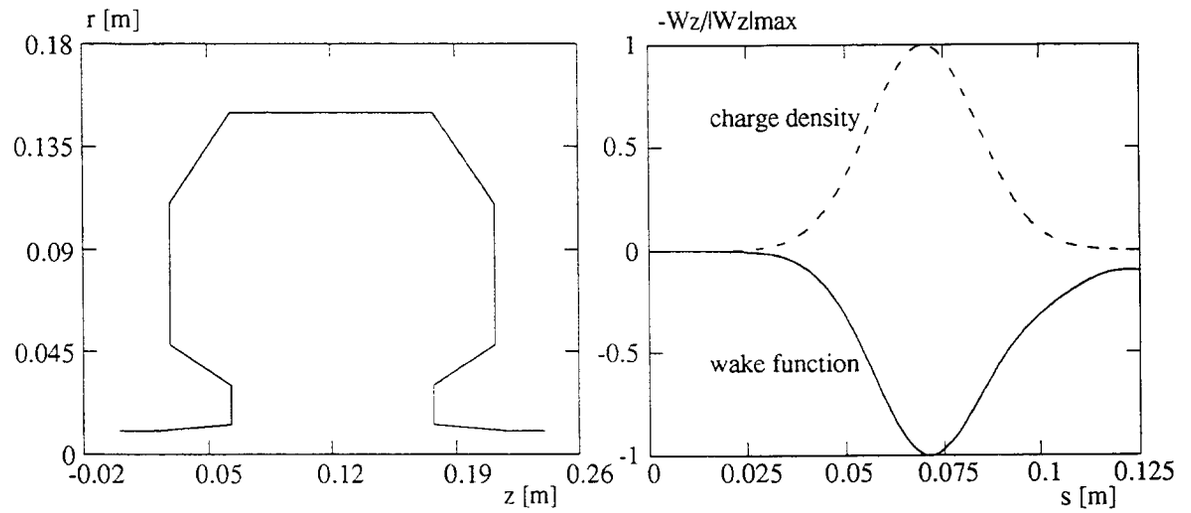


Fig. 33 Deep cavity and wake function for $c\sigma\tau = 1.4$ cm

For good resistors the expression for the loss factor has a simple form:

$$K(\sigma_\tau) = \frac{R}{2\sqrt{\pi}\sigma_\tau} \tag{233}$$

The third term has the structure of the resistive-wall impedance with the following wake-function (see Fig. 34 a):

$$W_z(\tau) = \frac{B|\tau|^{3/2}}{4\sigma_\tau^3}\left\{I_{-3/4}(\eta) - I_{1/4}(\eta) - I_{-1/4}(\eta) \pm I_{3/4}(\eta)\right\}e^{-\eta} \tag{234}$$

where $\eta = \left(\frac{\tau}{2\sigma_\tau}\right)^2$ and the $\pm$ sign corresponds to positive and negative "τ", respectively. I_ν are the modified Bessel functions of fractional order. Applying the loss factor definition we have:

$$K(\sigma_\tau) = \frac{B}{\pi \sigma_\tau^{3/2}} \frac{\Gamma(3/4)}{2} \tag{235}$$

where $\Gamma(3/4)/2 = 0.6127....$

The fourth term in (229) has the same dependence on ω as the impedance of a cavity with attached tubes at high frequencies:

$$Z(\omega) = \frac{(1 - j\,\mathrm{sign}(\omega))}{\sqrt{|\omega|}} \tilde{Z}_c \tag{236}$$

The wake-function corresponding to the impedance is given by:

$$W_z(\tau) = \frac{\tilde{Z}_c}{2\sqrt{\sigma_\tau}} \sqrt{\frac{|\tau|}{\sigma_\tau}} \left\{ I_{-1/4}(\eta) \pm I_{1/4}(\eta) \right\} e^{-\eta} \tag{237}$$

where the $\pm$ sign stands for positive and negative "τ". The function $-W_z(\tau)/|W_z|_{max}$ is shown in Fig. 34b.

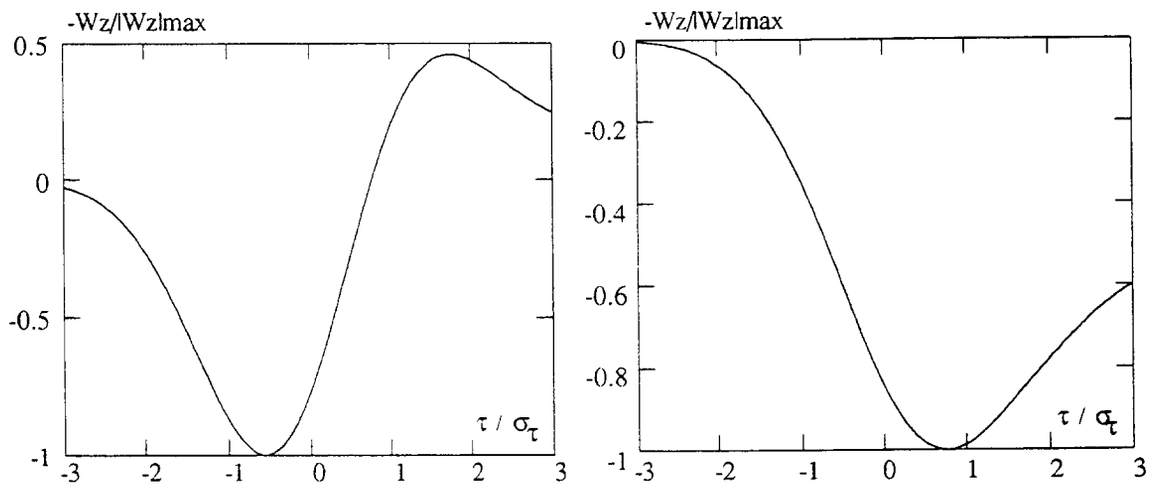


Fig. 34 Normalized wake functions corresponding to the cases:

left) $(1+j\,\mathrm{sign}(\omega))\sqrt{|\omega|}B$ right) $\frac{1-j\,\mathrm{sign}(\omega)}{\sqrt{|\omega|}} \tilde{Z}_c$

Using (26) we obtain the expression for the loss factor:

$$K(\sigma_\tau) = \frac{B}{\pi \sigma_\tau^{3/2}} \frac{\Gamma(3/4)}{2} \tag{238}$$

with $\Gamma(1/4)/2 = 1.8128...$

The parameters L, R, B, Zc of the broad-band model are extracted from TBCI, ABCI or MAFIA results by fitting the numerical functions $W(\tau)$ and $K(\sigma_\tau)$ to the above analytical expressions. In Fig. 35 we show an example of a wake function for a cavity with attached tubes where the high frequency behaviour is dominant.

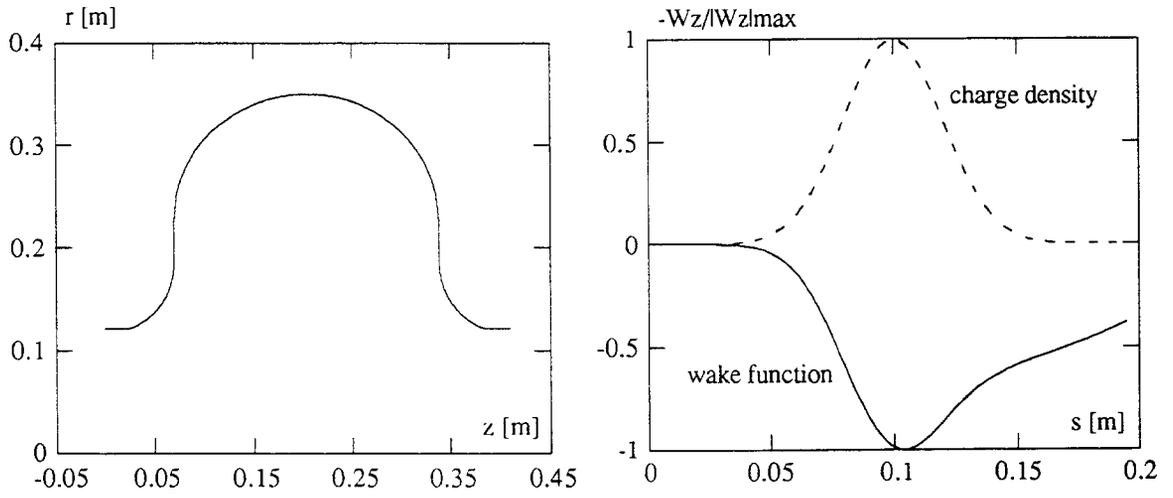


Fig. 35 Cavity with attached tubes and wake function for $c\sigma_\tau = 2$ cm

ACKNOWLEDGMENTS

We wish to thank M. Serio for some enlightening comments and suggestions.

REFERENCES

[1] "Computer Codes for Particle Accelerator Design and Analysis: A compendium", LA-UR-90-1766 (1990).
[2] L. Palumbo, V.G Vaccaro, "Wake Fields Measurements", 1988 Joint US-CERN Accelerator School, Lecture Notes in Physics, Springer Verlag Vol. 343, M. Month and S. Turner Eds. (1989).
[3] P.B. Wilson, SLAC PUB 2884 (1982).
[4] A.W. Chao, SLAC PUB-2946 (1982).
[5] K Bane, T. Weiland, P. Wilson SLAC PUB 3528 (1984).
[6] T. Weiland, "On the Computation of Resonant Modes in Cylindrical Symmetric Cavities," Nuclear Instrument and Methods Vol. 216, 329-349, (1983).
[7] T. Weiland, R. Wanzemberg, "Wake Fields and Impedances", DESY M-91-06, (1991).
[8] V.G. Vaccaro, CERN ISR-RF/66-35 (1966).
[9] W.K.H. Panofsky, W.A. Wenzel, "Transverse Deflection of Charged particles in Radio-Frequency Fields", Rev. Sci. Instrument, Vol 27 (1956).
[10] L. Palumbo, V.G. Vaccaro "Wake Fields, Impedance and Green's Function", CAS Advanced School, Oxford 1986, CERN 87-03 (1987).
[11] T. Weiland, "Transverse Beam Cavity Interaction", Nuclear Instruments and Methods, Vol. 212, 13 (1983).
[12] J.D. Jackson, "Classical Electrodynamics", Wiley NY (1975).
[13] L. Palumbo, V.G. Vaccaro, G. Wustefeld, "Coupling Impedance in Circular Accelerators, Circular Beam, Elliptic Chamber", IEEE Trans, NS-31 (1984).
[14] P.L. Morton, V.K. Neil, A.M. Sessler, JAP 37, 3875 (1966).
[15] L. Palumbo, V.G. Vaccaro, "Coupling Impedance between Circular Beam and a Lossy Vacuum chamber in particle Accelerators", Il Nuovo Cimento Vol. 89 (1985).
[16] R.L. Gluckstern, J. van Zeijts, B. Zotter, "Coupling Impedance of Beam Pipes of General Cross Section", CERN SL/AP 92-25 (1992).
[17] K. Yokoya, "Resistive Wall Impedance of Beam Pipes of General Cross Section", KEK preprints 93-196 (1993).
[18] S.A. Heifets, S.A. Kheifets, "Coupling Impedances in Modern Accelerators" Review of

Modern Physics, Vol. 63, No. 3 (1991).
[19] E. Gianfelice, G. Dôme, L. Palumbo, V.G. Vaccaro, L. Verolino, "A canonical Problem for the Understanding of Energy Diffraction Losses in High Energy Accelerators", Il Nuovo Cimento Vol. 104 A (1991).
[20] G. Dôme, L. Palumbo, V.G. Vaccaro, L. Verolino, "Longitudinal Impedance of a Circular Iris", Il Nuovo Cimento, Vol. A (1991).
[21] V.E. Bakalin, A.V. Novokhatsky, Proceedings of 12th Int. Conf. on High Energy Accelerators, Batavia, Illinois, F.T. Cole and R. Donaldson Eds. (1983).
[22] S.A. Kheifets and S.A. Heifets, "Radiation of a Charge in a Perfectly Conducting Cylndrical Pipe with a Jump in its Cross section", SLAC-PUB-3965 (1986).
[23] L. Palumbo, "Analytical Calculation of the Impedance of a Discontinuity", Particle Accelerators Vol. 25 (1990).
[24] E. Gianfelice, L. Palumbo, "Dipole Impedance of a Conducting Cylindrical Pipe with a Discontinuous Cross Section", IEEE Trans, NS, Vol. 37 (1990).
[25] Y.H. Chin, "User Guide for New ABCI", LBL 33091, CERN SL/92-49 (AP) (1992).
[26] S.A. Heifets, "Diffractive Model of High Frequency Impedance", Phy. Rev. D40 (1989).
[27] J.D. Lawson, "Radiation from a Ring Charge Passing Through a Resonator", Rutherford High Energy Laboratory Report RHEL/M144 (1968).
[28] G.D. Dôme, "Wake Potential of Relativistic Point Charge Crossing a Beam Pipe Gap: An Analytic Approach", IEEE Trans. Nucl. Sci. NS-23 (1985).
[29] K. Bane, M. Sands, "Wake-fields of Very Short Bunches in an Accelerating Cavity", SLAC PUB-4441 (1987).
[30] R. Palmer, "A Qualitative Study of Wakefields for Very Short Bunches", SLAC PUB 4433 (1987)
[31] V. Balakin and A. Novokhatsky, Proceedings of the 13th Conference on HEA, Novosibirsk (1986).
[32] R.L. Gluckstern, "Longitudinal Impedance of a Periodic Structure at High Frequency", University of Maryland (1988), also Phys. Rev. D39 (1989).
[33] S.A. Heifets and S.A. Kheifets, "High Frequency Limit of the Longitudinal Impedance of an Array of Cavities", SLAC PUB-4625, 1988 and Phys. Rev. D39 (1989).
[34] G. Dome, L. Palumbo, V.G. Vaccaro, L. Verolino, "A Method for Computing the Longitudinal Coupling Impedance of Circular Apertures in a Periodic Array of Infinite Planes", Particle Accelerators, Vol. 36 (1991).
[35] G. Dome, L. Palumbo et al. "The Frequency Behaviour of the Longitudinal Coupling Impedance of a Periodic Array of Diaphragms", EPAC Conference, Berlin (1992).
[36] S.S. Kurennoy, G.V. Stupakov, "A New Method for Calculation of Low-Frequency Coupling Impedance", SSC-preprint-332 (1993).
[37] K.-Y. Ng, Preprint FNAL FN-389, Batavia (1983).
[38] R.E. Shafer, "On the Coupling Impedance of Position Monitor Electrodes", IEEE Trans. on Nuclear Science, Vol. NS-32 (1985).
[39] K.-Y. Ng, "Impedances of Stripline Beam-Position Monitors", Particle Accelerators, 1988, Vol. 23, pp. 93-102.
[40] M. Serio, private communications.
[41] K.-Y. Ng, "Longitudinal Instabilities and Stability Criteria", In AIP Conference Proceedings 184 - NY, (1989).
[42] H.A. Bethe, Phys. Rev., V. 66, p. 163-182, (1944).
[43] S. Kurennoy, "Coupling Impedance of Pumping Holes", Particle Accelerators, V.39, No.1 (1992).
[44] R.L. Gluckstern, "Coupling Impedance of a Single Hole in a Thick Wall Beam Pipe",CERN SL/92-05 (AP) (1992).
[45] S. Kurennoy, "Beam Coupling Impedance of Holes in Vacuum Chamber Wall", IHEP Preprint 92-84, Protvino, 1992 (in Russian).
[46] S. Kurennoy, "Pumping Slots: Coupling Impedance Calculations and Estimates", SSCL-Preprint-636 (1993).
[47] S. Heifets, "Broadband Impedances of Accelerating Structures: Perturbation Theory", Particle Accelerators, V 42, No. 3-4 (1993).

[48] S. Kurennoy, "On Coupling Impedance of Pumping Holes", SSCL-Preprint-312, April (1993).
[49] R.L Gluckstern, "Coupling Impedance of Many Holes in a Liner within a Beam Pipe", CERN SL/92-06 (AP) (1992).
[50] M. Sands, "Energy Loss from Small Holes in the Vacuum Chamber", PEP-253 (1977).
[51] A. Hofman, J. Maidment, "Current Dependent Phenomena in LEP", CERN LEP note 169 (1979).
[52] A. Hofman, B. Zotter, "Improved Impedance Models for High Energy Accelerators and Storage Rings", CERN-LEP/TM-88-51 (1988).
[53] S. Heifets, "Broad Band Impedance of the B-Factory", SLAC/AP-93 (1992).
[54] K. Bane, "The Calculated Longitudinal Impedance of the SLC Damping Ring", SLAC-PUB-4618 (1988).
[55] K. Bane, "Bunch Lengthening in the SLC Damping Rings", SLAC-PUB-5177 (1990).
[56] S. Bartalucci et al., "Broad Band Model Impedance for DAΦNE Main Rings", Nuclear Instruments and Methods in Nuclear Research, A337 (1994).
[57] T. Weiland, Part. Accel. 15, 245, 292 (1984)
[58] R. Klatt, T. Weiland, SLAC Report 303 (1986).

CURES OF INSTABILITIES

D. Boussard
CERN, Geneva, Switzerland

Abstract
Careful design of machine elements, in particular the RF cavities, may be sufficient to avoid beam instabilities in circular accelerators, with the help of natural synchrotron radiation damping (e^+e^- machines) or Landau damping in the case of hadron beams. If this is not sufficient feedback systems, transverse or longitudinal, have to be employed. A review of the various techniques in use against beam instabilities induced by long-range wake fields is given in this paper.

1. INTRODUCTION

Accelerators are often plagued by various sorts of instabilities which limit their maximum beam currents. In the case of circular machines, which we consider here, one can distinguish between strong, turbulent instabilities driven by short range wake fields (the microwave instability, the transverse mode coupling instability are good examples) and weak multiturn instabilities which are, in most cases, due to long range fields. Against turbulent instabilities one can only consider "preventive" measures aimed at raising the threshold beam currents (the notable exception being the use of reactive feedback against mode coupling transverse instabilities, which we will not discuss here). Preventive measures comprise the minimization of the broad band impedance of the vacuum chamber by screening any abrupt cross section change (bellows, vacuum manifolds, etc...) and by a proper design of protruding elements like pick up and clearing electrodes.

Strong, turbulent instabilities must be avoided by a careful choice of parameters for any new machine design (the same is true also for instabilities with high, within bunch mode, numbers which are difficult to damp by active feedback).

In the following we shall concentrate on "slow" instabilities, for which effective "countermeasures" can be engineered.

2. DAMPING OF NARROW BAND IMPEDANCES

Most slow instabilities are driven by long-memory, narrow-band impedances (this is not the case however for the head-tail instability, transverse or longitudinal). In many cases, especially for e^+e^- machines which require a large RF voltage, the major source of narrow-band impedances are the higher-order modes (HOM's) of the accelerating cavities.

The problem here is to damp as much as possible all HOM's, without affecting the fundamental, accelerating mode. Higher-order mode dampers are essentially high-pass devices whose corner frequency is located above the fundamental and below the first HOM frequencies. To improve the rejection of the accelerating mode a notch filter is often combined with the HOM coupler. In this case one must pay attention to the high RF currents at the RF frequency which flow in the notch filter elements. A simple form of high-pass filter is a waveguide with a cut off frequency above the fundamental resonance of the cavity. The position of the waveguide(s) in the cavity can be optimized such as to couple to as many HOM's as possible [1]. Figures 1 and 2 shows an example of such a design with several waveguides attached to a copper cavity. Note the long tapers on either side of the cavity itself to minimize the contribution of the RF cavity to the broad band impedance of the ring (DAFNE) [2]. The waveguides can also be the vacuum chamber itself, as in Fig. 3, which

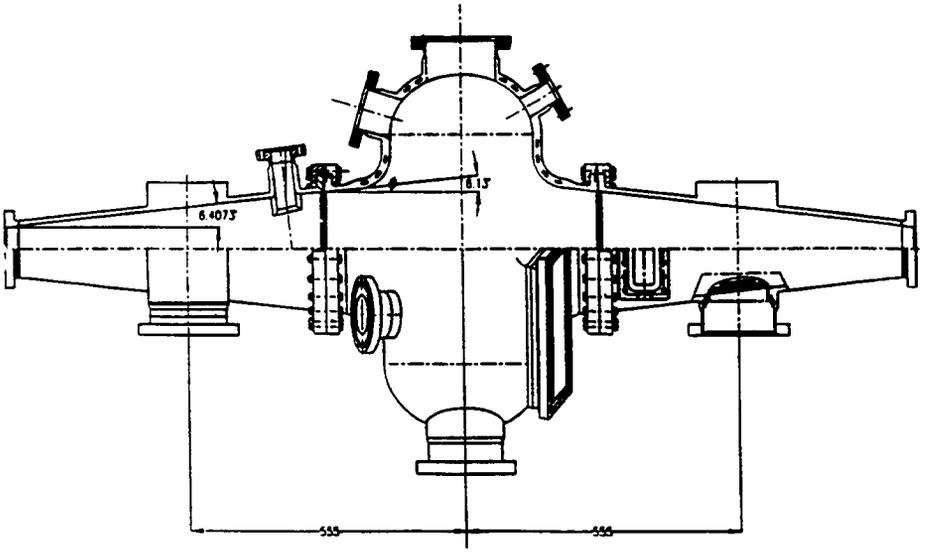


Fig. 1 Sketch of the DAFNE cavity with waveguide ports for HOM damping

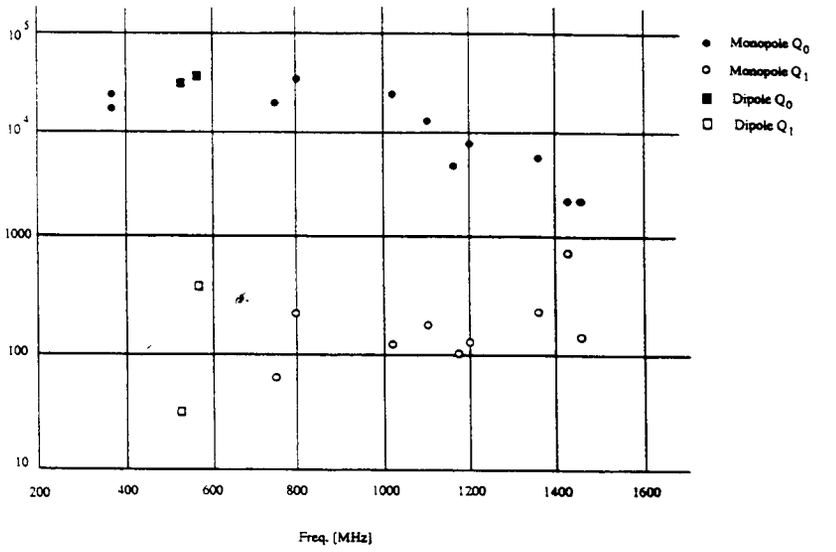


Fig. 2 Undamped and damped Q values of the lowest HOM's of the DAFNE cavity

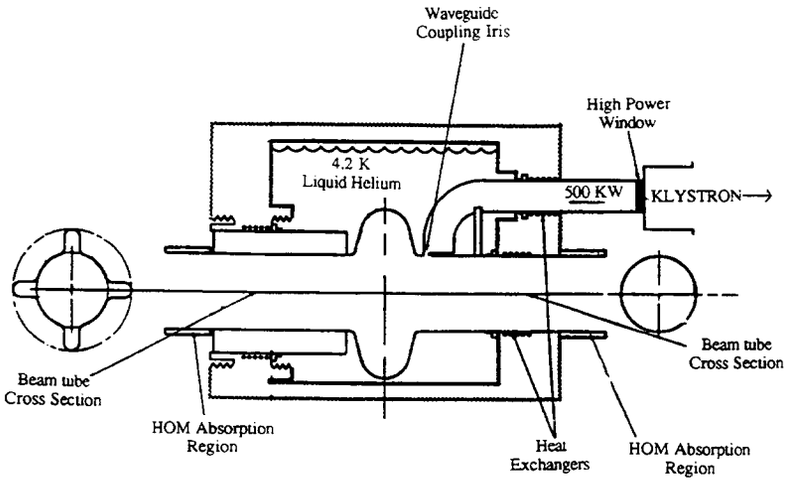


Fig. 3 Sketch of the CESR B superconducting cavity

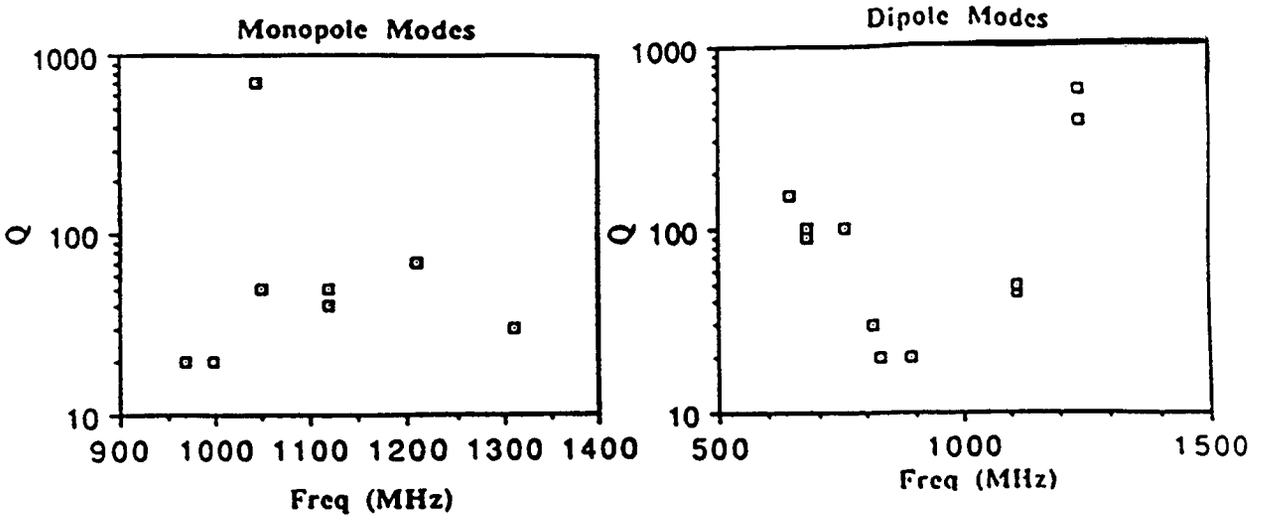


Fig. 4 Damped Q values for the CESRB cavity

shows the CESRB superconducting cavity design [3]. The so-called "fluted" vacuum chamber shape gives a lower cut off frequency as compared to a cylindrical waveguide and let also the first two transverse modes propagate outside the cryostat. In this case damping of the HOM resonators is accomplished with lossy material (ferrites) inside the vacuum chamber. With conventional waveguides, the damping elements can be inside vacuum, as waveguide terminations, in the case of small power dissipation (CEBAF, for instance). They can also be outside vacuum in coaxial form, if broad-band waveguide-coaxial transitions and coaxial vacuum feedthroughs are employed [4]. Figures 2 and 4 show the results of HOM damping for the two previous examples and illustrate the present state of the art (Q_{ext} of the order of 100 or less).

Many other damping techniques are in use, or have been proposed, notably tuned antennas to damp a particular mode in the cavity. For low-frequency cavities damping elements having a high pass characteristic can be inserted in "series" with the accelerating gap to provide damping of all longitudinal modes [5].

3. DAMPING MECHANISMS AND THRESHOLDS

3.1 Synchrotron radiation

In electron or positron machines, synchrotron radiation provides a natural damping to individual particle oscillations, and thus to any collective bunch motion. The instability threshold is given simply by the condition:

$$\Delta\omega_{mn} < 1/\tau_r$$

where $\Delta\omega_{mn}$ is the imaginary part of the frequency shift due to the instability and τ_r the radiation damping time, either the energy damping time if one considers longitudinal instabilities or the horizontal, vertical damping times in the transverse case.

For the longitudinal case, Table I gives typical numbers for two very different machines (the small 500 MeV DAFNE ring, and the very large ring of LEP). Even in the latter case

where synchrotron radiation is very strong, natural damping of longitudinal oscillations is fairly weak ($\Delta\omega_s / \omega_s \approx 10^{-3}$).

Table 1
Some parameters of two very different machines

Machine	Revolution frequency f_o (Hz)	Synchrotron time Q_s	Energy damping time τ_z(s)	Relative synchrotron frequency shift $\Delta\omega_s / \omega_s = 1/\tau_z\omega_s$
DAFNE	3.07 10^6	0.0123	17.8 10^{-3}	2.32 10^{-4}
LEP (45 GeV)	11246	0.085	17.81 10^{-3}	9.3 10^{-4}

3.2 Landau damping

There is always a spread of individual particles' frequencies in a bunch population, which is due to, for instance the non-linearity of the restoring force. This is notably the case for the synchrotron oscillations with a sinusoidal RF voltage or in the case of transverse oscillations when octupolar fields are present.

The obvious consequence of a spread of oscillation frequencies is that any coherent motion will ultimately result in a macroscopic increase of the beam emittance, a process often called filamentation. This has naturally a self stabilizing effect on any instability as the coherent oscillation resulting from the instability is transformed into an emittance increase which usually corresponds to a larger frequency spread. Ultimately the beam may become stable when the frequency spread is large enough for Landau damping to be effective.

Unfortunately this is a process difficult to predict (competition between instability growth and filamentation), it very often leads to unacceptable losses and in any case produces beam blow-up which may be undesirable. One has to mention, however, that many machines operate in this way; this is the case if beam losses are small enough and if emittance blow-up is not of great concern.

The spread of individual particles' oscillation frequencies also prevents instabilities developing below a certain threshold, a process improperly called Landau "damping". There is no damping as such, as in the case of synchrotron radiation, rather the instability cannot grow from initially microscopic disturbances. The threshold below which the beam stays stable depends on the details of the distribution of individual frequencies and the type of instability (mode number, etc...).

In the longitudinal plane, for the dipole mode and a smooth distribution, the approximate Landau damping criterion can be written [6]:

$$\frac{|\Delta\omega_s|}{\omega_s} < \frac{1}{4}\,\frac{s}{\omega_s} \qquad (1)$$

where $|\Delta\omega_s|$ is the magnitude of the instability frequency shift, and s/ω_s the relative synchrotron frequency spread inside the bunch.

For a non-accelerating bucket $(\phi_s = 0)$ one has, for a sinusoidal RF voltage:

$$\frac{s}{\omega_s} \approx \frac{\pi^2}{16}\left(\frac{\text{bunch length}}{\text{RF period}}\right)^2 \tag{2}$$

Landau damping is especially efficient for long bunches, usually the case with protons or heavy particles, for which synchrotron radiation is non existent. For non-accelerating buckets $(\phi_s = 0)$ a possible cure [7] of coupled bunch longitudinal instabilities is to reduce the RF voltage V; this increases dramatically s/ω_s when the beam edge approaches the separatrix. For a stationary bucket this method does not work because the instability growth rate $(\Delta\omega_S)/\omega_S$, proportional to 1/V, increases faster than s/ω_s.

An illustrative picture of Landau damping is given on Fig. 5 where the distribution of individual particles' oscillation frequencies is shown. For the dipole mode, the zero intensity coherent frequency ω_c is located inside the distribution (in the case of the so-called waterbag model, with uniform particle density in phase space up to a limit amplitude, and zero beyond, ω_c is just on the edge of the distribution). When the beam current is increased, the coherent frequency ω_c is shifted by an amount $\Delta\omega_c$, usually complex. Landau damping is effective if the new value ω'_c is still inside the initial distribution of incoherent frequencies. This is a condition similar to Eq. (1) except for the numerical factor which depends on the particular case of interest (mode number, type of distribution, type of impedance).

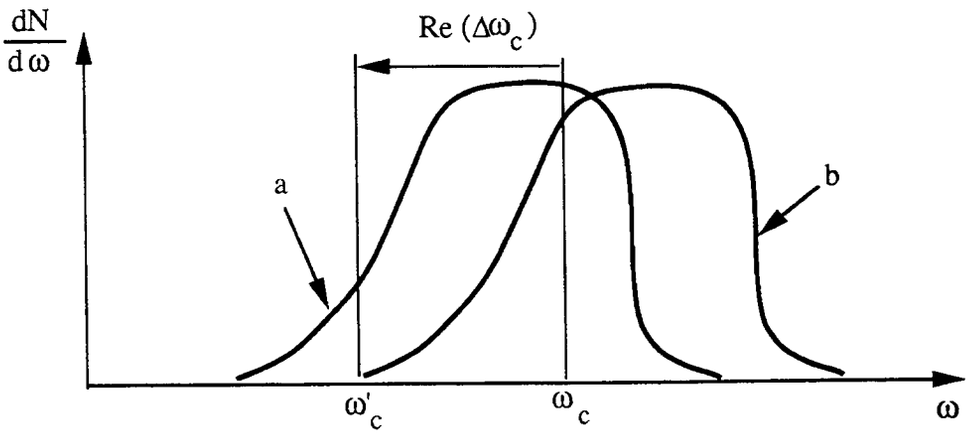


Fig. 5 An illustrative picture of Landau damping: a) without reactive broad band impedance, b) with reactive broad band impedance

This picture also illustrates the effect of the reactive part of the machine broad band impedance (space charge, inductive wall), which is to change the spectrum of incoherent frequencies (Fig. 5b). The coherent line ω_c may or may not be displaced depending on the interaction considered. In particular, the frequency of the longitudinal dipole mode is not shifted when one considers the effect of the inductive wall impedance. In the new situation, illustrated in Fig. 5b, the instability line may fall outside the spectrum of incoherent frequencies, resulting in a complete loss of Landau "damping". One then observes coherent beam oscillations sustained for very long periods, much longer than would be anticipated from the spread of incoherent frequencies. In conclusion the effect of the broad band impedance, even though it does not produce beam instabilities in itself, may be to provoke a complete loss of Landau damping. The rule of thumb criterion for this to happen is when the frequency shift due to the broad-band impedance becomes comparable to the natural frequency spread in the zero intensity beam.

By making the restoring force highly non-linear one can considerably increase the natural frequency spread within the bunch. "Landau" cavities operating at a multiple of the RF frequency can be used for this purpose. The combination of RF voltages from the normal accelerating cavities and the "Landau" cavities can be such that the synchrotron frequency goes to zero at the bunch centre giving maximum synchrotron tune spread, and therefore maximum Landau "damping". It must be pointed out, however, that Landau cavities have been found not very effective when the bucket is already almost full, presumably because the tune spread which is already comparable to the synchrotron frequency does not change dramatically. The control of harmonic cavities, in the presence of strong beam loading, may also be difficult [8].

3.3 Bunch-to-bunch frequency splitting

The idea here is to reduce the coupling between bunches and hence the instability growth by making their unperturbed oscillation frequencies slightly different. This is reminiscent of Landau damping, although there we have a limited number of resonators (the number of bunches) with different frequencies. This makes computer simulation relatively easy unless the number of bunches is very large.

Indeed one observes a reduction of the growth rate or the suppression of the instability depending on the spread in oscillation frequencies, the frequency splitting pattern and the instability parameters. In all cases the stabilizing effect is weak; only instabilities with $\Delta\omega_s / \omega_s \leq 0.01$ can be suppressed by this method.

In the transverse plane, betatron frequency splitting can be achieved with an RF quadrupole excited at a multiple of the revolution frequency. Synchrotron frequency splitting is obtained with RF cavities driven at p f_0 (f_0: revolution frequency, $p \neq h$, h: harmonic number). Although this method is fairly simple to implement (for instance one can excite one of the normal RF cavities at a frequency $(h\pm1)\ f_0$), it reduces significantly the machine acceptance in the longitudinal plane, as some bunches see a reduced RF voltage. Similarly a large area is necessary in the transverse tune diagram to accommodate the additional bunch-to-bunch tune splitting.

Figure 6 shows an old example of longitudinal instability suppression by synchrotron frequency splitting (two different frequencies) [7].

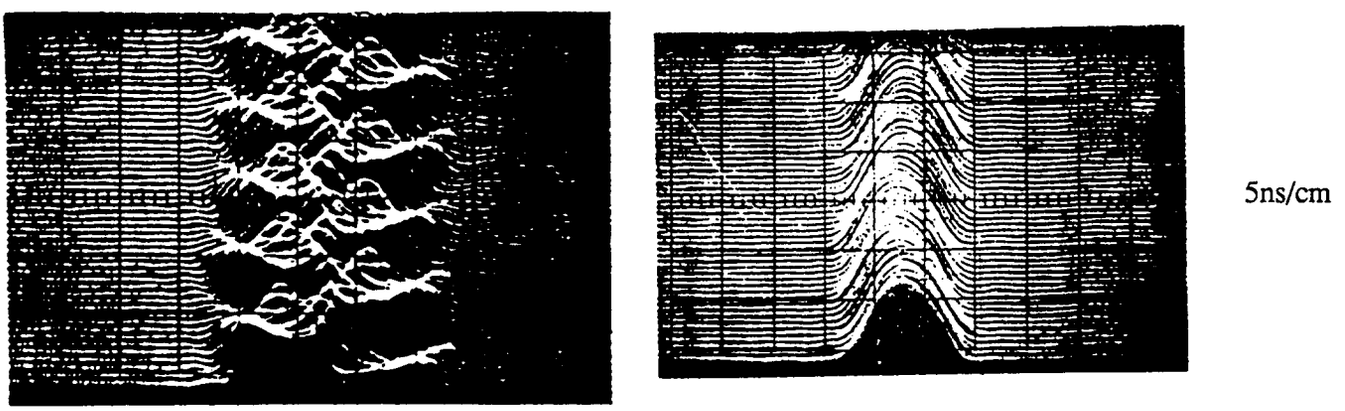


Fig. 6 Longitudinal instability in the CERN PS suppressed by frequency splitting (1970)

4. TRANSVERSE FEEDBACK SYSTEMS

The most effective cure against coupled bunch instabilities, transverse or longitudinal is the use of feedback systems [9]. We shall concentrate on the analysis of such systems in the following.

4.1 Architecture of a transverse feedback system

Figure 7 shows the schematics of a typical transverse feedback system. The beam displacement resulting from the instability is detected by pick-up electrodes, then amplified and applied as a correction to the beam with a downstream transverse deflector (or kicker).

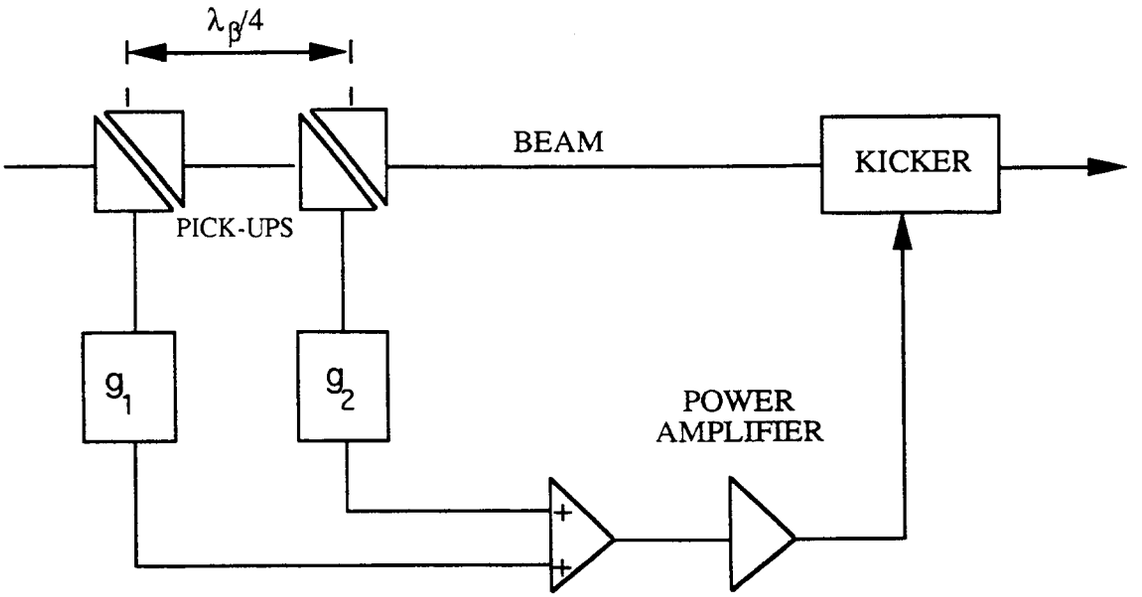


Fig. 7 A typical transverse feedback system

The transit time of the signal from pick-up electrodes to the deflector is arranged such that it matches the beam time of flight. In this way the correction is applied to the same particles which have generated the pick-up signal. To be insensitive to energy oscillations, the pick-up electrodes are better placed in a region of minimum dispersion.

Figure 8 shows the normalized transverse phase space diagram of a bunch (or ensemble of particles). Assume that the maximum displacement occurs at the pick-up location like indicated on the figure. A quarter betatron wavelength downstream, the correction is applied as a change of slope, along the x' axis. The result is a new trajectory which will spiral inwards and ultimately reach the origin.

The distance between pick-up and kicker should ideally be such that the corresponding betatron change is ± 90°. This can be achieved by combining the signals of two transverse detectors spaced by a quarter betatron wavelength (Fig. 7). For large machines (high Q values), it is easier to position pick-ups and kickers close together; in this case combining of two pick-up signals is also useful to correct for small tune changes; the overall delay is in this case close to one machine turn.

The feedback gain G is defined by the relation:

$$G = \frac{\Delta\theta}{x}\beta \qquad (3)$$

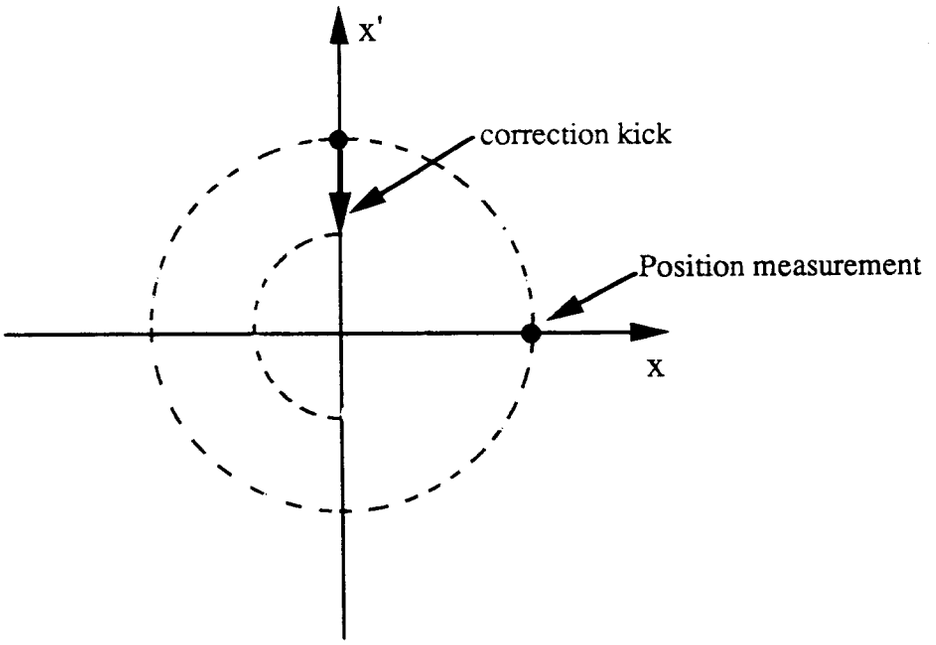


Fig. 8 Phase space diagram of a transverse bunch oscillation

where $\Delta\theta$ is the change of slope produced by a measured displacement x, assuming the same β value at the pick up and at the kicker.

After one turn the displacement has been corrected by:

$$\Delta x = \beta \, \Delta\theta = Gx \tag{4}$$

which corresponds, per unit time (smooth approximation) to:

$$\frac{dx}{dt} = Gf_0 x \tag{5}$$

and to a damping time $\tau = 1/Gf_0$. This would be correct if, at each turn, the beam would pass in the detector with maximum amplitude. Obviously this is not true because the tune cannot be an integer; as a result an additional factor 1/2 comes from averaging over all possible betatron phases at the pick-up position. Finally one obtains in the ideal case (optimum pick-up - kicker phase):

$$\frac{1}{\tau} = \frac{Gf_0}{2} \tag{6}$$

If the center of gravity of each bunch has to be damped individually (ideal bunch-to-bunch damping of the dipole mode), the required bandwidth of the system is, from the Nyquist theorem, $f_b/2$ where f_b is the bunch frequency ($f_b = Mf_0$ for M equidistant bunches). In many cases however, the feedback gain may not be constant in the 0 - $f_b/2$ frequency interval; it can be tailored to the frequency dependent instability growth rate. In the case of the resistive-wall instability which falls off rapidly with frequency, the feedback system can be limited in bandwidth to a few revolution harmonics. It is then not necessary to measure the position of each bunch individually: averaging over several bunches or undersampling is possible.

In the case of a few unstable modes, driven by narrow-band impedances, one can also use band pass filters and treat each mode separately. This "mode-by-mode" damping offers more flexibility (phase and gain can be adjusted independently for each frequency) but becomes cumbersome for many bunches and many possible modes of instability.

4.2 Rejection of closed-orbit offsets – Filtering

Without any signal processing in the amplifier chain, the offset of the closed-orbit in the pick-up will be amplified by G, with two undesirable effects:

- a change of the overall closed-orbit in the machine, dependent on G
- a saturation of the kicker amplifier.

To reject the unwanted closed-orbit component in the pick-up signal one discriminates between the quasi-static closed-orbit component and the fast varying betatron oscillation signal. More precisely the closed-orbit component is a periodic signal at the revolution frequency, which results from the uneven structure of the beam current along one machine turn. In particular, for a debunched beam only the DC component is present in the closed-orbit signal.

Several techniques are in use to reject the closed-orbit component of the pick-up signal. The self balanced front-end (Fig. 9) is a purely analog solution to the problem. The closed-orbit DC component (obtained after base line restitution of the pick-up signal) is used to balance the two electrodes and to bring the electrical center of the detector on the position of the closed orbit.

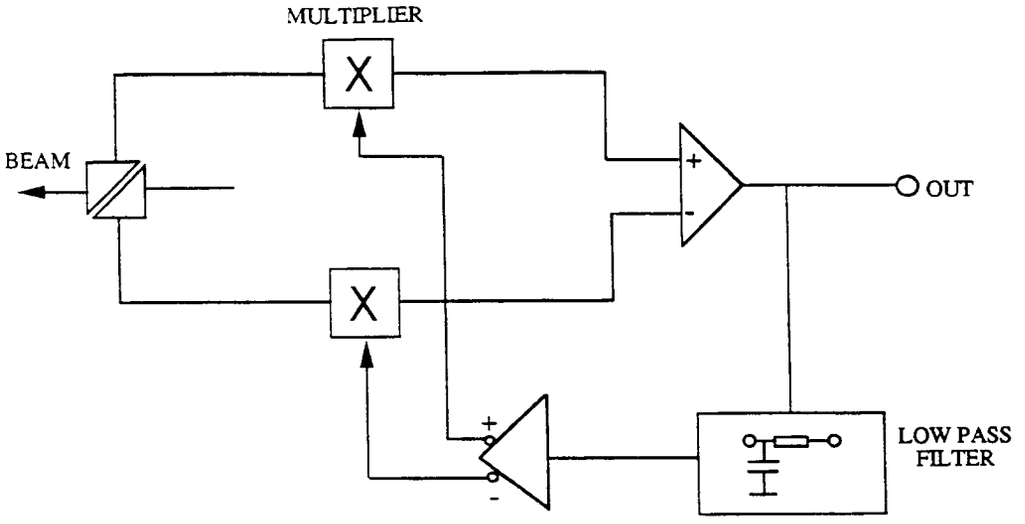


Fig. 9 Self-balanced frontend to reject closed orbit component

The more modern technique is based on periodic filters, which are conveniently constructed with digital technology. In its simplest form this filter makes the difference between the position measurements at successive turns according to the equation

$$u_n = x_n - x_{n-1} \tag{7}$$

feedback signal	measurement at turn n	measurement at turn n-1

The closed-orbit signal being almost constant from turn to turn is rejected by the filter.

The frequency response of the filter, schematized on Fig. 10 where the position at turn n-1 is obtained by a one turn delay, is simply:

$$\frac{u}{x} = 1 - \exp(-j\omega T_0) \tag{8}$$

Its amplitude and phase responses are displayed on Fig. 10. Closed-orbit signals which appear at multiples of the revolution frequency are rejected by the periodic notches (at every f_0 harmonic) of the filter. The gain and **phase** of the feedback signal (whose spectral components are close to the betatron lines f_0) are changed by the filter. In particular the optimum condition (90° phase difference between pick-up and kicker) is no longer fulfilled.

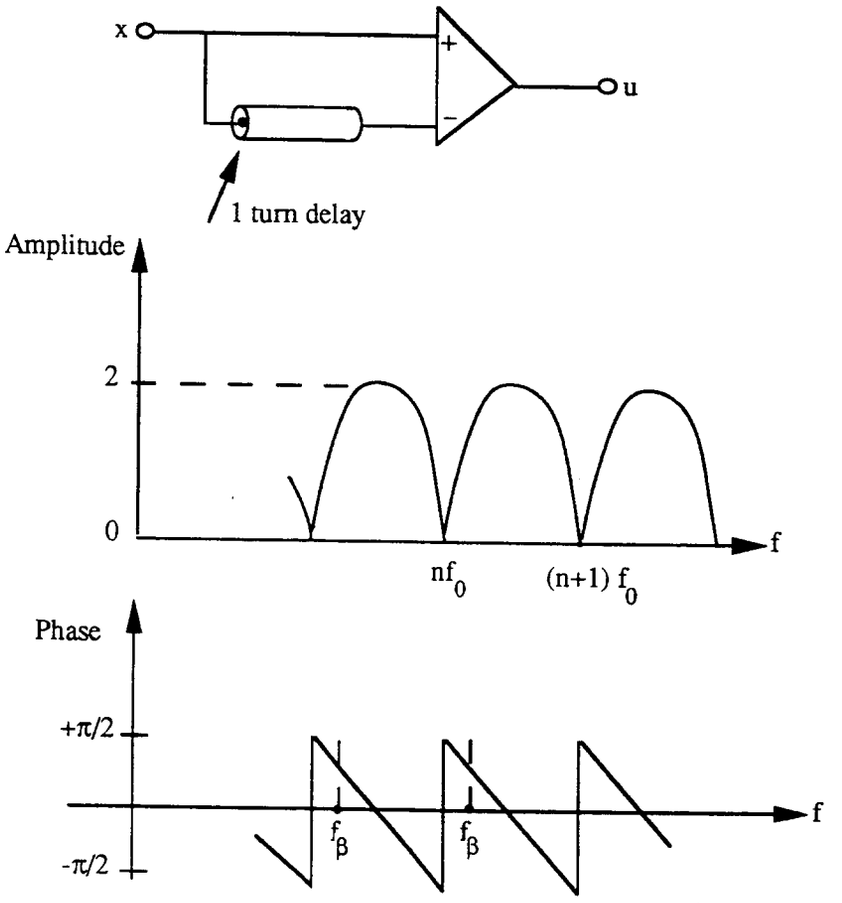


Fig. 10 Rejection of closed orbit with a single periodic filter

On the contrary, with more complicated filters, one can adjust the betatron phase to be optimum, even if the pick-up – kicker distance does not correspond to $\pi/2$ betatron phase difference.

The architecture of the filter ensures that each bunch is treated indepedently of all others; this is because the memory element in the filter is exactly one turn, which means that the signal of each bunch only combines with itself, not with any other bunch signal. One can also regard the filter as being used in a time multiplexing mode (one time slot per bunch): each bunch signal is processed by the same filter elements (adder and delay), but the memory (delay) contains at a given moment the information of all bunches.

The general expression for filters having this property is given by the expression:

$$u_{m,n} = a_o x_{m,n} + a_1 x_{m,n-1} + \ldots + a_k x_{m,n-k} + b_1 u_{m,n-1} + b_2 u_{m,n-2} + \ldots + b_p u_{m,n-p} \qquad (9)$$

where m is the bunch number, n the turn number and k and p are integers.

The filter is **non recursive** (FIR or Finite Impulse Response) if b_1, b_2 ... $b_p = 0$, otherwise it is **recursive** (IIR, or Infinite Impulse Response), according to the terminology in use in the digital filter theory. The only difference here is that, unlike a classical digital filter which is limited to the Nyquist frequency ($f_0/2$ here), the feedback filter works beyond, because of the particular nature of the beam signal. There is no aliasing effect up to half bunch frequency $f_b/2$ with one measurement per bunch per turn.

In other words, the filter behaves like a linear system up to $f_b/2$. The condition to obtain a notch filter response (zero at n f_0) is simply:

$$\sum_{n=0}^{k} a_n = 0$$

The complex frequency response of the filter is given by:

$$H(j\omega) = \frac{\sum_{n=0}^{k} a_n \exp(-jn\omega T_0)}{1 - \sum_{n=1}^{p} b_n \exp(-jn\omega T_0)} \qquad (10)$$

as illustrated on Fig. 11. This is characteristic of an ideal **bunch to bunch** feedback system for which all bunches are damped independently, each with the same damping rate. It also corresponds to a constant feedback gain for each betatron line.

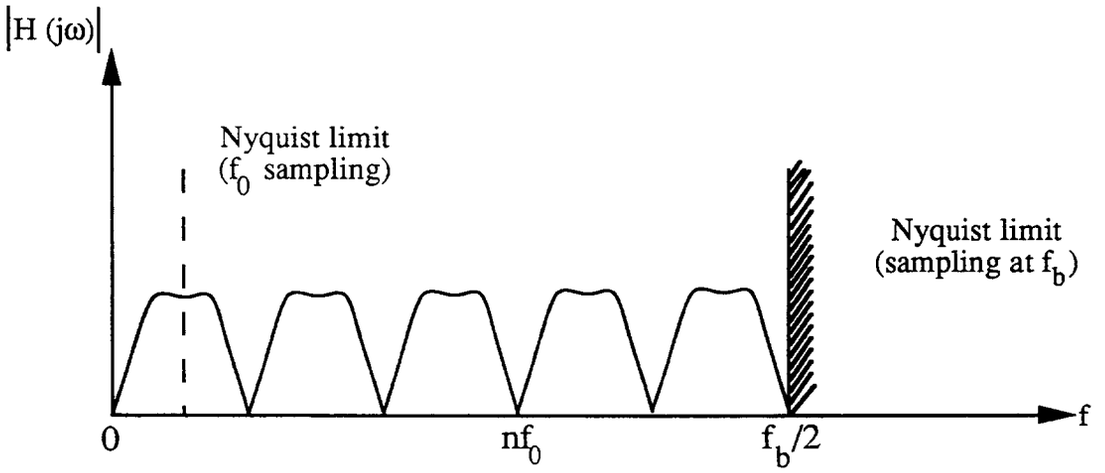


Fig. 11 Amplitude response of a periodic notch filter for a transverse feedback

In most cases however the overall feedback gain G does not need to be constant in the $0 = f_b/2$ frequency interval. It can be tailored to the behaviour of the instability growth rate versus frequency. In particular this is the case for the resistive-wall instability, which is peaked at low frequency (Fig. 12). If the maximum frequency of interest f_{max} is much smaller than $f_b/2$, sampling the position of every bunch is no longer necessary: one can

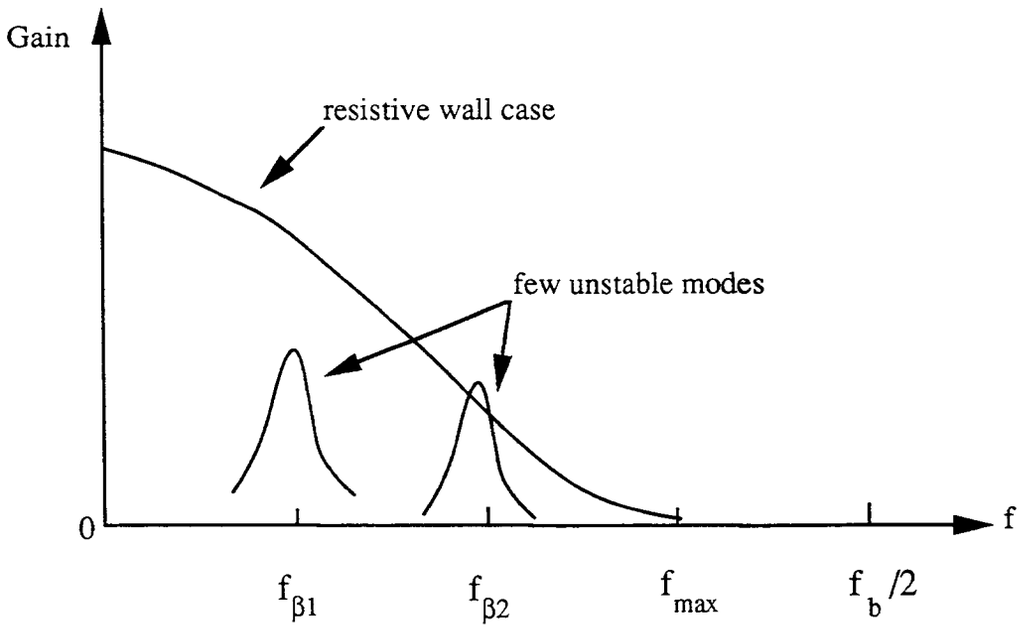


Fig. 12 Feedback gain tailored to instability characteristic

average the position measurements over several bunches (front-end bandwidth limitation) or (and) use undersampling.

When there are only a few unstable modes driven by narrow-band impedances, it may be sufficient to use selective filtering (Fig. 12). If the latter becomes very narrow to cover only one mode (filter bandwidth < $f_0/2$), the feedback technique is referred to as "mode-by-mode" damping.

As explained before the **one turn delay line** is the basic building block of the transverse feedback systems. At a fixed revolution frequency f_0, analog delay lines can be used, with the benefit of simplicity. However, the product delay-bandwidth of the line rapidly becomes a limitation, especially for large machines with many bunches. Sampled filters (at f_b) and in particular the family of *M* path filters can provide the desired transfer function of Fig. 11 (ideal bunch-to-bunch feedback system). Subsequent analog filtering (including the kicker frequency response) may be needed to obtain the overall gain vs. frequency curve. Analog filters with quasi constant delay (linear phase shift) are better suited to keep the phase of each betatron line close to its optimum value.

The rapid advances in digital filter technology have made this technique increasingly popular in constructing the filters of transverse feedback systems. The basic clock frequency must be a multiple of f_0 (up to f_b), which ensures delays of exactly one turn. The delay element itself is a memory, or shift register, the so-called FIFO (First In – First Out) version being the exact digital replica of an analog delay.

Sampling and digitizing at f_b is generally the most critical part of the filter from the point of view of speed (f_b can be as high as several hundred MHz) and number of bits. The latter must be compatible with the closed-orbit offset in addition to the betatron signal; it must also not introduce an unacceptable level of quantization noise. There is generally a trade off between the memory capacity and the arithmetic complexity of the processor (number of adders, multipliers, etc...). FIR filters usually need more memory capacity than their IIR counterparts for similar transfer functions. The filters with fixed parameters (a_n, b_n coefficients) can be synthesized in a simple way (shift of binary words) if a_n, b_n can be approximated with 2^p or $1-2^p$ (p integer).

Changes of f_0 are automatically accounted for in a digital filter clocked at mf_o, except for the unavoidable fixed part of the pick-up – kicker delay (cables, analog filter, kicker delay). Compensation of fixed delays in the chain against a change of f_0 is possible if the input (ADC) and output (DAC) clocks are connected to the mf_o source with different delays [10]. The difference should be equal to the fixed delay to be compensated.

4.3 Power requirements

Pick-ups and kickers are basically the same electromagnetic devices. They interact with the beam through the longitudinal component of their electric field [11] (Panofsky Wenzel Theorem).

A very common example considers the strip line pick-up and its kicker counterpart (Fig. 13), composed of two matched transmission lines (length l, spacing d, width w).

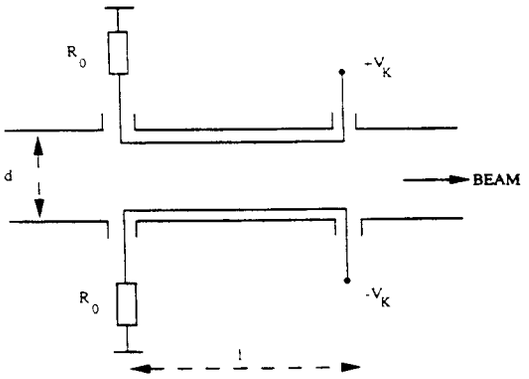


Fig. 13 Strip-line kicker

The kicker sensitivity is defined by:

$$K_{\perp} = \frac{1}{V_k}\frac{v}{e}\Delta p_{\perp} \tag{11}$$

where V_k is the applied voltage, v the particle's velocity, e the electronic charge and $\Delta p_{\perp}$ the transverse kick impressed on the beam.

The single passage beam deflexion $\Delta\theta$ (Eq. 3) is related to the kicker sensitivity by the relation:

$$\Delta\theta = \frac{\Delta p_{\perp}}{p} = \frac{K_{\perp} V_K e}{E\beta^2} \tag{12}$$

From the reciprocity and Panofsky Wenzel theorems one can derive the following relation between pick-up and kicker sensitivities, for the same device:

$$K_{\perp} = 2\frac{V}{\omega}\frac{S_{\Delta}}{R_0} \tag{13}$$

in which S_{Δ} is the pick-up sensitivity and R_0 the characteristic impedance of the line. The sensitivity of the strip line kicker of Fig. 13 is given by:

$$K_{\perp} = 2\frac{1}{d} \underbrace{\tanh\left(\frac{\pi w}{d}\right)}_{\text{geometry factor}} \underbrace{\frac{\sin(2\pi l/\lambda)}{2\pi l/\lambda}}_{\text{frequency response}} \tag{14}$$

For a given kicker cross section, sensitivity and bandwidth depend on the length l in opposite ways. This limits the operating frequency of the kicker to about c/4l.

Strip-line kickers can also operate in an RF band (from mf_b to $(m+\frac{1}{2})f_b$) using the "meander" technique [12]. Each section is $\lambda/4$ long (at the center frequency m f_b), and is rotated by 180° along the beam axis with respect to its neighbours.

At low frequencies ($f_b << c/2l$) magnetic or electrostatic deflectors can be used as well. The sensitivity of the electrostatic kicker is given by:

$$\Delta\theta = V_K \frac{l}{d} \frac{e}{E\beta^2} \tag{15}$$

When driven through a tetrode V_K becomes limited at high frequencies by the maximum current of the tube (loaded by the electrode capacitance). As a result the deflexion capability decreases with frequency as in the strip line case.

In most cases the kicker power is determined by the injection transients multiplied by the feedback gain. The latter is imposed by the specified damping rate needed to suppress the instability, or to avoid undesired blow-up in case of filamentation. Random injection errors come from fluctuations in beam steering or in magnetic elements in the machine. Systematic errors may be due to uncorrected closed-orbit which cannot be rejected instantaneously by the feedback notch filter. During its setting time closed-orbit error may dominate and ultimately determine the kicker power, unless *G* can be reduced during this limited period of time.

Linear operation of the transverse feedback system is not necessary during injection transient damping. The fastest damping rate, for a given kicker power is achieved in saturation (bang-bang) mode [13]. In this case digital limiters provide the best performance.

In a system where maximum deflexion can only be provided at low frequencies (slew rate limited), it is important to process the beam signal in such a way as to avoid any large high frequency components which are non significant [14]. This is the case in particular when a new batch is injected with some error, after a long gap in the bunch train: special signal processors may be used to take advantage of the gap length to reduce the required rate of change of the kicker voltage.

4.4 Emittance blow-up from damper noise

The noise present in the electronics of the beam detector will be amplified and applied to the beam by the transverse feedback system. One expects a coherent effect on the beam (corrected to some extent by the feedback system itself) and an incoherent one leading to beam blow-up. This is a crucial issue in colliders where small emittances are needed for high luminosity.

The overall damper configuration, including detector noise, is the same as that of a transverse cooling system, and therefore under some not too restrictive assumptions (debunched beam theory applied to a bunched beam, tune spread dominated by beam-beam effect, no loss of Landau damping) the classical cooling formula can be applied [15] [16]:

$$\frac{1}{\tau_{x^2}} = \frac{1}{2}\frac{f_0}{N}\sum_n \left[2\frac{g_n}{1+S_n} - M_n \frac{g_n^{\,2}}{(1+S_n)^2} - \frac{U_n\, g_n^{\,2}}{(1+S_n)^2} \right] \tag{16}$$

where $1/\tau_{x^2}$ is the growth rate of the beam emittance, N the total number of particles, g_n the feedback gain for band n, M_n the mixing factor, S_n the feedback via the beam factor and U_n the noise /Schottky signal ratio, for full mixing. The first two terms of the right side correspond to the cooling and mixing effects, which can be shown to be negligible in the case of a transverse damper. Only remains the last term, for which U_n can be expressed as the sum of the detector noise and the external excitation (to be suppressed by the damper). One obtains:

$$\frac{1}{\tau_{x^2}} = -f_0 \sum_{n=0}^{j} \frac{\left(\frac{\alpha^2 g_n^{\,2}}{12j} + r_n^{\,2} \right)}{\left(1 + \frac{g_n}{4\delta Q} \right)^2} \tag{17}$$

where α is the ratio of the Last Significant Bit of the ADC to the r.m.s beam size (the detector noise is assumed to be due to quantization in the ADC, and is assumed also to be additive), r_n is the external excitation in band n, $W = jf_0$ is the total bandwidth of the feedback and δQ the beam-beam tune spread.

The family of curves of Fig. 14 show the effect of g_n (feedback gain) in various cases. The term $g_n/4\delta Q$ must normally be larger than unity to achieve a damping time significantly shorter than the decoherence time, and avoid blow-up. For high gains the blow-up rate only depends on detector noise, as expected. At lower gains the blow-up rate is either smaller in case detector noise dominates over external excitation, or higher in the opposite situation, (which should be the normal operating mode of the damper during coast).

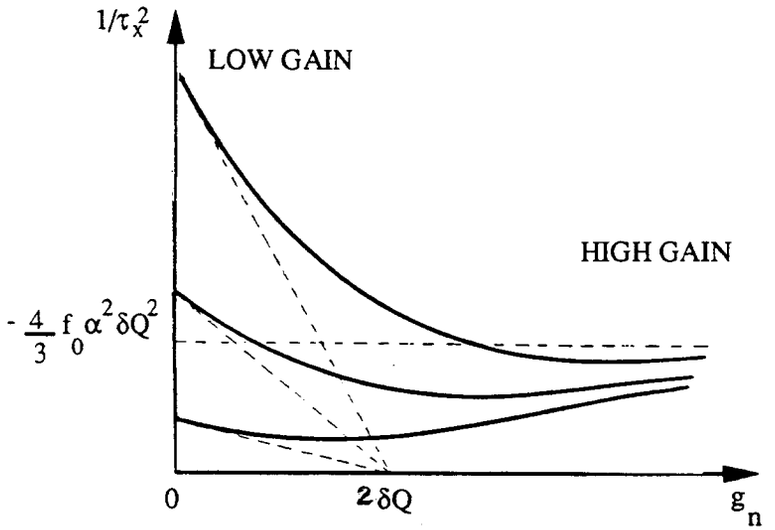


Fig. 14 Growth rate of the beam emittance due to feedback noise

5. LONGITUDINAL FEEDBACK SYSTEMS

5.1 Architecture of a longitudinal feedback system

A very simple longitudinal feedback system is sketched on Fig. 15 [7]. The radial detector is located at a position of large dispersion, such that any energy error can be easily detected and applied back to the same bunch via a longitudinal kicker. No phase shift is necessary here as the same variable (the energy deviation) is measured by the detector and applied to the beam by the kicker. The beam and signal transit times from pick-up to kicker should be identical, but, contrary to the transverse case, the distance pick-up to kicker is unimportant, because the "synchrotron wavelength" is much longer than one turn ($Q_s \ll 1$).

With two radial detectors located half a betatron wavelength apart one can reduce the undesired effect of betatron oscillations and closed-orbit error. Nevertheless, the sensitivity of the energy measurement via the beam radial position cannot be very high and hence noise contamination limits severely the performance of such a system.

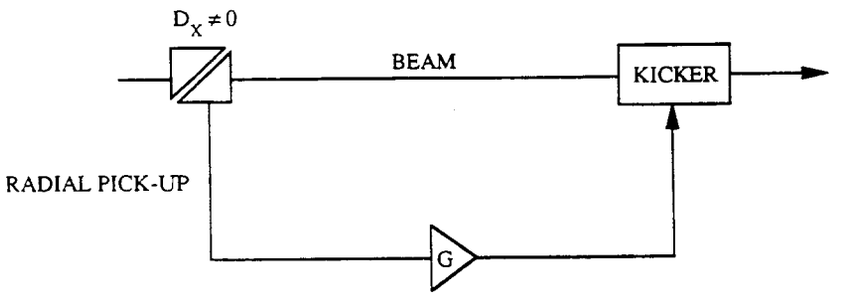


Fig. 15 A single longitudinal feedback system

Instead of a radial measurement one can use a phase measurement to detect any longitudinal oscillation of the beam. This however requires a phase shift of 90° at the synchrotron frequency f_s which can only be obtained by filtering, contrary to the transverse case. This is again because Q_s is much smaller than unity.

A "bunch-by-bunch" damping system is sketched in Fig. 16. The phase error signal, from a fast phase detector, is demultiplexed into M channels corresponding to M bunches. Each bunch signal is processed independently at low frequency (up to $f_0/2$) such as to provide a 90° phase shift at the synchrotron frequency. The M signals are then multiplexed into a common kicker drive signal.

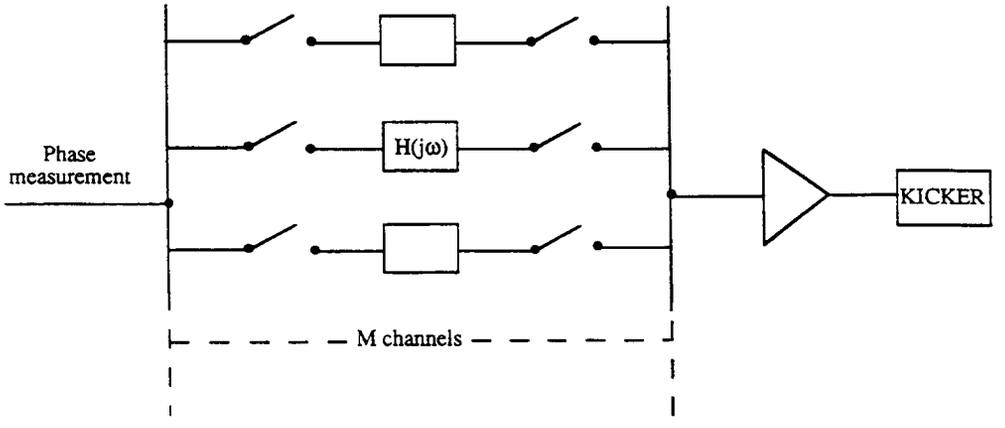


Fig. 16 Longitudinal bunch by bunch damping system

Each individual channel should have a filter characteristic H (jω) which not only provides high gain and 90° phase shift at f_s, but also negligible gain at zero frequency and $f_0/2$. The overall response of the "*M* path filter", including multiplexing and demultiplexing is periodic at f_0, each half band being the replica of the H(jω) transfer function (Fig. 17). The system is linear (no aliasing effects) up to $f_b/2$. As shown in Fig. 17 feedback gain is provided at the frequencies $nf_0 \pm f_s$ of all possible instability modes. Again there is complete equivalence between a bunch by bunch feedback system as depicted on Fig. 16 and a "mode-by-mode" feedback system with a constant gain for all instability modes.

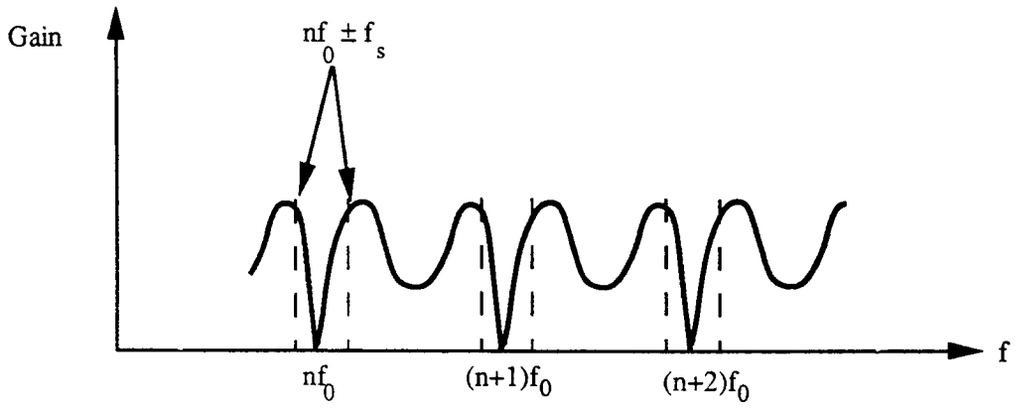


Fig. 17 Overall response of a longitudinal feedback system

5.2 Filtering

Periodic filters can be realized in analog technology (*M* path filters, Fig. 16) when the number of bunches is small. Modern digital technology is better suited in the case of many bunches. To obtain the required phase shift at f_s ($f_s \ll f_0$), one needs a large delay, or in digital technology a large memory capacity for a given sampling frequency (usually f_0 per bunch). On the other hand, sampling at f_0 may be redundant if Q_s is very small, opening the possibility of downsampling [17] (the sampling frequency being a submultiple of f_0, per bunch). This however increases the noise in the system.

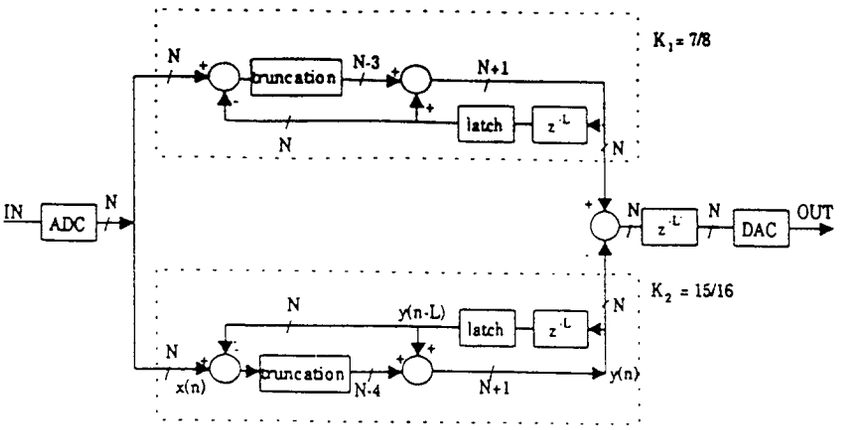
Non-recursive filters based on the difference of two phase measurements of the same bunch distant by m turns automatically provide notches at f_0 harmonics (rejection of constant or slowly varying phase offsets) [18] [19]. If m is small ($m \gg 1/Q_s$), the peak of $|H(j\omega)|$ happens at a frequency much smaller than f_s. This type of operation, sometimes referred to as the "differentiator mode" provides a phase shift close to 90° at f_s with a small memory capacity, but at the expense of more noise coming from the unnecessarily large gain at frequencies higher than f_s. The typical duration of the transient response of the filter is given by the inverse of the peak frequency, and can be made short if necessary. This is typically the case if the feedback must quickly damp injection oscillations.

By selecting $m = 1/2Q_s$, the peak of $|H(j\omega)|$ coincides with f_s, which is the best situation as far as noise is concerned. In this "peak gain mode", the 90° phase shift must be provided by an additional delay of length $T_s/4$. The total memory capacity needed thus corresponds to a total delay of $3T_s/4$, which is also the duration of the transient response of the filter.

Another technique, based on the difference of two recursive filters suppresses the undesirable lobes of $|H(j\omega)|$ which appear with non recursive filters. The transfer function:

$$H(j\omega) = \frac{1-K_1}{1-K_1 \exp(-j\omega T_0)} - \frac{1-K_2}{1-K_2 \exp(-j\omega T_0)} \qquad (18)$$

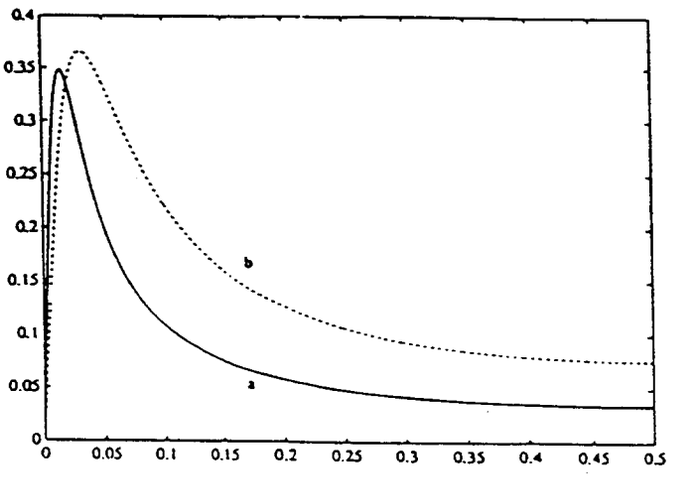
can be adjusted with two parameters K_1 and K_2. In the architecture of Fig. 18 K_1 and K_2 are of the form $1\text{-}2^{-p}$ and multiplication can be synthesized by adding and truncation, which makes processing very fast and cheap. The frequency and transient responses of this filter are given in Fig. 18.



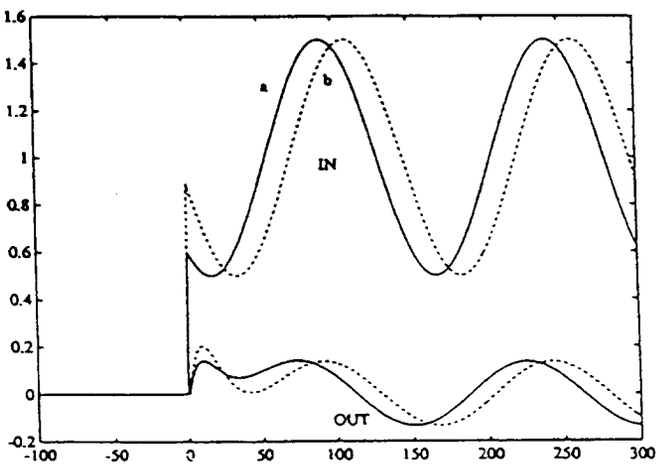
z^{-L}: delay of L clock periods (one turn)
$z^{-L'}$: delay of L' clock periods (one turn minus a fixed delay)
N: number of bits

Equation of the branch of the filter with K=15/16:
y(n)=y(n-L)+(x(n)-y(n-L))/16 or
y(n)=15/16*y(n-L)+x(n)/16

Architecture of the filter



Amplitude response of the filter
a: K_1=7/8, K_2=15/16
b: K_1=3/5, K_2=7/8



Rejection of injection offset
top: f_s oscillation (2 phases a, b) superimposed on unit step offset
bottom: filter output

Fig. 18 A simple longitudinal filter based on the difference of two recursive filters

5.3 Synthesis of a longitudinal impedance

A longitudinal dipole, coupled bunch instability is driven (above transition) by an impedance which has the property:

$$\text{Re } Z(kf_0 + f_s) > \text{Re } Z(kf_0 - f_s) \tag{19}$$

Conversely, the instability mode number n can be damped if an impedance with the opposite property:

$$\text{Re } Z(pf_b + nf_0 + f_s) < \text{Re } Z(pf_b + nf_0 - f_s) \tag{20}$$

is inserted in the beam. To avoid strong interaction with the revolution frequency line at $pf_b + nf_o$, one imposes the additional condition:

$$Z(pf_b + nf_0) \approx 0$$

which leads to the design objective:

$$\text{Re } Z(pf_b + nf_0 + f_s) \approx -\text{Re } Z(pf_b + nf_0 - f_s) \tag{21}$$

It can be shown that damping of quadrupole, sextupole modes can be achieved as well if the real part of the impedance is symmetric around $pf_b + nf_0$, over a band $\pm 2f_s$, $\pm 3f_s$ respectively.

The problem here is to synthesize an impedance such that its real part changes sign every f_0 harmonic to damp all possible instability modes. With a broadband detector and kicker (bandwidth $\approx f_b$) the problem can be solved with periodic filters similar to those described earlier. The difference here is that the beam current signal is directly filtered (without phase detection) and transmitted to the longitudinal kicker.

The system being intrinsically sensitive to all internal bunch modes (dipole, quadrupole, etc ...) the shape of the filter in a $0\text{-}f_0/2$ interval is important. In particular due to the unavoidable one turn delay, Re Z usually changes sign not only at nf_0, but also between nf_0 and $\left(n+\frac{1}{2}\right)f_0$ making high internal bunch modes potentially unstable. These modes are however generally strongly Landau damped.

The periodic filters which cover a frequency band of width 0 to f_b are translated in frequency by coherent mixing (Fig. 19) to keep the system linear (no aliasing). The mixers (input and output, in-phase and in-quadrature), whose reference frequency is a multiple of f_b can be regarded as phase and amplitude detectors (input mixers) and phase and amplitude modulators (output mixers). In this respect the architecture of the synthesized impedance looks like a combined dipole (phase detector – phase modulator) and quadrupole (amplitude detector – amplitude modulator) longitudinal damper, with the two periodic filters attributed to each bunch mode (dipole and quadrupole).

5.4 Pick-ups and kickers

Longitudinal beam detectors are straightforward devices, with large bandwidth and dynamic range. In the case of closely spaced bunches the problem of independent phase measurement of each bunch needs special treatment. The bunch signal is transformed into an RF burst of finite duration (ideally), which is shorter than the bunch to bunch distance. This signal is used to make the phase comparison with a reference frequency at a multiple of f_b.

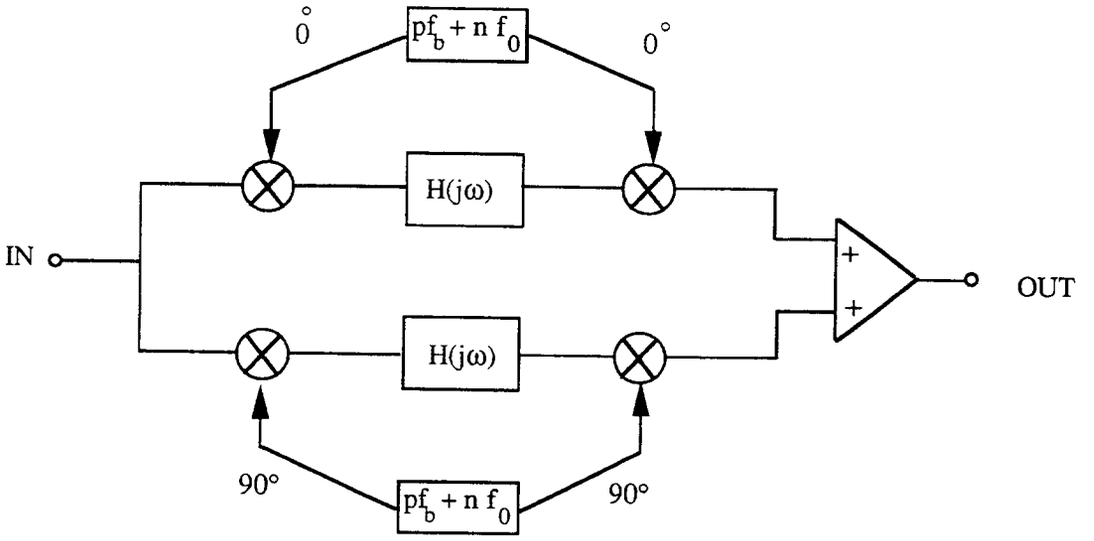


Fig. 19 Coherent mixing (in phase and quadrature) fro frequency translation

The RF burst generation can be made directly by combining the signals of several pick-up electrodes located at wavelengths intervals or with an array of directional couplers. A typical output of the RF burst generator developed for the SLAC B factory is given on Fig. 20.

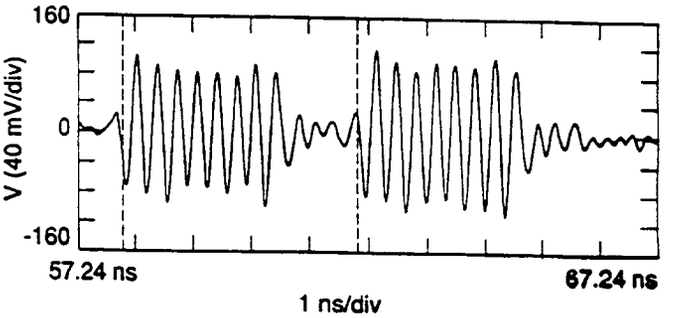


Fig. 20 Independent phase measurement of closely spaced bunches: PEPII burst generator output

The feedback voltage V_f applied to the beam is determined by the required gain (or damping time τ_f) and the initial phase error, according to:

$$V_f = 2\,\frac{\Delta\omega_s}{\omega_s}\,V\cos\phi_s\Delta\phi \qquad (22)$$

where $\Delta\omega_s=1/\tau_f$, $\omega_s=2\pi f_s$, V: RF voltage, ϕ_s: stable phase angle and $\Delta\phi$: initial phase error.

The initial phase error is usually large if the feedback system must damp injection oscillations (this is typically the case for proton machines), whereas $\Delta\phi$ can be chosen not much above the detection threshold of the phase measurement in case the feedback system only prevents instabilities from developing.

The RF cavities can be used as longitudinal kickers if their bandwidth is large enough. The feedback voltage appears there on the skirts of the resonance curve of the cavity, whose transfer function must be properly accounted for by filtering. This technique is used in the CERN PS Booster, and is particularly well suited there because of the large frequency sweep

during acceleration. Dedicated feedback cavities are also considered, for instance for LHC, with a flattened frequency response provided by RF feedback.

With short bunches, as in e^+e^- machines, very high frequency kickers or cavities [21] are interesting because of their small size. The travelling wave kicker for the SLAC B factory [1] is composed of an array of gaps connected in series with half wavelength lines (radial and azimuthal) (Fig. 21).

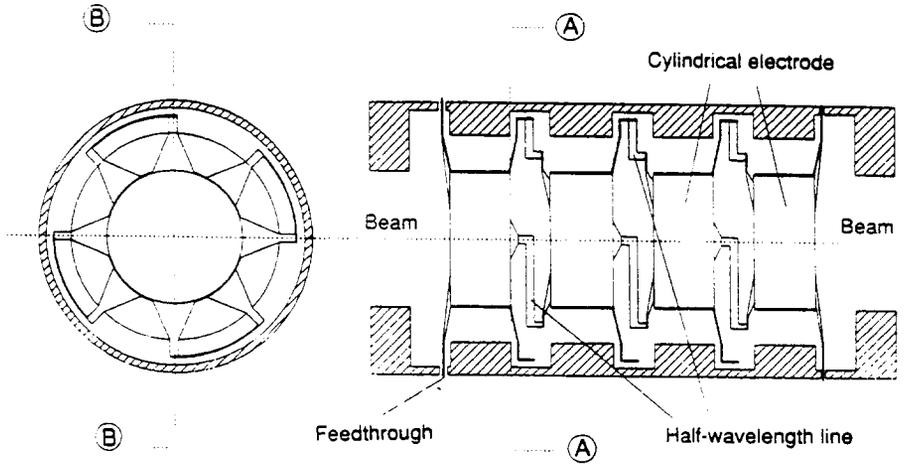


Fig. 21 Travelling wave longitudinal kicker for the SLAC B factory

5.5 Instabilities driven by the fundamental mode of RF cavities

In this particular case, the availability of an already existing kicker, namely the RF cavity and its associated power amplifier calls for a special treatment of the problem.

The stability of the n = 0 mode is governed by the well known Robinson stability criterion, valid without loops. It sets limits on the detuning angle and maximum current. In the presence of loops (e.g. tuning loop, amplitude loop, etc ...) the stability criterion is strongly pertubed and in general difficult to evaluate precisely. The RF feedback technique is often used in this case to dramatically increase the stability margin.

For large rings (f_0 small), the cavity detuning may exceed f_0. This is the case if:

$$\frac{1}{2}\,\frac{R}{Q}\,f_{RF}\,\frac{I_b}{V} \geq f_0 \qquad (23)$$

where R/Q is the cavity geometric parameter and I_b the RF component of the beam current. In this case mode n = 1 (f_0 away from f_{RF}) can sit on the cavity resonance and be strongly excited. A possible cure is with RF feedback to reduce the cavity dynamic impedance to an acceptable value. This is limited, however by the RF amplifier delay and peak power capability of the tube, which must compensate the n = 1 component at I_b [21].

Another type of approach is to limit cavity detuning below f_0, by using cavities having a parameter $\frac{R}{Q}\cdot\frac{1}{V}$ small. This parameter is inversely proportional to the charge stored on the equivalent cavity capacitor. It can be reduced by at least one order of magnitude using large aperture superconducting cavities, as compared to conventional copper structures. Another solution is with storage cavities outside the beam line and coupled to the accelerating cavity; in this case however new resonant modes appear, which must be properly corrected [23].

REFERENCES

[1] An asymetric B factory based on PEP - SLAC 372, Feb. 1991.

[2] R. Boni, A. Gallo, Status of the DAFNE RF system, 4th European Part. Acc. Conf., EPAC 94, London.

[3] H. Padamsee et al., Accelerating cavity development for the Cornell B factory CESR B, 1991 Part. Acc. Conf., San Francisco, p. 786-788.

[4] R. Boni et al., Update on the broadband waveguide to 50Ω coaxial transition for parasitic mode damping in the DAFNE RF cavities, 4th EPAC, London June 1994.

[5] W.R. Smythe et al., A versatile RF cavity mode damper, EPAC 1990, Nice, p. 976-978.

[6] F. Sacherer, A longitudinal stability criterion for bunched beam, 1973, Part. Acc. Conf., p. 825.

[7] D. Boussard, J. Gareyte, Damping of the longitudinal instability in the CERN PS, Proc. 8th International Conf. of High Energy Accelerator, CERN 1971.

[8] J.M. Baillod et al., A second harmonic (6-16 MHz) RF system with feedback reduced gap impedance for accelerating flat topped bunches in the CERN PS booster, IEEE Trans. Nucl. Science , NS30 n°4, Aug. 1983, p. 3499.

[9] F. Pedersen, Feedback systems, CERN PS/90-49 (AR).

[10] D. Boussard, G. Lambert, Reduction of the apparent impedance of wideband accelerating cavities by RF feedback, PAC 1983, Sante Fe, and CERN SPS/83-18.

[11] D. Boussard, Schottky noise and beam transfer function diagnostics, CERN Accelerator School, Oxford 1985, CERN 87-03.

[12] I.N. Ivanov, V.A. Melnikov, A.S. Scheulin, The meander kickers for the transverse damping system of the cycle accelerators, JINR, Dubna 1993.

[13] R. Bossart, R. Louwerse, J. Mourier, L. Vos, Operation of the transverse feedback system at the CERN SPS, Part. Acc. Conf. Washington DC, 1987, p.763-765.

[14] D. Boussard, E. Onillon, Damping of phase errors at injection in the LHC, IEEE, Trans. Nucl. Science, NS-40 (1993), p. 2379-2381.

[15] D. Boussard, Evaluation of transverse emittance growth from damper noise in the collider, CERN SL/Note 92-79 (RFS).

[16] W. Chou, J. Peterson, The SCC transverse feedback systems in "Supercolliders and Superdetectors", Editors W.A. Barletta, H. Leutz, World Scientific 1993.

[17] Hindi et al., Down sampled signal processing for a B factory bunch by bunch feedback system, EPAC, Berlin, March 1992.

[18] L. Ma, S. Kurokawa, E. Kikutani, FIR filters for the bunch by bunch feedback system of TRISTAN II, KEK Report 1993.

[19] F. Pedersen, Multibunch feedback, Lecture Notes in Physics, 425, Springer-Verlag, Berlin (1994).

[20] D. Briggs et al., Prompt bunch by bunch synchrotron oscillation detection via fast phase measurement, PAC, San Francisco , May 1991.

[21] M. Ebert et al., Transverse and longitudinal multibunch feedback systems for PETRA, DESY 91-036.

[22] D. Boussard, RF power requirements for a high intensity proton collider, CERN SL/91-16 revised.

[23] Y. Yamazaki, T. Kagemaya, A three cavity system which suppresses the coupled bunch instability associated with the accelerating mode, KEK Preprint 93-15 (1993).

BEAM LOADING

D. Boussard
CERN, Geneva, Switzerland

Abstract
Beam loading on RF cavities may seriously limit the performance of high-intensity circular accelerators or storage rings. The RF power requirements to correct for beam loading will be first examined in several typical cases (lepton and hadron machines). Then, the methods to control the RF system (feedback and feedforward) and to achieve stability under heavy beam loading conditions will be reviewed.

1. INTRODUCTION

In accelerator language, beam loading usually refers to the effects induced by the passage of the beam in the radio frequency cavities. As such, it could be considered to be one particular example of the more general problem of the beam interaction with its surroundings, in this case the cavity impedance.

However the beam loading problem deserves a special treatment, for several reasons. Firstly, the RF cavities are very often the largest contributor to the total ring impedance (in the following we shall concentrate on circular machines) and, consequently, power considerations play a very important role in beam-loading problems. Secondly, contrary to many other machine elements, the RF cavities are well known items being carefully designed and measured, from the RF point of view, and are easily accessible from the outside world via the RF power amplifier. Dedicated correction techniques can therefore be used where not only the cavity but also its associated RF amplifier are included.

In the following we shall first consider the stationary situation established in the beam-cavity system, the two extreme cases being when the bunches are wide apart and when every bucket is filled. The case of a long train of bunches followed by a gap is of particular importance in high intensity machines (ion clearing gaps in e^+e^- machines, beam dump gap for hadron colliders, for instance) and will be examined separately. Travelling-wave cavities with their inherent advantages as far as beam loading is concerned will be examined in this context.

Before settling to the stationary situation, the beam-cavity system undergoes a transient phase which may be very harmful to the beam, especially for hadron machines without natural damping. To circumvent this problem, it will be shown that RF power must be available. Finally the various methods used to control the RF amplifier-cavity combination in order to suppress beam-loading effects will be reviewed.

2. SINGLE-BUNCH PASSAGE IN A CAVITY

When the distance between bunches is very large compared to the filling time of the cavity, the fields induced by the previous bunches, or the previous bunch passages of the same bunch, have decayed sufficiently and can be neglected. Consequently, before the bunch passage the RF waveform is a pure sinewave produced by the RF generator (Fig. 1a).

The effect of the bunch passage is to excite an additional field in the cavity (Fig. 1b). For a short bunch (short compared to the RF period) and considering only the fundamental resonance of the cavity, the excited waveform is an exponentially decaying sinewave oscillating at the resonant frequency of the cavity ω_c.

Combining the generator driven and beam driven waveforms, one obtains the total voltage $V(t)$ at the cavity gap (Fig. 1c).

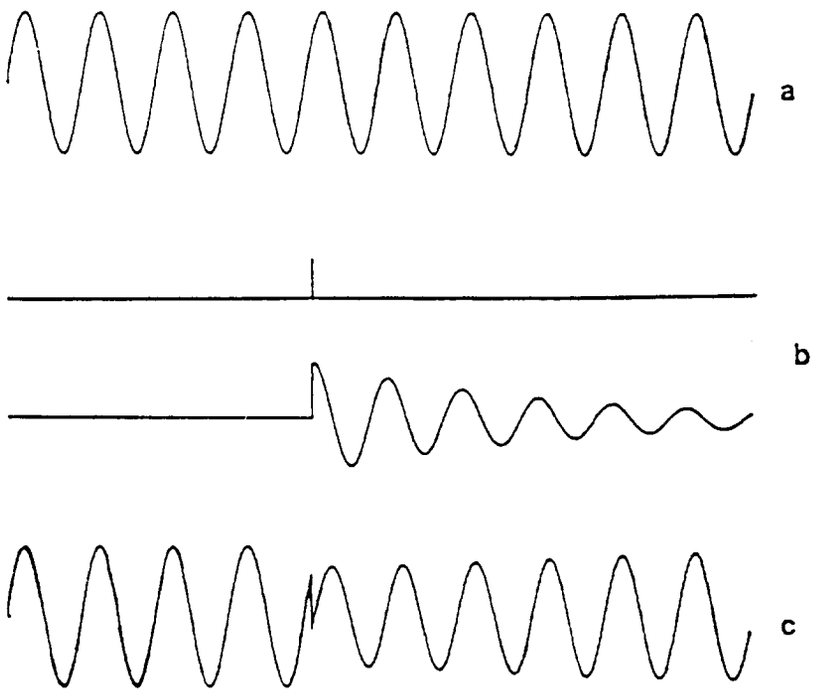


Fig. 1 Single-bunch passage in a cavity

Assuming a negligible beam current (unperturbed voltage), the power delivered to the beam by the RF generator is simply:

$$P = -\frac{\vec{V}_g \cdot \vec{i}_b}{2} \tag{1}$$

where $\vec{V}_g$ is the generator driven voltage and $\vec{i}_b$ is the fundamental component of the beam current.

When crossing the gap the charge q induces the voltage V_{b0}, and loses a fraction of its energy which is finally transformed into heat in the cavity walls before the next bunch passage. In this situation the effective gap voltage for the beam is modified by the effect of the beam. We can define an effective gap voltage V evaluated in the following way.

In the transient phase (assuming a short bunch length compared to the RF period), the cavity gap impedance can be represented by a single capacitor C related to the cavity parameters by:

$$\frac{1}{C} = \frac{R}{Q_0}\omega_c \tag{2}$$

where ω_c is the resonant frequency of cavity, Q_0 the unloaded cavity quality factor and R the shunt impedance of the cavity (circuit convention). Obviously $V_{b0} = q / C$, and the energy lost by the bunch and stored in the cavity just after the bunch passage amounts to:

$$W = \frac{1}{2} C \; V_{b0}^2 = \frac{1}{2} q \; V_{b0} \; . \tag{3}$$

This corresponds to an average power loss of:

$$\frac{W}{T_b} = \frac{1}{2} i_{bDC} V_{b0} = \frac{1}{4} i_b V_{b0}$$

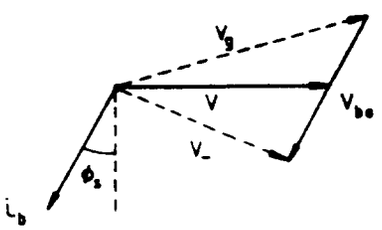
where T_b is the bunch distance and $i_{bDC} = i_b/2$ (short bunch approximation). The net power received by the beam P' is simply, remembering that $\vec{i}_b$ and $\vec{V}_{b0}$ are in phase:

$$P' = -\frac{\vec{V}_g \cdot \vec{i}_b}{2} - \frac{1}{4} \vec{i}_b \cdot \vec{V}_{b0} \tag{4}$$

$$P' = -\frac{1}{2} \left(\vec{V}_g + \frac{1}{2} \vec{V}_{b0} \right) \cdot \vec{i}_b = -\frac{1}{2} \vec{V} \cdot \vec{i}_b \tag{5}$$

Here $\vec{V}$ is the effective RF voltage, delivering the net power P' to the beam.

Equation (5) leads to the vector diagram of Fig. 2, which shows the voltages before ($\vec{V}_g$) and after $\left(\vec{V}_-\right)$ the bunch passage and their relations with the bunch current. Obviously the voltage V_g to be delivered by the generator is higher for the same effective voltage V, than in the case of no beam loading. The excess power can be easily computed from the cavity shunt resistance and beam current.



V_g: generator driven voltage

V_-: voltage after bunch passage

V: net voltage experienced by the beam

Fig. 2 Vector diagram – single-bunch passage in a cavity

One can also remark that the beam "sees" only one-half of its own induced voltage:

$$\vec{V}_b = \frac{1}{2} \vec{V}_{b0}$$

where

$$\vec{V} = \vec{V}_g + \vec{V}_b \; .$$

This result is sometimes quoted as "the fundamental theorem of beam loading", and can be demonstrated more generally (P. Wilson [1]) using linearity and superposition. Similarly, it is easy to show that, in fact, V_{b0} represents the sum of all beam induced voltages for all cavity modes.

3. MULTIPLE-BUNCH PASSAGES [1,2]

We look for a stationary solution, when an infinite train of bunches, spaced by h_b RF periods, crosses the cavity gap. Following P. Wilson's analysis [1] we should replace V_g in Eq. (5), which represents the voltage just before the bunch passage, by the combination of the generator-driven voltage and the voltage resulting from all previous bunch passages. The decay of the voltage between two successive bunch passages is simply $\delta = T_b / T_f$ where T_f is the cavity time constant ($T_f = 2Q_L / \omega_c$, Q_L: loaded cavity quality factor), and the phase shift with respect to the RF generator amounts to $\psi = \omega_c T_b - 2\pi h_b$. The relation $\vec{V} = \vec{V}_g + 1/2\vec{V}_{b0}$ will therefore transform into:

$$\vec{V} = \vec{V}_g + \vec{V}_{b0}(e^{-\delta}e^{j\psi} + e^{-2\delta}e^{2j\psi} + \ldots) + \frac{1}{2}\vec{V}_{b0} = \vec{V}_g + \vec{V}_b \tag{6}$$

in which the term in brackets represents the contributions of all previous bunch passages, whereas the last one reflects the effect of the bunch on itself (Fig. 3).

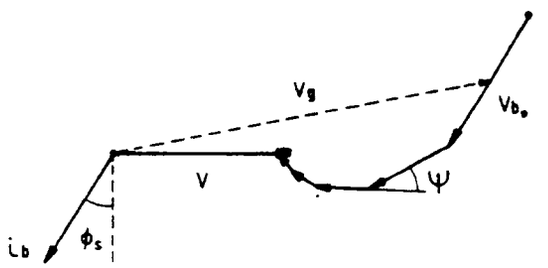


Fig. 3 Vector diagram – multiple-bunch passages

Using the sum of the geometric series:

$$V_{b0}(1 + e^{-\delta}e^{j\psi} + e^{-2\delta}e^{2j\psi} + \ldots) = \frac{V_{b0}}{1 - e^{-\delta}e^{j\psi}} \tag{7}$$

one obtains:

$$\vec{V}_b = \vec{V}_{b0}\left(\frac{1}{1 - e^{-\delta}e^{j\psi}} - \frac{1}{2}\right) \tag{8}$$

which, when separating real and imaginary parts leads to:

$$V_b = V_{b0}\left[F_1(\delta, \psi) + jF_2(\delta, \psi)\right] \tag{9}$$

with:

$$F_1(\delta, \psi) = \frac{1 - e^{-2\delta}}{2(1 - 2e^{-\delta}\cos\psi + e^{-2\delta})}$$
$$F_2(\delta, \psi) = \frac{e^{-\delta}\sin\psi}{1 - 2e^{-\delta}\cos\psi + e^{-2\delta}} \tag{10}$$

If we introduce now the more usual cavity parameters:

$$\tan\phi_c \text{ (detuning angle)} = 2Q_L\frac{\omega_c - \omega}{\omega_c} \tag{11}$$

$$\beta \text{ (coupling coefficient)};\quad Q_L = Q_0 \frac{1}{1+\beta} \tag{12}$$

and $\delta_0 = T_b / T_{f0}$ (T_{f0} being the filling time of the unloaded cavity), Eq. (9) becomes:

$$V_b = 2 i_0 R \delta_0 \left[F_1(\beta, \phi_c) + j F_2(\beta, \phi_c) \right] \tag{13}$$

where i_0 is the DC beam current.

The functions F_1 and F_2 are given by:

$$F_1(\beta, \phi_c) = \frac{1 - e^{-\delta_0 (1+\beta)}}{2D} \tag{14}$$

$$F_2(\beta, \phi_c) = \frac{e^{-\delta_0 (1+\beta)} \sin \delta_0 (1+\beta) \tan \phi_c}{D} \tag{15}$$

$$D = 1 - 2 e^{-\delta_0 (1+\beta)} \cos\left[\delta_0 (1+\beta) \tan \phi_c \right] + e^{-2\delta_0 (1+\beta)} \,. \tag{16}$$

From these expressions, it is possible to calculate the generator power needed to produce a given accelerating voltage V. For a generator which is assumed to be matched, by using, for instance, a circulator between generator and cavity, one obtains [1]:

$$P_g = \frac{(1+\beta)^2}{4\beta} \frac{V_g^2}{2R} \frac{1}{\cos^2 \phi_c} \{A^2 + B^2\} \tag{17}$$

where A and B are complicated functions of cavity and beam parameters [1]. Numerical computations are required to optimize the various parameters in order to minimize P_g.

4. LIMITING CASE $\delta_0 \equiv 0$

When the bunch distance T_b is short compared to the unloaded cavity filling time (Fig. 4), Eqs. (10) simplify to:

$$F_1(\delta_0, \beta) = \frac{-1}{\delta_0 (1+\beta)(1 + \tan^2 \phi_c)} \tag{18}$$

$$F_2(\delta_0, \beta) = \frac{-\tan \psi}{\delta_0 (1+\beta)(1 + \tan^2 \phi_c)} \,. \tag{19}$$

Combined with (13), one obtains:

$$V_b = \frac{i_b R}{1+\beta} \frac{1}{1 - j \tan \phi_c} \,. \tag{20}$$

In this case the cavity gap waveform is approximately sinusoidal (Fig. 4), and the equivalent circuit of Fig. 5, where the beam current is represented by its component at the RF frequency i_b can be used. There the coupling coefficient β is simply related to the cavity and generator shunt

resistances by: $\beta = R / R_g$. Obviously, V_b given by Eq. (20) is the cavity voltage (sinusoidal in the approximation $\delta_0 \cong 0$) developed when i_g (generator current) = 0.

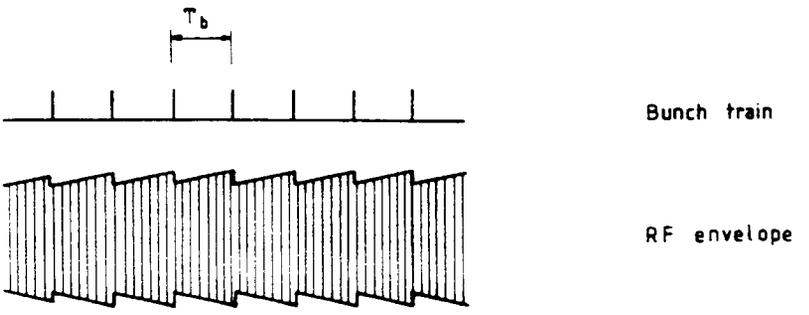


Fig. 4 Case $\delta_0 \cong 0$. The RF waveform is a quasi sinusoid

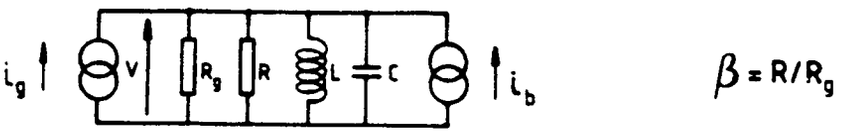


Fig. 5 Equivalent circuit for the case $\delta_0 \cong 0$

In the vector diagram of Fig. 6a, the total current $\vec{i}_t = \vec{i}_g + \vec{i}_b$ drives the RLC circuit and produces the gap voltage V. For a given V, the vector $\vec{i}_t$ follows the dotted line in Fig. 6a, when the detuning angle ϕ_c is varied. This is because the admittance of the equivalent RLC circuit has a constant real part.

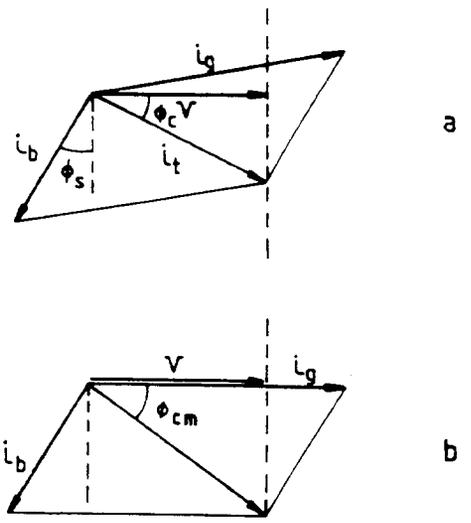


Fig. 6 Vector diagrams for the case $\delta_0 \cong 0$. Optimum tuning in (b)

The required RF power:

$$P_g = \frac{1}{2}\vec{V}\cdot\vec{i}_g \tag{21}$$

is a minimum for given V, i_b and ϕ_s, if the two conditions:

$$\tan\phi_{c0} = -i_b\frac{R\cos\phi_s}{(1+\beta)V} \tag{22}$$

$$\beta_0 = 1 + i_b\frac{R\sin\phi_s}{V} \tag{23}$$

are fulfilled. The minimum RF power for $\phi_c = \phi_{c0}$ and $\beta = \beta_0$ is given by:

$$P_{gm} = \frac{V^2}{2R} + Vi_b\sin\phi_s \ , \tag{24}$$

the first term corresponding to the cavity losses and the second to the power delivered to the beam. The optimum phase condition ($\phi_c = \phi_{c0}$) corresponds to i_g and V being in phase (Fig. 6b). Usually there is a servo-tuner which measures the phase difference between RF drive and gap voltage, and controls the cavity tune via a mechanical tuner or ferrite bias, for instance. At equilibrium of the servo-tuner, Eq. (22) is automatically satisfied. It corresponds to a cavity detuning:

$$\Delta\omega_0 = \omega_c - \omega = \frac{1}{2}\frac{R'}{Q_L}\omega_c i_b\frac{\cos\phi_s}{V} = \frac{1}{2}\frac{R}{Q}\omega_c i_b\frac{\cos\phi_s}{V} \tag{25}$$

where R' is the equivalent resistance of R and R_g in parallel.

On the contrary, the cavity coupling is usually fixed by construction, and can only be optimized for a given value of i_b and ϕ_s. However for a hadron storage ring, where $\phi_s = 0$, the critical coupling ($\beta = 1$) corresponds to the optimum situation.

For very large machines and high beam currents, RF cavity detuning may become equal to or larger than the revolution frequency f_{rev}. When cavity resonance coincides with a revolution harmonic one can expect large modulations of the RF voltage if all RF buckets are not equally populated (see § 5), but also a strong coupled bunch instability driven by the fundamental RF cavity impedance.

The critical cavity parameter here is $\frac{R}{Q}\cdot\frac{\omega_c}{V}$ (equal to the charge stored on the equivalent C of the cavity) which should be minimized if such effects are to be avoided. For this purpose wide aperture cavities are interesting (small R/Q), especially superconducting (large V, without sacrificing Q). Another solution is to increase the stored energy (and consequently the equivalent stored charge) with additional resonators coupled to the accelerating cavity [3].

5. THE CASE OF MISSING BUNCHES

We assume again $\delta_0 \cong 0$, but consider a gap in the train of bunches, corresponding to a number of consecutive buckets being empty. Beam loading is, in this case, periodic at the revolution frequency and will result in a periodically modulated RF waveform.

We shall try to find directly the periodic equilibrium situation in the very simple case $\phi_s = 0$ (no acceleration) and a lossless cavity ($R' = \infty$). In the rotating frame at ω, the RF voltage vector $\vec{V}$ evolves under the effects of: a) cavity detuning $\Delta\omega$, b) beam current (we assume here that the generator current is unmodulated at f_{rev} harmonics).

Cavity detuning produces a phase variation $d\varphi_1$ of $\vec{V}$:

$$d\varphi_1 = \Delta\omega \ dt \tag{26}$$

Beam current $i_b(t)$ produces an additional quadrature $(\phi_s = 0)$ cavity voltage

$$dv = \frac{1}{2}\frac{R}{Q}\omega \ i_b(t)dt \tag{27}$$

The total phase variation:

$$d\varphi = \left(\Delta\omega - \frac{1}{2}\frac{R}{Q}\omega\frac{i_b(t)}{V}\right)dt \tag{28}$$

integrated over one turn must be zero for a stationary situation. As the amplitude of V remains constant (only quadrature contributions) it follows:

$$\Delta\omega_0 = \frac{1}{2}\frac{R}{Q}\frac{\omega}{V}\frac{1}{T}\int_0^T i_b(t)dt = \frac{1}{2}\frac{R}{Q}\frac{\omega}{V}\bar{i}_b \tag{29}$$

where T is the revolution period and $\bar{i}_b$ the average RF component of the beam current.

One obtains again Eq. (25) as expected, from which the phase variation of the RF voltage can be expressed as:

$$\varphi(t) = \int_0^t d\varphi = \frac{1}{2}\frac{R}{Q}\frac{\omega}{V}\int_0^t \left(\bar{i}_b - i_b(t)\right)dt \tag{30}$$

In the simple, but very useful case of a continuous beam of constant intensity, with a gap of length $(T - t_0)$, one finds a linear phase variation (Fig. 7) with a maximum excursion:

$$\Delta\varphi_{max} = \frac{1}{2}\frac{R}{Q}\frac{\omega}{V}\bar{i}_b(T - t_0) = \Delta\omega_0(T - t_0) \tag{31}$$

Interestingly enough Eq. (31) shows no resonance even if $\Delta\omega = 2\pi / T$ as one would expect. This is because the AM modulation of the beam current is exactly compensated, for the side band falling on the cavity resonance, by the phase modulation given by Eq. (31). This result can also be obtained in a more complete way by frequency analysis [4, 5]. For optimum detuning $\Delta\omega_0$ one obtains:

$$\frac{p_v}{a_b} = -\frac{\Delta\omega_0(s + \sigma - \Delta\omega_0 \tan^3\phi_s)}{(s+\sigma)^2 - \Delta\omega_0^2 \tan^2\phi_s} \tag{32}$$

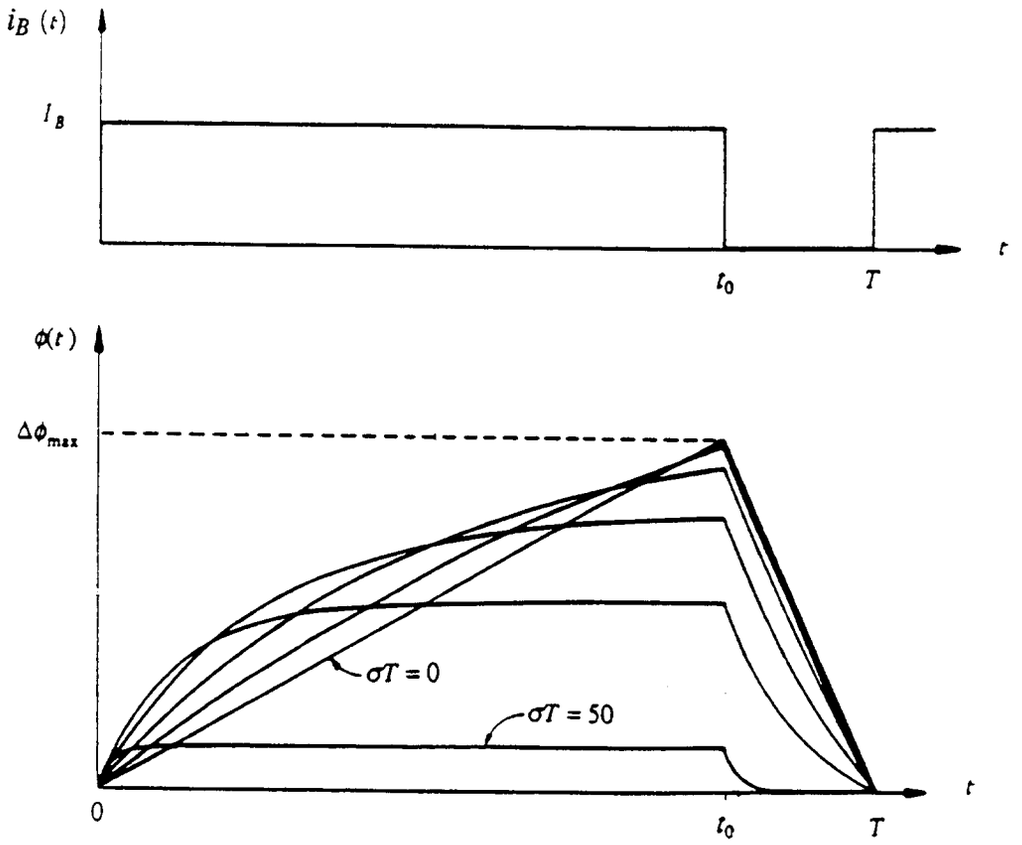


Fig. 7 Phase modulation of V_{RF} due to a gap in the bunch train for various values of the cavity bandwidth

$$\frac{a_v}{a_b} = -\frac{\Delta\omega_0 (s + \sigma + \Delta\omega_0 \tan\phi_s)\tan\phi_s}{(s+\sigma)^2 - \Delta\omega_0^2 \tan^2\phi_s} \tag{33}$$

where p_v, a_v and a_b are the phase and amplitude modulation indices of V and a_b the beam modulation amplitude due to the gap. s is the Laplace variable and $\sigma = \omega_c / 2Q_L$ the half cavity bandwidth. In the case $\phi_s = 0$, the amplitude modulation of V disappears; only remains the phase modulation given by:

$$\frac{p_v}{a_v} = -\frac{\Delta\omega_0}{s+\sigma} \tag{34}$$

and illustrated in Fig. 7 (low pass filter characteristics).

Phase modulation along the batch may have important consequences, in particular in the case of colliding beams: the azimuthal position of the beam crossing point depends on the bunch positions in the trains and may become excessive [4, 5]. Here again a cavity with a large stored energy (small $\Delta\omega_0$) is interesting. Alternatively one can also increase the cavity bandwidth (in particular with RF feedback, section 8.3), but at the expense of more RF power.

6. THE CASE OF A TRAVELLING-WAVE STRUCTURE

It is known that in a long chain of coupled resonators travelling waves can propagate within some frequency limits, i.e. passbands of the structure. In the travelling mode of operation, the structure is terminated by its characteristic impedance and behaves like a transmission line (Fig. 8). At synchronism, the phase velocity v_φ of the wave equals the particle velocity v_p, giving maximum voltage seen by the beam, like an RLC circuit at resonance.

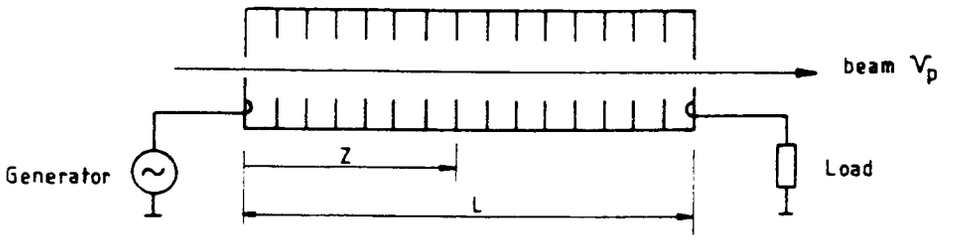


Fig. 8 Schematic of a travelling-wave structure

For a single-bunch passage, it is usually possible to neglect the cavity coupling as the energy transfer from cell to cell is much slower than the bunch velocity $\left(v_g << v_p;\ v_g\text{: group velocity}\right)$. The previous analysis can therefore be applied to the quasi uncoupled resonators. It is generally applied also for the standing-wave mode of operation of multicell cavities, which are non-terminated structures [6]. However, for a repetitive train of many bunches, the RLC equivalent circuit model would fail in the travelling-wave mode because the waves excited by previous bunch passages also propagate along the structure.

For instance, at exact synchronism $\left(v_\varphi = v_p\right)$, the waves excited in each cell by the beam passage add linearly in the forward direction, and, on average, cancel in the reverse direction, in a forward travelling-wave structure. In other words, the decelerating electric field E_z is simply proportional to the distance along the structure counted from the feed point.

If the synchronism is not perfect, we must introduce a phase factor $\exp\ j\left(\omega t - \beta_\varphi z\right)$ for each individual wave, where $z = v_p t$ and β_φ is the wave propagation constant, with the result that the induced field $E_z(z)$ is proportional to the integral:

$$E_z(z) \sim \int_0^z \exp\ j\left(\omega t - \beta_\varphi z\right) dz = \int_0^z \exp\ j\theta dz \ . \tag{35}$$

Expanding θ around the synchronous point $\left(\omega_0, \beta_0\right)$ one obtains:

$$\theta = \omega t - \beta_\varphi z = \left(\frac{\omega_0 + \delta\omega}{v_p} - \beta_0 - \Delta\beta\right) z = \left(\frac{\delta\omega}{v_p} - \Delta\beta\right) z \ . \tag{36}$$

Introducing $v_g = \delta\omega / \Delta\beta_\varphi$ and the phase slip angle τ defined by:

$$\tau = L\frac{\delta\omega}{v_g}\left(1 - \frac{v_g}{v_p}\right) \tag{37}$$

L being the structure length, one obtains:

$$\theta = -\tau \frac{z}{L} \tag{38}$$

and:

$$\int_0^z \exp\ j\theta\ dz = \frac{1 - \exp\left(-j\frac{\tau}{L}z\right)}{j\frac{\tau}{L}} . \tag{39}$$

In particular, for $z = 0$, the integral vanishes: the beam induced field is zero on the upstream end of the structure (generator side). This is a very important result as it shows that, for a travelling-wave structure, there is no beam loading effect seen by the RF generator, which always remains matched without the need for a circulator. In the case of a backward-wave structure, where the generator is connected to the downstream end of the structure, this result is still valid. Beam loading only changes the field on the load side: not all the generator power goes into the load, some fraction is transferred to the beam.

The total voltage V_b seen by the beam is obtained by integrating the electric field, given by (39) along the structure:

$$V_b \sim \int_0^L \frac{1 - \exp\left(-j\frac{\tau}{L}z\right)dz}{j\frac{\tau}{L}} . \tag{40}$$

It gives finally:

$$V_p = i_b R_2 \frac{L^2}{8}\left[\left(\frac{\sin\tau/2}{\tau/2}\right)^2 - 2j\frac{\tau - \sin\tau}{\tau^2}\right] \tag{41}$$

where the proportionality factor R_2, called the series impedance of the structure, is characteristic of its geometry [7]. Figure 9 shows a plot of Eq. (41) in the complex impedance plane.

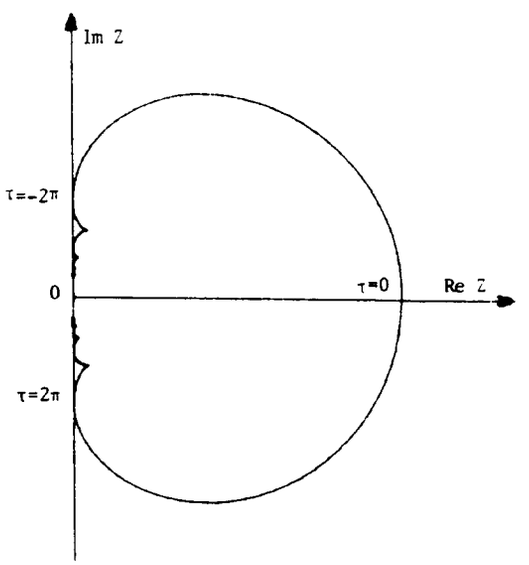


Fig. 9 Impedance seen by the beam of a travelling-wave structure

7. TRANSIENT CORRECTION

Consider again the case of a cavity represented by its RLC equivalent circuit. Even in the case $\delta_0 \cong 0$ (quasi sinusoids) the stationary solution of Section 4, where only the RF frequency component is considered, cannot describe transient situations, when V or i_b change rapidly.

The worst case situation corresponds to a sudden change of $\bar{V}$ (e.g. transition) or $\bar{i}_b$ (injection of a prebunched beam, fast ejection of part of the beam). The resulting unwanted transient must of course be damped for the stationary solution described above to settle down properly, but it must also be short compared with the synchrotron period T_S. This condition will ensure that the effects on the beam such as mismatch and subsequent blow-up, or even loss of particles, will be minimum, or in other words that beam loading will be properly corrected.

We shall now consider the example of a prebunched beam i_b injected into an empty machine. Before injection the servo-tuning keeps $i_t = i_g$ and V in phase. Immediately after injection the new vector i_b destroys the equilibrium, and V changes by a large amount until the tuning loop retunes the cavity to a different value. Unless one uses very fast tuners, which may lead to multiloop stability problems [8], it will take more than a small fraction of a synchrotron period for the tuning loop to settle at its new value, the result being a strong distortion of the longitudinal phase plane.

The only way to maintain V constant during the transient phase of the tuner is to act via the RF power generator which provides a fast control of V. The obvious solution (Fig. 10) is to change i_g into i_g' when the beam is injected. If we make:

$$i_g' = i_g - i_b \tag{42}$$

the total current in the cavity does not change and, at constant tuning, V stays constant.

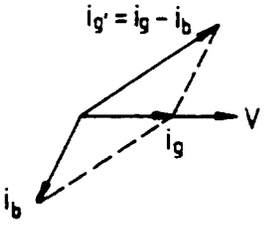


Fig. 10 Correction of beam-loading transient with the power generator

In the simple case of no acceleration, the amplitude of the peak current i_g' which must be delivered by the RF power tube during the transient phase of the tuner, is given by:

$$|i_g'|^2 = |i_g|^2 + |i_b|^2 \tag{43}$$

or, for a high beam loading

$$i_g' \cong i_b \ .$$

The RF generator must be able to deliver the current i_b and sustain the voltage V. This determines the required *installed* power $\left(\frac{1}{2}Vi_b\right)$ for transient beam loading correction, and

more precisely the RF tube characteristics (max V, max i_g') and the transformation ratio between tube and cavity.

In the case of a matched generator (Fig. 11) with a circulator inserted between RF power and cavity, the additional current means also additional *real* power. In the $\phi_s = 0$ case, the peak power $\hat{P}$ required during the transient is given by [4]:

$$\hat{P} = \frac{Z_0}{8}\left(V^2\left(\frac{1}{Z_0}+\frac{1}{R}\right)^2 + \left(\frac{V}{x} - i_b(t)\right)^2 \right) \tag{44}$$

where Z_0 is the line impedance transformed to the gap and x is the reactive part of the cavity impedance

$$x = \frac{1}{2}\frac{R}{Q}\frac{\omega_0}{\Delta\omega} \quad .$$

Assuming $\Delta\omega = 0$ before injection, one can optimise $\hat{P}$ by adjusting the transformation ratio according to:

$$\frac{1}{Z_0^2} = \left(\frac{i_b}{V}\right)^2 + \frac{1}{R^2} \tag{45}$$

One obtains in this case, and for heavy beam loading ($R\, i_b >> V$):

$$\hat{P} \cong \frac{V i_b}{4} \tag{46}$$

Remark that the installed power (which is also real power here) is only half of what would be required without circulator.

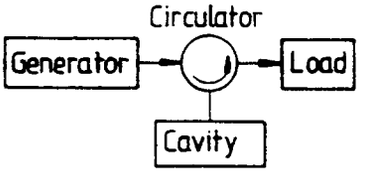


Fig. 11 A circulator to match the RF power generator

However, the correction described above corresponds to the worst case situation. In certain cases it is possible to minimize the required peak power or peak current. In particular, by pretuning the cavity before injection, one can make the two powers, before and after injection, equal and obtain in this case $\hat{P} = |V||i_b|/8$ (for $\Phi_s = 0$) again with a circulator between power source and cavity. One can also reduce the transient on i_b with multiple injections of smaller currents, or by adjusting the bunching factor of the injected beam.

In the above analysis, we assumed that all bunches are submitted to the same RF voltage. This may not be true in the case of unequal filling of the ring which will give a modulation of V at f_{rev} and its multiples. The same analysis applies here: at each "batch" passage transient beam loading must be corrected to make all bunches see the same RF voltage. This effect is particularly important in large machines not only at injection but also at transition. As before, condition (44) is valid in the worst case situation, i_b being now the batch current [4, 9].

8. RF DRIVE GENERATION

During the transient phase of the tuner, we must synthesize i_g' to meet condition (42) and correct for the effect of beam loading at injection or at transition. The same is true in the case of uneven filling of the ring. It obviously implies that i_g' (or the corresponding power $\hat{P}$) is available from the RF generator, otherwise transient beam loading cannot be corrected completely. Various techniques used to generate the proper i_g' will now be examined. As will become apparent in the following, they are closely related to the stability of the entire RF system.

8.1 Amplitude and phase servo loops

The synthesis of i_g in order to keep V constant irrespective of the beam loading can be done with two servo loops (Fig. 12): the first acting on the amplitude of i_g (amplitude loop) controls $|V|$, and the second maintains the relative phase of V and i_b constant through the control of the phase of i_g (phase loop). The cut-off frequency f_c of the loops must be much larger than the synchrotron frequency f_s, which means very strong damping of beam oscillations. This justifies the simplified stability analysis [8] in which the beam transfer function is neglected. The cut-off frequency f_c is obviously limited by the delays in the system, including the cavity bandwidth, but more fundamentally by the revolution frequency f_{rev}. The simple configuration of Fig. 12 with high loop gains cannot correct transient beam loading at f_{rev} and its multiples.

Steady beam loading with its associated cavity detuning could excite mode $n = 0$ (Robinson instability [10]) if it were not heavily damped by the phase loop. However, mode $n = 1$ (one wavelength per turn) which is not damped may show up also due to cavity detuning and must be suppressed by dedicated feedback circuitry acting through the RF cavity itself.

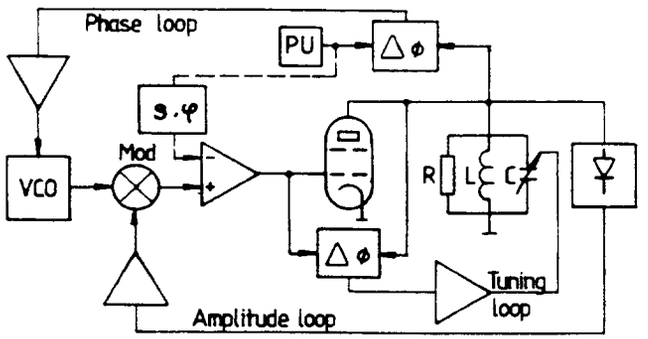


Fig. 12 Tuning, amplitude and phase loops. Feedforward correction (dotted line).

Independent amplitude and phase control of V is a well known technique for proton machines. It works satisfactorily for relatively small beam currents, i.e. when the gap voltage is predominantly determined by the generator current (typically $|i_b| < |i_g|$). For higher beam currents, a variation of the amplitude of i_g, for instance, not only results in a variation of the amplitude of V but also of its phase. In other words, the two loops, which were independent at low beam currents, become coupled together and an unstable behaviour of the system results above a certain beam current threshold. Pedersen's detailed analysis [8], confirmed by experiments on the CERN PS booster, leads to the generalized Robinson stability criterion, valid for $\Phi_s = 0$, optimum detuning, σ large, and assuming a negligible beam response in the frequency range of interest:

$$\frac{|i_b|}{|i_g|} < \sqrt{2 + \frac{f_a}{f_T} + \frac{f_T}{f_a} + \frac{f_p}{f_T} + \frac{f_T}{f_p} + \frac{f_a}{f_p} + \frac{f_p}{f_a}} \tag{47}$$

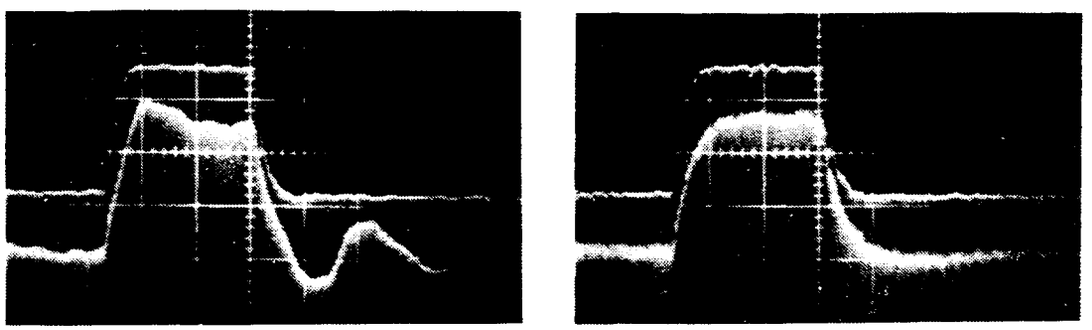
where f_a, f_p and f_T are the unity gain frequencies of the loops (amplitude, phase and tuning respectively). Although the threshold is weakly dependent on the loop cut-off frequencies, it might be dangerous in this configuration to increase the servo-tuner bandwidth.

In the more general case, no analytical simple stability criterion could be found. Simulations [15] have shown that in almost all cases if the quantity $|i_b|/|i_g|$ is kept below 2 stability of the RF system is ensured with some safety margin. A safe design criterion for a new RF system would be $|i_b|/|i_g| < 1$ which can be expressed by the simple rule of thumb: beam induced voltage not larger than RF voltage.

Although it is in principle possible to compensate loop coupling by an additional decoupling circuitry so increasing the instability threshold, much simpler solutions are offered by feedforward correction and RF feedback.

8.2 Feedforward correction

With a pick-up electrode followed by a filter centred at f_{RF}, one can obtain a signal proportional to $-i_b$ independently from the RF system, and generate i_g' (RF drive with beam) according to (32) with a simple adder. Applied to the amplitude and phase servo loops described in section 7.1, the method consists of injecting into the input of the RF amplifier the pick-up signal, with proper amplitude and phase (g, φ), to generate the $-i_b$ current at the gap (Fig. 12). The amplitude and phase loops now act on the quantity i_g, corresponding to no beam loading, instead of i_g', and the cross couplings between loops are removed, as can be shown analytically and experimentally [11] (Fig. 13). As a result, the instability threshold can be considerably increased and, for instance, stable operating conditions have been observed in the CERN PS for $|i_b|/|i_g| \cong 8$ to 10.



(a) (b)

Fig. 13 Transient response of amplitude loop with (b) and without (a) feedforward correction (CERN PS machine). The loop response becomes oscillatory at high intensity (bottom trace) without feedforward correction

The signal corresponding to $-i_b$ does not need to be synthesized with the ultimate precision as it only removes the loop couplings and restores stability. For a varying RF frequency, the pick-up to cavity delay must be continuously adjusted, and the variations in gain and phase of the RF power amplifier (assumed linear) corrected.

Feedforward correction can also be considered as a means to reduce the effective impedance of the cavity seen by the beam. At the RF frequency, the beam induced voltage on the cavity amplifier combination is zero for a perfect correction. From this point of view, high amplitude and phase loop gains at f_S are no longer required to correct beam loading as V is

automatically kept constant by the feedforward compensation. Application of this technique (low loop gains) was, for instance, used on the Brookhaven AGS during adiabatic capture.

It is interesting to mention a variant of the feedforward technique derived from the Alvarez linear accelerator technology. If the generator is a gridded tube (tetrode or triode), its output impedance is high if maximum RF power is to be extracted from the tube. When connected to the cavity by a long line, it fully reflects the beam loading wave travelling from the cavity to the generator. One can choose the length of the line to make the reflected wave cancel the beam induced voltage at the gap, the high impedance of the generator is then transformed into a quasi-short circuit at the cavity. Note that, even with no voltage induced on the gap, the generator sees a mismatched load with beam and must be able to deliver the current under this condition. This technique is in use on the CERN PS 200 MHz RF system, with trombones inserted on the feeder lines of the fixed-tune cavities.

If the pick-up to cavity delay is adjusted to be exactly one turn (T), beam loading cancellation can be achieved, not only at f_{RF}, but also at frequencies $f_{RF} \pm nf_{rev}$. This is relatively easy at fixed RF frequency, for example in the CERN ISR [12], but with modern sampled or digital filters and variable delays it is also possible to follow a varying RF frequency. The overall result is a rapidly changing impedance, ideally zero at frequencies nf_{rev}, but twice as large at intermediate frequencies, $(n+1/2)f_{rev}$, where there are no beam current components (Fig. 14). With a one turn delay and perfect cancellation, the voltage perturbation only lasts T which is small compared with T_S since $Q_s = (f_{rev}T_s)^{-1}$ is usually << 1. In other words the reduction of the magnitude of the cavity impedance at the synchrotron satellites $nf_{rev} \pm mf_S$ is also large (factor $(2\sin m\pi Q_s)^{-1}$) for a small Q_S.

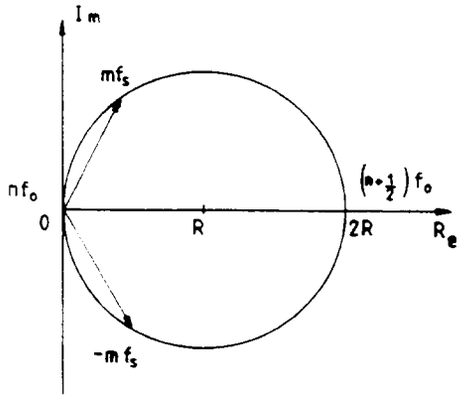


Fig. 14 Residual impedance at synchrotron sidebands for a one turn delay feedforward correction

8.3 RF feedback

We can consider the cavity itself as a beam pick-up tuned at f_{RF} and obtain the $-i_b$ signal from the gap itself. This leads to the configuration of Fig. 15 in which one obviously recognizes a feedback loop built around the RF power amplifier. From the loop equations one obtains:

$$i_g' = i_g - \frac{GZ\ i_b}{1+GZ} \tag{48}$$

which, for $GZ >> 1$ (GZ: loop gain, Z cavity impedance) reduces to Eq. (32):

$$i_g' \cong i_g - i_b$$

i_g being here the generator current with no beam.

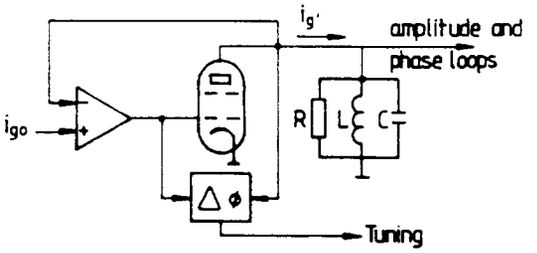


Fig. 15 RF feedback around the power amplifier

The feedback loop automatically generates the correct compensating signal, which is another way of saying that it keeps the controlled parameter V constant. One can consider RF feedback as a means to reduce the output impedance of the RF amplifier, a well known design being the cathode follower with its low output impedance which shunts the cavity.

Even simpler, but of limited efficiency, is the use of a triode instead of a tetrode as the RF power tube, the internal plate to grid feedback reducing the output impedance. In the same way pulsing the DC current of the RF tube or powering a second tube, in parallel [13], has been used to reduce the output impedance of the RF amplifier for short periods.

In the case of Fig. 15, the cavity parameters (pole at $f_{RF}/2Q_L$) and the total delay of the feedback path determine the loop stability. The preamplifiers which are selected for the shortest propagation delay must be located very close to the power amplifier-cavity combination. As an example Table 1 gives the parameters for the CERN PS booster second harmonic system, operating between 6 and 16 MHz [14].

Table 1
Feedback parameters of the CERN PSB second-harmonic system

Preamplifier gain	25 x
Bandwidth	150 MHz
Power	130 W (1 dB compression)
Propagation delay	5 ns
Impedance reduction factor	21 dB at 6 MHz
	14.5 dB at 16 MHz

For a varying RF frequency one could adjust the delay of the return path to keep the 180° phase condition at f_{RF}. However, in many designs, for instance the second harmonic PS booster and the future PS RF system, a wide bandwidth preamplifier is used to keep the total delay short enough to ensure stability over the entire RF frequency range, without programming the phase. In this case it is extremely important to damp the higher resonances of the cavity or to reject the corresponding signals in order to avoid parasitic oscillations of the feedback system at high frequencies.

The RF feedback technique is very attractive since it reduces the effective impedance of the cavity not only at the RF frequency but also over a large bandwidth. This feature is particularly helpful to avoid self-bunching instabilities in storage rings for debunched beams and was used at the CERN ISR and AA for instance.

In such a feedback system the total phase slip should be less than about $\pm\pi/4$ over the unity gain bandwidth $2\,\Delta\omega$ of the open loop system, giving the condition [15]:

$$\Delta\omega = \pi / 4\tau \tag{49}$$

where τ is the overall delay in the feedback path. For a fixed tuned cavity and a small detuning angle, the cavity impedance (RLC approximation) far from the ω_c resonance is given by $Z = R / 2jQ_L(\Delta\omega / \omega_c)$. The overall loop gain, GZ, at the $\pm\Delta\omega$ points is of the order of unity: this gives an upper limit for GZ and a minimum value of the impedance seen by the beam, $R_{\min}$, given by:

$$R_{\min} \cong \frac{2}{\pi}\tau\frac{R\omega_c}{Q} = 4\frac{R}{Q}f_{RF}\tau \tag{50}$$

Equation (39) shows that the ultimate performance of wideband RF feedback only depends on τ and the cavity geometry (R/Q parameter).

When a gap is present in the circulating beam, RF feedback will automatically correct for it and, in the limit of very high gain suppress phase and amplitude modulations given by Eqs. (32) and (33). This may be undesirable if the corresponding RF power is not available. In this case, one may program in phase the RF reference (i_g in Fig. 14) as the expected phase production and greatly reduce the power requirements [4].

If a servo tuner is used in conjunction with RF feedback, it may be necessary to control it by the normalized reactive power of the amplifier [16] instead of the classical phase detection between RF drive and cavity voltage.

RF feedback is now widely used in high intensity RF systems [5, 17-20]. The dynamic properties and stability of the RF system are critically dependent on the delay of the power amplifier. This requirement has led, in particular to the development of short delay high power klystrons for such applications. RF feedback opens the way to using superconducting cavities also in high intensity applications. As a practical advantage, RF feedback can easily be set up without the presence of beam, contrary to the feedforward technique; it finally makes the behaviour of the RF system far less sensitive to the beam current.

8.4 The RF feedback with long delay [5, 21]

In large RF systems, the CERN SPS for instance, long delays may be unavoidable and the conventional RF feedback would have a too restricted bandwidth, much smaller than the cavity bandwidth itself in the SPS case. Transient beam loading at multiples of f_{rev} would not be corrected, leading to phase oscillations of fractions of the beam and possible coupled bunch instabilities.

In order to solve the problem, we observe that a large gain G is only needed in the vicinity of the revolution frequency harmonics where beam current components exist. Outside these bands, the phase rotation due to the excessive delay will be unimportant if G can be made small enough. With a return path transfer function having a comb-filter shape with maxima at every f_{rev} harmonic, this condition can be satisfied. In addition, the overall delay of the system must be extended to exactly one machine turn (T) to ensure a zero phase at the $f_{RF} + nf_{rev}$ frequencies.

The comb filter transfer function (Fig. 16) is of the form:

$$H(j\omega) = \frac{G_0}{1 - K\exp(-j\delta\omega T)} \tag{51}$$

where G_0 and K are constants ($0 < K < 1$).

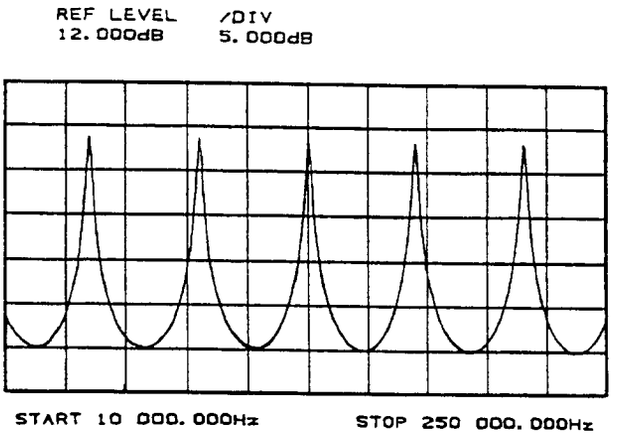


Fig. 16 Comb-filter transfer function $K = 7/8$, $N = 462$

Combined with the one turn delay (transfer function: $\exp(-j\delta\omega T)$), the overall open loop transfer function becomes:

$$G(j\omega)Z(j\omega) = \frac{G_0 Z(j\omega)}{\exp(j\delta\omega T) - K} \tag{52}$$

represented in the complex plane by a circle for a slowly varying $Z(j\omega)$ as shown in Fig. 17. The complex plane origin is encircled and therefore the gain of the system is limited by the stability condition. In the vicinity of the cavity resonance, where Z is maximum and real (note that for a travelling wave structure Z is always real [4], the circle crosses the negative real axis at a distance $-G_0 Z / (1+K)$ from the origin.

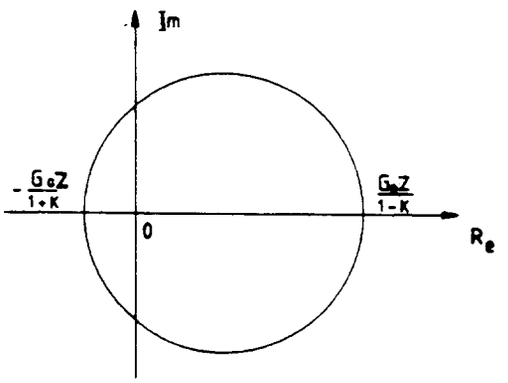


Fig. 17 Open-loop transfer function for RF feedback with long delay

Stability obviously requires that $|G_0 Z| < 1 + K$, and it can be shown that this condition is also sufficient even outside resonance for an RF cavity approximated by a single RLC equivalent circuit.

For Z real, the apparent impedance of the cavity Z':

$$Z' = Z\frac{\exp(j\delta\omega T) - K}{\exp(j\delta\omega T) - K - G_0 Z} \tag{53}$$

is real for frequencies:

$$f_{RF} + nf_{rev}; \quad Z' = Z\frac{1-K}{1-K+G_0Z} << Z \tag{54}$$

and:

$$f_{RF} + \left(n+\frac{1}{2}\right)f_{rev}; \quad Z' = Z\frac{1+K}{1+K-G_0Z} \tag{55}$$

To stay at a reasonable distance from the stability limit, take for instance $G_0Z = (1+K)/2$. This gives, at frequencies $f_{RF} + \left(n+\frac{1}{2}\right)f_{rev}$, $Z' = 2Z$ as in the case of feedforward correction, whereas for the revolution frequency harmonics one obtains:

$$Z' \cong Z(1-K) \tag{56}$$

for $(1 - K) << 1$.

By making K close to unity, RF feedback approaches the theoretical performance of the feedforward correction but with all the inherent advantages of closed loop systems; in particular no critical adjustments are needed. Similarly, the time response of the RF feedback is entirely determined by the one turn delay as in the feedforward case. Note that the unity gain frequency of the servo in this case is of the order of $f_{rev}/2$.

The residual impedance at the synchrotron sidebands is approximately the same as for a one turn delay feedforward correction (for $K \cong 1$ and $G_0Z \cong 1$); its phase changes sign at each nf_{rev} harmonic resulting in a rotation of the complex synchrotron frequency shift curve. The coupled-bunch, cavity-driven, instability thresholds must be obtained numerically [18].

Except for relatively small machines with fixed RF frequency, long delay feedforward or feedback techniques could only be envisaged with the help of modern signal processing technology, i.e. sampled or digital filters. The digital comb filter is derived from the well known first-order low-pass recursive filter shown in Fig. 18. With a sampling frequency Nf_{rev} locked to a subharmonic of the RF frequency, the theoretical bandwidth of the filter is $Nf_{rev}/2$, corresponding to $N/2$ maxima in the comb filter response (N = 462 in the SPS design). Implementation of the one-turn delay is straightforward in digital technology with a memory (R.A.M. or first-in first-out type), or with more modern arithmetic processors.

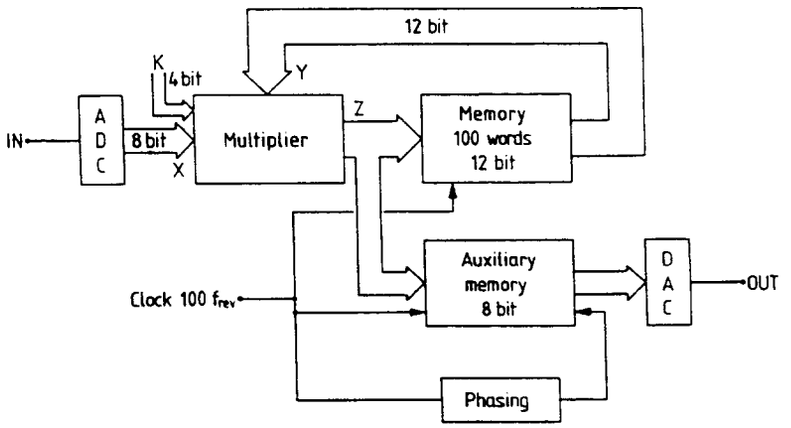


Fig. 18 The digital filter and delay

It is also possible to combine short and long delay RF feedbacks on the same cavity. The additional impedance reduction at the f_{rev} harmonics is however limited to factors 3 to 5, because of the non-linear phase response of the cavity with short delay feedback. The SPS superconducting cavities [19] are now equipped with combined RF feedback circuits to cope with the highest beam intensities.

In e^+e^- machines where Q_S is not so small, one can also use more complex periodic filters, the objective being to reduce the equivalent cavity impedance preferentially at the synchrotron sidebands $f_{rev} \pm f_s$ to avoid coupled bunch instabilities [5].

The speed of the various elements, limited by the cycle time (T/N), may become very critical requiring the fastest A-D converters (flash converters), memories and multipliers (parallel multipliers).

The RF signals may have to be translated in frequency to be conveniently processed. Coherent mixing with separate channels for in-phase and in-quadrature components is necessary to reject the unwanted image frequencies (measured rejection > 35 dB), and to make the overall electronic chain look a linear system. For a varying RF frequency the correct phase can even be maintained with an artificial delay inserted between the output and input local oscillators as in Fig. 19.

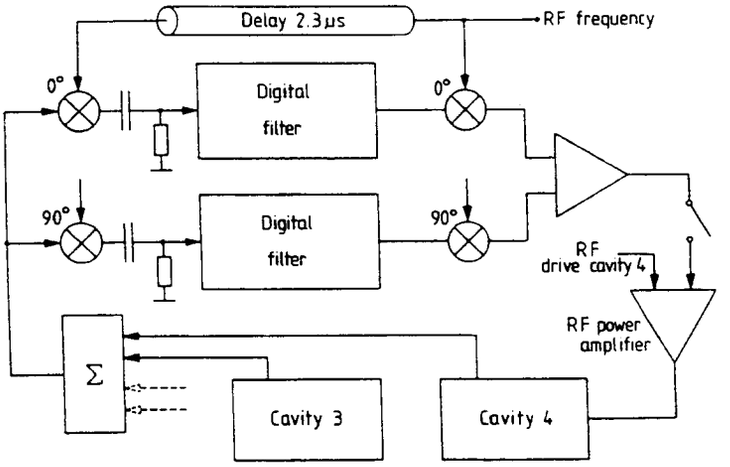


Fig. 19 Layout of the SPS RF feedback system

* * *

REFERENCES

[1] P. Wilson, CERN ISR-TH/78-23 (1978).

[2] P. Wilson, IXth Int. Conf. on High Energy Accelerators, SLAC, Stanford (1974), p. 57.

[3] K. Akai and Y. Yamazaki, Proc. EPAC 94, London, World Scientific Singapore, p. 2131.

[4] D. Boussard, RF power requirements for a high intensity proton collider Parts 1 and 2. CERN SL/91-16 (RFS).

[5] F. Pedersen, RF cavity feedback, CERN/PS 92-59 (RF).

[6] H. Henke, CERN ISR-RF/78-22 (1978).

[7] G. Dôme, 1976 Proton Linac Conference, Chalk River, Canada, p. 138.

[8] F. Pedersen, IEEE Trans. Nucl. Sci. NS-22 (1975), p. 1906.

[9] D. Boussard, IEEE Trans. Nucl. Sci. NS-32 (1985), p. 1852.

[10] K.W. Robinson, CEA report CEAL - 1010 (1964).

[11] D. Boussard, CERN/SPS/ARF Note 78-16 (1978).
[12] H. Frischolz and W. Schnell, IEEE Trans. on Nucl. Sci. NS-24 (1977), p. 1683.

[13] G. Gelato et al., IEEE Trans. on Nucl. Sci. NS-22 (1975), p. 1334.

[14] J.M. Baillod et al., IEEE Trans. on Nucl. Sci. NS-30 (1983), p. 3499.

[15] D. Boussard, Design of a ring RF system, CERN SL/91-2 (RFS).
[16] F. Pedersen, IEEE Trans. on Nucl. Sci. NS-32 (1985), p. 2138.

[17] R. Garoby et al., RF system for high beam intensity acceleration in the CERN PS, Proc. 1989 IEEE Part. Acc. Conf. Chicago, Argonne Nat. Lab.

[18] R. Boni and A. Gallo, Proc. EPAC 94 London, World Scientific Singapore, p. 1862.

[19] H. Schwarz and R. Rimmer, ibid, p. 1882.

[20] D. Boussard, H.P. Kindermann and V. Rossi, EPAC Rome, June 1988, World Scientific, Singapore, p. 985.

[21] D. Boussard and G. Lambert, IEEE Trans. on Nucl. Sci. NS-30 (1983), p. 2239.

HIGH-FIELD ELECTRON LINACS

J. Le Duff
Laboratoire de l'Accélérateur Linéaire, Orsay, France

Abstract
High-field electron linacs are considered as potential candidates to provide very high energies beyond LEP. For almost twenty-five years little improvement has been made on linac technologies as they have been mostly kept at low and medium energies to be used as injectors for storage rings. Today, both their efficiency and their performances are being reconsidered, and for instance the pulse compression scheme developed at SLAC and introduced to upgrade the energy of that linac is a first step towards a new generation of linear accelerators. However this is not enough in terms of power consumption and more development is needed to improve both the efficiency of accelerating structures and the performances of RF power sources.

1. INTRODUCTION

After introducing briefly the needs for higher gradient electron linear accelerators by showing that simple extrapolation of present technologies would fail in trying to reach much higher energies, we shall review different ways of improving these conventional techniques and their related problems. Since high gradient means also high RF power we shall also present RF power source progress towards higher peak power and higher efficiencies.

Notice that the present contribution was originally written in 1986 and has only been updated for the purpose of these new proceedings. The intent remains to provide some ideas about the constraints related to high gradient linacs rather than following the exact evolution of the performance of existing linacs or the R&D worked out in different laboratories. Neither new technologies such as 2-beam accelerators and superconducting linacs, nor new accelerating concepts such as laser-plasma or wake-field accelerators, are discussed here.

2. EXTRAPOLATION OF PRESENT TECHNOLOGIES

There are at least three large electron/positron linacs operating in the world (Table 1) either as injectors for storage rings or as high-energy physics facilities.

Table 1
Large linacs in the world

Laboratory	Energy (GeV)	Accelerating length (m)	Overall length (m)
LAL/Orsay	2.3	230	360
KEK/Tsukuba	2.5	320	400
SLAC/Stanford	24	3050	3350

The accelerating gradient lies in the range of 8 to 14 MeV/m. Recently the SLAC linac has been upgraded to 50 GeV with new klystrons followed by a pulse compression system [1,2]. In this last mode of operation the accelerating gradient is as much as 17 MeV/m. Two bunches, electrons and positrons, are simultaneously accelerated, then transferred in the two arms of a circular transport system in such a way that they collide once at a given location. This is the first linear collider (SLC) operating in the world, at an energy level comparable with LEP

stage 1. It will serve as a test bed for future linear colliders as well as for studying the intermediate boson Z_0.

In order to reach many hundred GeV or a few TeV in the centre of mass with electrons and positrons, it appears that linacs are better suited than storage rings since circular machines would lead to enormous power being radiated in the bends; remember that in LEP, operating at 100 GeV per beam, each particle will lose 2.8 % of its energy per turn. Clearly the SLC scheme, with a single linac plus a circular transport system, will be also avoided for higher energies and future linear colliders will consist of two linacs firing against each other.

Consider now a first step in increasing the linac energy, by roughly one order of magnitude, using SLAC technology for the accelerating sections. Table 2 gives the resulting constraints and three possible schemes have been considered, knowing that :

$$E_{\text{acc}} \propto \left[P_{\text{input}}\right]^{1/2}$$

$$P_{\text{a.c.}} \propto f_{\text{rep}} N_K \eta_K^{-1} P_K$$

where E_{acc} is the accelerating gradient, P_{input} the input RF power at the structure, f_{rep} the linac repetition frequency, N_K, P_K, η_K being respectively the total number of klystrons, the peak power and the efficiency of each klystron.

Table 2
Extrapolation of SLAC up to 300 GeV

	SLAC*	Super SLAC (1)	Super SLAC (2)	Super SLAC (3)
E [GeV]	30	300	300	300
L [km]	3	30	3	3
N_K	240	2400	240	960
P_K [MW]	38	38	3800	950
E_{acc} [MV/m]	10	10	100	100
f_{rep} [Hz]	180	180	180	180
sections/klystron	4	4	4	1
pulse length [μs]	2.5	2.5	2.5	2.5
$P_{\text{a.c.}}$ [MW]	13.7	137	1370	1370

*Note that for several years SLAC has routinely operated at 50 GeV beam energy with new 65 MW klystrons and a pulse power compression system (SLED).

Clearly a higher gradient keeps the linac length to a reasonable value but introduces strong constraints on the power. For instance new power sources need to be developed at the level of 1 GW peak. The state of the art in operational klystrons is 65 MW peak with 3.5 μs RF pulse length which can be compressed to increase the effective peak power by a factor 3, as used on SLC [3].

Consider another example of a high gradient linear collider using the CERN site. On Fig. 1 two linacs have been drawn along a LEP diameter as if the circumference of LEP was the ultimate possible size for a new accelerator (this is not a fact). Table 3 shows the possible stages in energy that can be considered for this example [4]. The first case uses conventional technologies such as iris-loaded travelling-wave structures and 50 MW, 5 μs klystrons followed by a pulse compression system. However the power consumption has been minimized by matching the accelerating structures to the pulse compressor (see next section), but even so it is too high to compete with the LEP storage ring at the same energy. The second

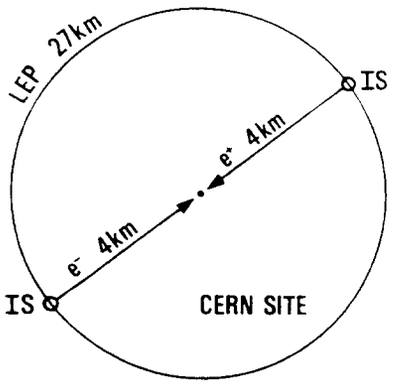


Fig. 1 LEP site with a linear collider

case uses an accelerating gradient of 125 MeV/m which has been already reached on experimental structures but seems to be realistic only at RF frequencies much higher than 3 GHz. However it would work if one could use 1 GW, short-pulse RF power sources which do not exist yet. In both cases a linac repetition rate of 1000 Hz has been considered.

Table 3
Possible stages for a linear collider along a LEP diameter

Gradient (MV/m)	Energy/beam (GeV)	a.c. Power/linac (MW)	Comments
25	100	200	Conventional
125	500	1500	Need for new power sources
250	1000	-	Need for new power sources and new structures
1250	5000	-	New accelerating methods

Improvements on the power consumption may come from improvements in the efficiency of accelerating structures and also from some tricks such as for instance the use of pulse current trains that can lower the repetition rate for the same luminosity [5]. However, designers now consider that to be realistic the a.c. power should not exceed 100 MW.

3. RF COMPRESSION SCHEME

3.1 Present situation

Although this scheme can be considered as an already existing technique [2], it is worthwhile recalling the principles since one can expect to improve the efficiency of the system in the near future.

The pulse compression is schematically represented on Fig. 2; the first part of the long pulse from the klystron is stored in a couple of low loss cavities. At a given time t_1 the input signal to the klystron is rapidly π shifted so that the energy is now reflected at the entrance of the storage cavities and directly goes to the structure. In addition the stored energy flows out of the cavities and also goes to the structure making the peak energy during the time interval [t_1, t_2] much higher than it would have been if the structure had been fed directly. The method can be either used to increase the energy of an existing linac, or to save on the total number of power sources for a given output energy.

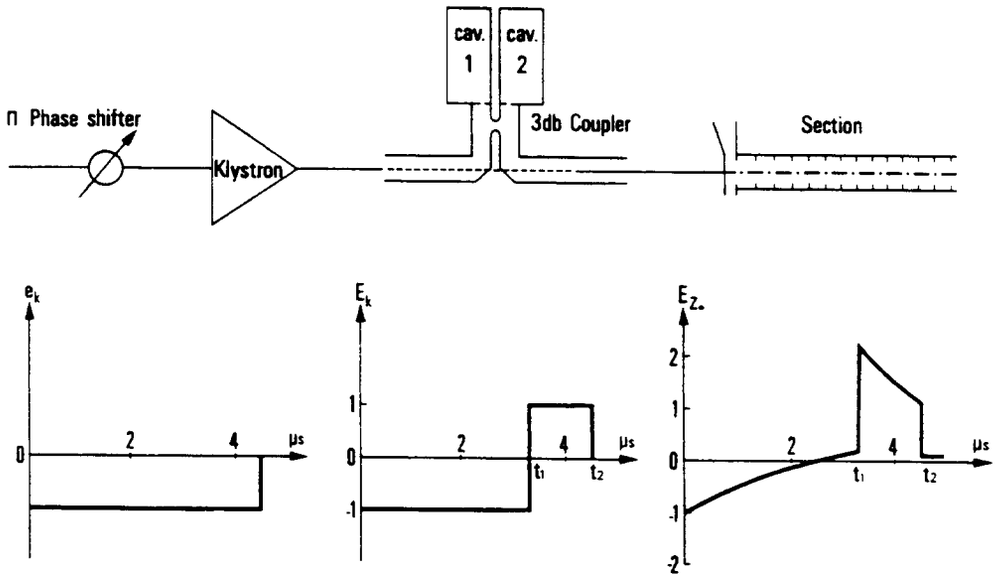


Fig. 2 Pulse-compression scheme

SLED (SLAC Energy Doubler) [2]

Storage cavities are placed between the klystrons and the accelerating structures. The RF pulse length is 5 μs and since the filling time of the structure is 0.8 μs one can adjust the switching time such that $t_2 - t_1 = 0.8$ μs. However the compression scheme has a poor efficiency and the maximum improvement factor on the peak power that one can expect is about 3 leading to an improvement factor $\sqrt{3}$ on the accelerating gradient.

LIPS (LEP Injector Power Saver) [6]

The scheme is used to reduce the number of klystrons by a factor 2 as shown on Fig. 3. Instead of feeding two structures from a single klystron, one can feed four structures with the same total beam energy if the system is adjusted to increase the effective peak power by a factor 2. This improvement factor is lower than the previous one and can be obtained from a 3.5 μs klystron pulse length, the filling time of the LEP injector Linac (LIL) structures being 1.2 μs. It will be seen later that the improvement factor mainly depends on the initial klystron pulse length when all other parameters, such as for instance the filling time of the accelerating structures, have been optimized.

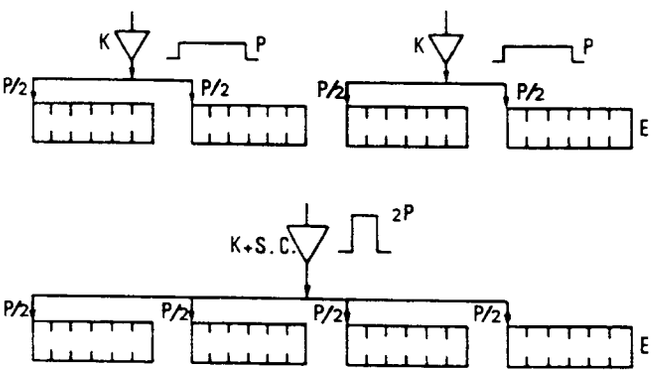


Fig. 3 The LIPS scheme

Up to here the maximum improvement factor one can reach with a conventional electron linac, by adding a compression system, is of the order of 1.7 and from there existing linacs might be able to operate with an accelerating gradient of the order of 35 MV/m, taking into account also the up-to-date klystrons and assuming that each klystron feeds a single structure. As will be seen in the next section, an additional improvement factor can be obtained by optimizing the parameters of the accelerating structures to match properly the compression system.

3.2 Optimization of TW accelerating structures for SLED operation [7]

The RF pulse shape due to the compression scheme, worked out by Z.D. Farkas et al. [2], is shown on Fig. 4. There are two regions of interest: region 1 which corresponds to a continuous increase of the stored energy in the cavities and region 2 which, after a π phase shift at the klystron input occuring at time t_1, corresponds to a relatively fast decay of the stored energy in these cavities to the benefit of the accelerating field.

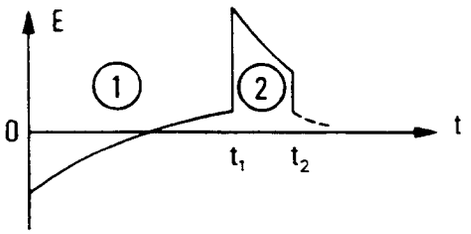


Fig. 4 Pulse shape due to compression

For a unit rectangular klystron pulse the combined field entering the accelerating section is:

$$E_1(t) = (\alpha - 1) - \alpha\ e^{-t/\tau_c} \quad \text{for} \quad 0 < t < t_1$$

$$E_2(t) = \gamma\ e^{-\frac{t-t_1}{\tau_c}} - (\alpha - 1) \quad \text{for} \quad t_1 < t < t_2$$

where

$\tau_c = 2Q_c / \omega(1+\beta)$ is the filling time of the storage cavities
Q_c is the unloaded quality factor of the storage cavities
β is the coupling coefficient of the storage cavities
$\alpha = 2\beta/(1+\beta)$
$\gamma = \alpha\left(2 - e^{-t_1/\tau_c}\right)$.

For a given peak power P from the klystron (rectangular pulse) the accelerating field at the input of the section would be:

$$E = E_0 \cdot E_{1,2}$$

$$E_0 = \left(P \frac{\omega r_0}{v_{g0} Q} \right)^{1/2}$$

where r_0 , v_{g0} , Q are respectively the shunt impedance per unit length, the group velocity and the quality factor of the first accelerating cell.

In a constant impedance structure all the cells are identical, and hence r, v_g , Q will remain constant along the structure. Due to power dissipation in the cells the amplitude of the propagating field will decrease exponentially. At a given azimuth z the field becomes:

$$E(z,t) = E_{1,2}[t - \Delta t(z)]e^{-(\omega/2v_g Q)z}$$

where index 1,2 refers to the two different time intervals as previously defined. Here again the expression needs to be multiplied by E_0 for a given peak power P from the klystron. $\Delta t(z)$ is the wave propagation time from the origin up to z:

$$\Delta t(z) = \int_0^z \frac{dz}{v_g(z)} = \frac{z}{v_{g0}} \ .$$

It looks interesting to use the normalized variable $z' = \frac{z}{L}$ where L is the length of the structure. Then:

$$\Delta t = \tau_a z'$$

with

$$\tau_a = \frac{L}{v_{g0}} \ .$$

Depending whether the time t - Δt appears to be below or above t_1, the field E_1 or E_2 should be used. That tells us that a field discontinuity will appear at some location z_1' in the structure such that:

$$t - \Delta t = t_1$$

or

$$t - t_1 = \tau_a z_1' \ .$$

If $z_1' < 1$ the energy gain along the structure is the contribution of two field integrals:

$$V(t) = \int_0^{z_1'} E_2[t - \Delta t(z')]e^{-\frac{\omega\tau_a}{2Q}z'} dz' + \int_{z_1'}^1 E_1[t - \Delta t(z')]e^{-\frac{\omega\tau_a}{2Q}z'} dz'$$

where now t represents the time at which the particle traverses the structure (the transit time of the particle is negligible compared to the filling time of the structure).

Let us call V_1 and V_2 the integrals relative to E_1 and E_2. One gets:

$$V = V_1 + V_2$$

$$V_1(z_1') = -(\alpha - 1)\frac{T_1}{\tau_a}\left[e^{-\frac{\tau_a}{T_1}} - e^{-\frac{\tau_a}{T_1}z_1'}\right] - \alpha\frac{T_2}{\tau_a}e^{-\frac{t_1}{\tau_c}}\left[e^{\frac{\tau_a}{T_2}-\frac{\tau_a}{\tau_c}z_1'} - e^{-\frac{\tau_a}{T_1}z_1'}\right]$$

$$V_2(z_1') = (\alpha - 1)\frac{T_1}{\tau_a}\left[e^{-\frac{\tau_a}{T_1}z_1'} - 1\right] + \gamma\frac{T_2}{\tau_a}\left[e^{-\frac{\tau_a}{T_1}z_1'} - e^{-\frac{\tau_a}{\tau_c}z_1'}\right]$$

with:

$$\frac{1}{T_1} = \frac{\omega}{2Q}$$

$$\frac{1}{T_2} = \frac{1}{\tau_c} - \frac{\omega}{2Q} .$$

It is interesting to look at the behaviour of the function $V(z_1')$ in the interval $0 < z_1' < 1$. It can be shown numerically that for each value of z_1' there is a value of β which maximizes the energy gain. This has been taken into account in the plots of Fig. 5 where it appears that the maximum energy gain corresponds to $z_1' = 1$, which means that the beam should enter the structure at time $t = t_2 = t_1 + \tau_a$ and that the width of the compressed pulse must be equal to the filling time of the structure.

The study will continue by considering only this optimum case for which:

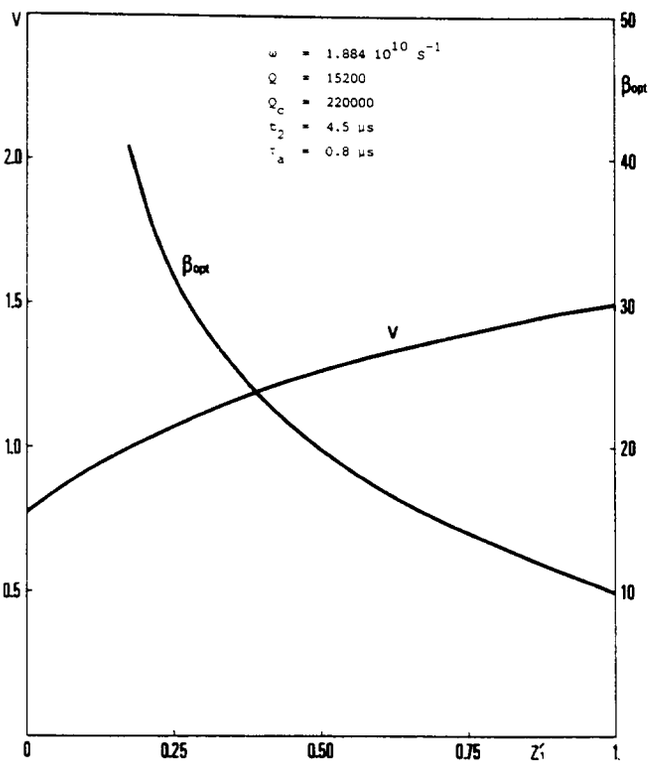


Fig. 5 Multiplication factor for a constant impedance structure as a function of the beam timing

$$V_1 = 0$$
$$V = V_2\left(z_1' = 1\right) .$$

This leads to:

$$V_M = (\alpha - 1)\frac{T_1}{\tau_a}\left[e^{-\frac{\tau_a}{T_1}} - 1\right] + \gamma\frac{T_2}{\tau_a}\left[e^{-\frac{\tau_a}{T_1}} - e^{-\frac{\tau_a}{\tau_c}}\right] .$$

However this is not the exact energy multiplication factor since for a unit pulse entering a constant impedance structure the energy gain over a unit length is:

$$V_0 = \frac{T_1}{\tau_a}\left(1 - e^{-\frac{\tau_a}{T_1}}\right) .$$

Hence the real multiplication factor is the ratio V_M/V_0. For each value of τ_a there is a value of β, hence a value of τ_c, which maximizes this multiplication factor as seen on Fig. 6.

A similar treatment for the case of a constant gradient structure would show that the efficiency of this type of structure is either the same, at low filling time, or slightly less, at high filling time, than the efficiency of a constant-impedance structure. Hence we shall proceed with constant-impedance structures in the following paragraphs.

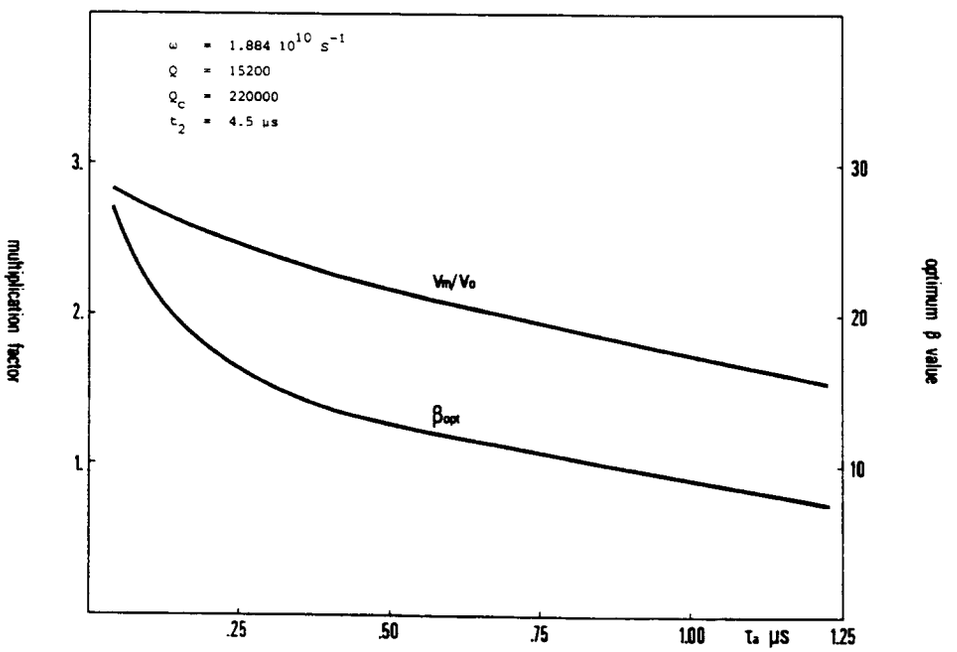


Fig. 6 Multiplication factor versus the filling time of a constant impedance structure

It has been seen that for a given structure length there was an ensemble of optimum values for β, τ_c and τ_a which produce the correct matching between the SLED pulse and the accelerating structure. It is interesting to look in more detail at the performances of these structures versus different parameters, like the settings of the storage cavities (Q_c, β), the length

and the aperture of the accelerating structures, and the width of the direct peak power pulses from the klystrons.

For a constant impedance structure, fed by a klystron peak-power pulse P, t_2, through a couple of storage cavities with a π phase shift at time $t_1 = t_2 - \tau_a$, the energy gain is :

$$V = \left(P \frac{R\omega}{Q} \tau_a \right)^{1/2} \left\{ \frac{\gamma}{\tau_a} \left(\frac{1}{\tau_c} - \frac{\omega}{2Q} \right)^{-1} \left(e^{-\frac{\omega\tau_a}{2Q}} - e^{-\frac{\tau_a}{\tau_c}} \right) + (\alpha - 1) \frac{2Q}{\omega\tau_a} \left(e^{-\frac{\omega\tau_a}{2Q}} - 1 \right) \right\}$$

where $R = rL$ is the total shunt impedance of the structure, and $\tau_a = L/v_g$ its filling time.

The fact that for a given length there is an optimum value for τ_a means that there is an optimum value for v_g, hence for the iris aperture $2a$ of the structure. To illustrate this point let us consider the cell characteristics of the LEP injector linacs (LIL) [8] which operate at 3 GHz in the $2\pi/3$ mode:

$Q = 15200$
$r = 86\text{-}3.6\ (2a)^2$
$v_g/c = (2a)^{3.23}/891$

where $2a$, the iris diameter, is expressed in cm while the shunt impedance r is in MΩ/m.

Figure 7 shows the evolution of the RF performances versus the iris diameter, for different structure lengths. As the length decreases the iris diameter also decreases in order to get the maximum gain corresponding to the right matching value for τ_a . In all cases β and τ_c have been optimized.

The maximum energy gains obtained for each structure length are plotted on Fig. 8 as well as the corresponding values of τ_a and τ_c which clearly remain constant. A systematic study of the energy gain as a function of the other parameters, like t_2, Q_c and $Q\left(\frac{r}{Q} = \text{cte}\right)$ leads to the following conclusions:

- neither Q nor Q_c influence the optimum value of τ_a. Both effect slightly the optimum energy gain. The optimum value of τ_c changes with Q_c .

- the optimum value of τ_a changes with the width t_2 of the direct klystron wave. For long pulses one can hold a longer filling time, but that means a smaller aperture for a fixed structure length. An important increase of the energy gain follows an increase of t_2.

- one of the most important features, considering the results plotted on Fig. 8, is that the total energy gain from one klystron source will be higher if the power is shared between smaller structures, for the same total length. This fact is illustrated on Fig. 9, assuming no power losses in the RF networks, and knowing that the energy gain follows the square root of the input power. Of course smaller structures, when optimized, will have smaller apertures and the interesting result is that the minimum structure length will directly depend on the beam aperture requirement. For instance a minimum aperture of 1.8 cm would lead to a design length of 1.8 m for LIL type cells, according to Fig. 10.

In order to design an optimum linac structure under SLED operation, it is useful to draw design curves having the main design parameters, P_{klystron}, Q, Q_c and t_2. Such a design

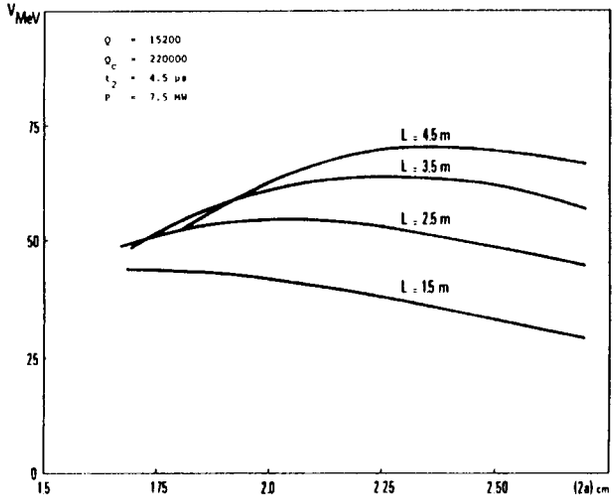


Fig. 7 Energy gain of a constant impedance structure versus the iris diameter and the structure length

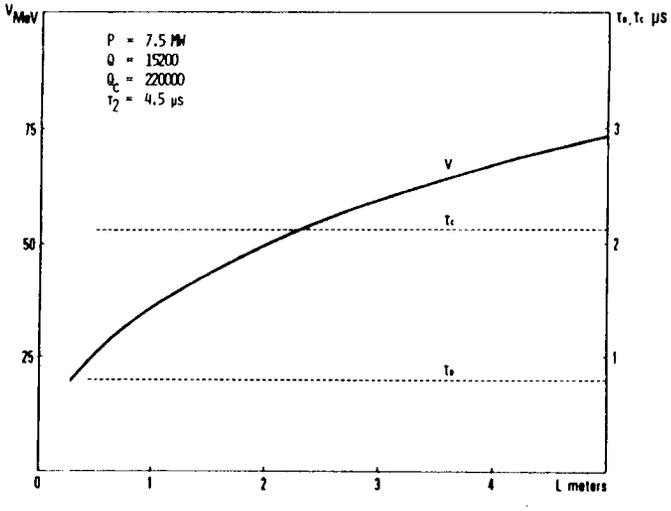


Fig. 8 Maximum energy gain as a function of the structure length

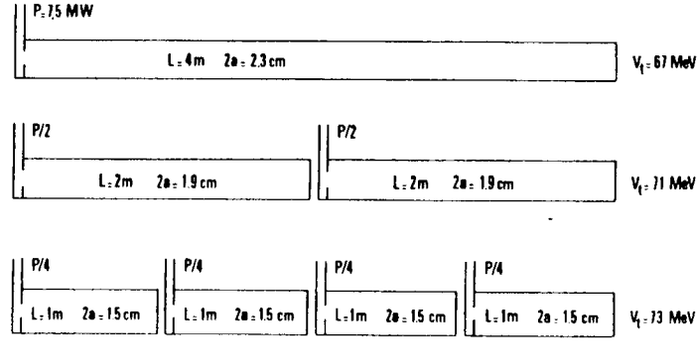


Fig. 9 Total energy gain from a single klystron as a function of the number of structures, for a given total length

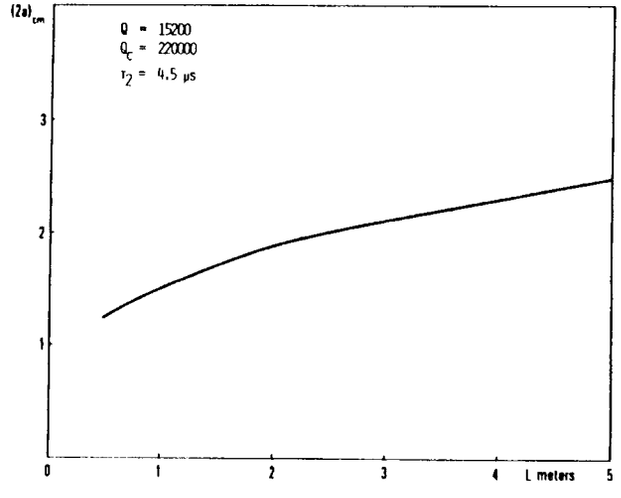


Fig. 10 Optimum aperture of an iris loaded structure versus the structure length

example is shown on Fig. 11. If one introduces a design constraint such as $(2a)_{min}$ = 2.0 cm one gets directly the remaining design parameters which in the present case are:

$$\begin{aligned} & \tau_c = 2.12\mu s \\ L = 2.5 \text{ m} \quad & \tau_a = 0.8\mu s \\ \beta = 8 \quad & t_1 = 4.2\mu s \ . \end{aligned}$$

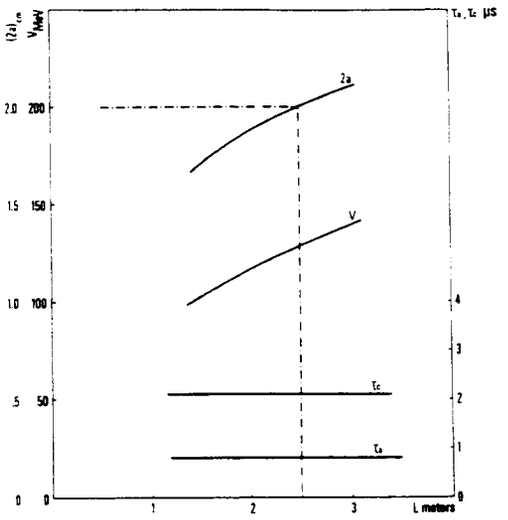


Fig. 11 Design example: P = 40 MW, Q = 15200, Q_c = 180000, t_2 = 5 μs

If a single structure is fed by one klystron the average accelerating gradient becomes 52 MeV/m. A smaller value for $(2a)$ would lead to a higher gradient, for instance $2a$ = 1.65 cm gives 75 MV/m and the corresponding structure length is 1.3 meters. Finally with short constant impedance structures optimized to match the SLED conditions, and commercially available klystrons one can get close to 100 MeV/m in the near future.

4. ULTIMATE ACCELERATING GRADIENTS IN CONVENTIONAL STRUCTURES

A question can be raised now: can we reach in practice the gradient previously mentioned and can we go even further?

The answer to the first part of the question is mainly related to breakdown limits in warm structures and will be treated in this section. If no limitation occurs one way to go further consists of improving both the efficiency of accelerating structures and the peak power of RF sources (their efficiency too) and this will be treated in the next two sections.

4.1 The Kilpatrick criterion

Breakdown phenomena may occur at high field level on the walls of accelerating structures and they are not very well understood at microwave frequencies. The study done by Kilpatrick [9] is one of the few investigations of breakdown phenomena and was in the past very often referred to by accelerator designers. He empirically derived a relationship between frequency and maximum electric field:

$$f = 1.643 E_{max}^2 \exp(-8.5 / E_{max})$$

where f is the RF frequency in MHz and E_{max} is the maximum electric field in MV/m. At f = 3000 MHz this relation predicts E_{max} = 46.8 MV/m.

The corresponding maximum accelerating field now depends on the type of structure. For instance disk-loaded waveguides have a ratio E_{wall}/E_{acc} of the order of 2, hence the maximum expected gradient would be 23 MV/m. This could be one reason why accelerating gradients have been kept below this value for a long time, but certainly another good reason was that the klystron peak power was still low and long accelerating structures were making a better use of this power in terms of maximum beam energy per klystron (no SLED). The overall linac length was not a big worry at that time.

Accelerating structures with higher shunt impedances would lead to lower maximum accelerating gradients since, as a matter of fact, the ratio E_{wall}/E_{acc} increases when the shunt impedance increases.

Recently the need for higher gradients became more and more obvious and new checks of the Kilpatrick criterion become of real concern at several places. As a result it is now believed that the Kilpatrick criterion is pessimistic, at least under pulsed RF conditions.

4.2 The experiment at VARIAN [10]

The experimental set up is shown on Fig. 12, where a single nose cone cell is fed by a magnetron (2.6 MW, 4.4 μs). In this experiment the repetition rate could be varied between 70 and 300 pps while the output peak power could be varied from 0.2 to 2.6 MW.

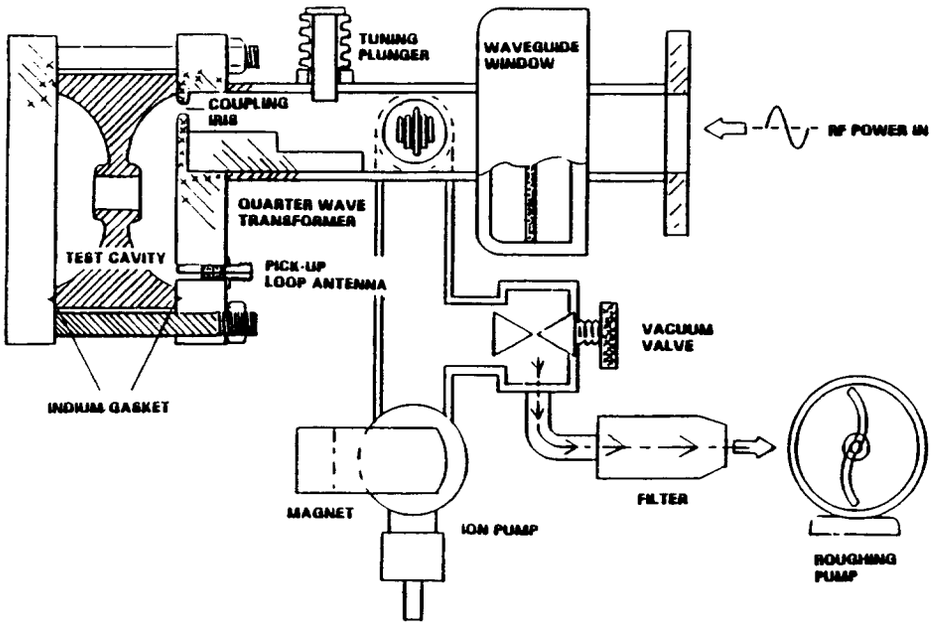


Fig. 12 A cross-sectional view of the cavity test system

The type of cavity which was used has a high shunt impedance, as much as 130 MΩ/m, at 3 GHz and the corresponding E_{wall}/E_{acc} ratio is of the order of 8. The observed breakdown limit corresponded to an accelerating field of 30 MV/m and a maximum field of 240 MV/m on the inner surface of the cell. With different geometries corresponding to different E_{wall}/E_{acc} ratios the maximum field was roughly the same.

It was also observed that above a certain level of wall polishing there was no effect on the breakdown limit. The limit also was found to be independent of the repetition frequency in the range previously mentioned.

From this experiment one can conclude that the maximum surface electric field can be at least as high as five times the limit predicted by Kilpatrick. Extrapolating this result to disk-loaded cavities one can expect at least accelerating gradients close to 120 MV/m.

4.3 Experiments at SLAC [11,12]

The first high gradient test at SLAC was done on the normal SLAC accelerating structures. In order to increase the gradient two klystrons operating in the SLED mode were combined, so that each of the four sections normally fed by one klystron could receive an input peak power as high as 87 MW (Fig. 13).

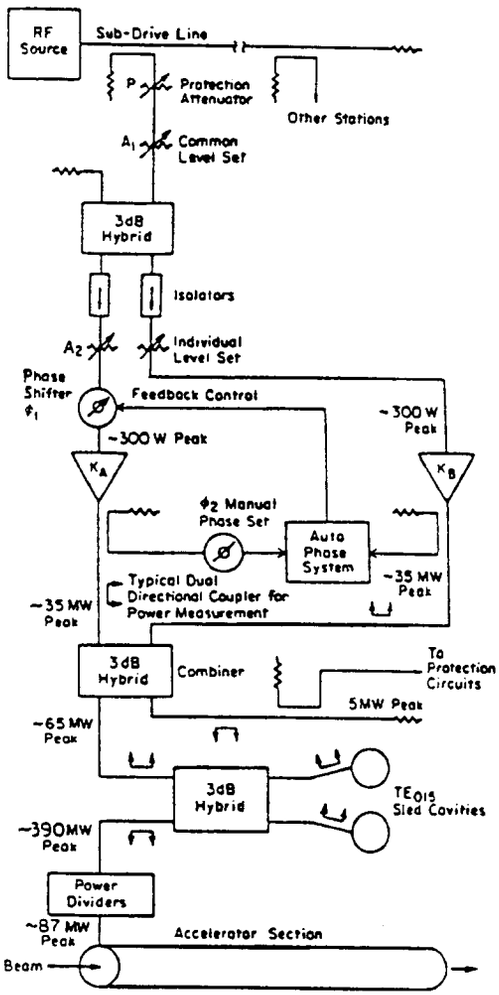


Fig. 13 The first experimental configuration

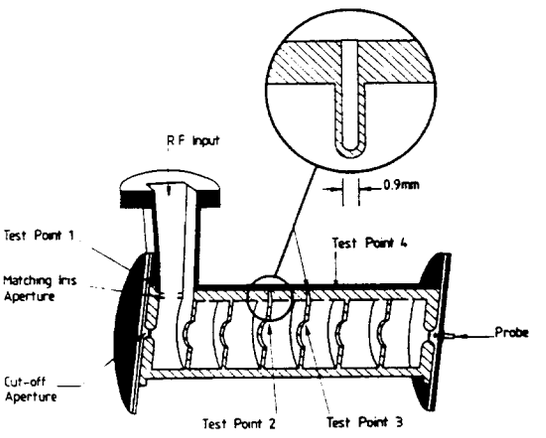


Fig. 14 Resonant structure used for the second experiment. Test points indicate locations of thermocouples.

The corresponding SLED field in the sections was then up to 32 MV/m on the axis and 65 MV/m on the walls. At this level no breakdown occured in the sections. In order to increase the gradient a second experiment was set up in which a short disk-loaded structure was designed to operate in the $2\pi/3$ SW mode (Fig. 14).

The cavity fed by 30 MW RF peak power did not show breakdown problems after a short processing. The maximum equivalent travelling-wave accelerating and surface fields in these conditions were respectively 133 MV/m and 259 MV/m. However it should be noticed that in this experiment considerable X-ray radiation was detected around the section corresponding to a strong field emission.

4.4 Dark current studies

The strong field emission that was detected in the previous experiment at SLAC, led to more systematic experimental works, on typical travelling-wave accelerating structures in order to study the operational limitation under high accelerating gradients.

At LAL, Orsay a test facility was built especially for high gradient experiments at 3 GHz. This facility used a 38 MW, 5 μs klystron that could be combined (or not) with an RF pulse compression system. Two short TW accelerating structures were studied [13], the first one, 0.5 m long, being derived from the LEP Linac structure and the second one, 1.2 m long, of the backward type with magnetic coupling between cells developed by CGR-MeV in collaboration with LAL.

It was observed that, after patient RF conditionning, no dramatic breakdown would occur at accelerating gradients below 100 MV/m. However important field emitted current could be detected at the exit of these structures for accelerating gradients above 50 MV/m, and this current was increasing dramatically, almost exponentially, when increasing the gradient. Systematic experimental studies, together with computer simulations [14] have shown that field emitted electrons are captured and accelerated by the RF field. The main measured properties of this dark current at the exit of the structures, are:

- a wide energy spectrum, showing the contribution from all cells
- an important peak current at the beginning of the RF pulse, which extends over two filling times, and which corresponds to a transient effect during the rise time of the RF pulse. This is followed by a small almost constant current during the RF pulse itself.

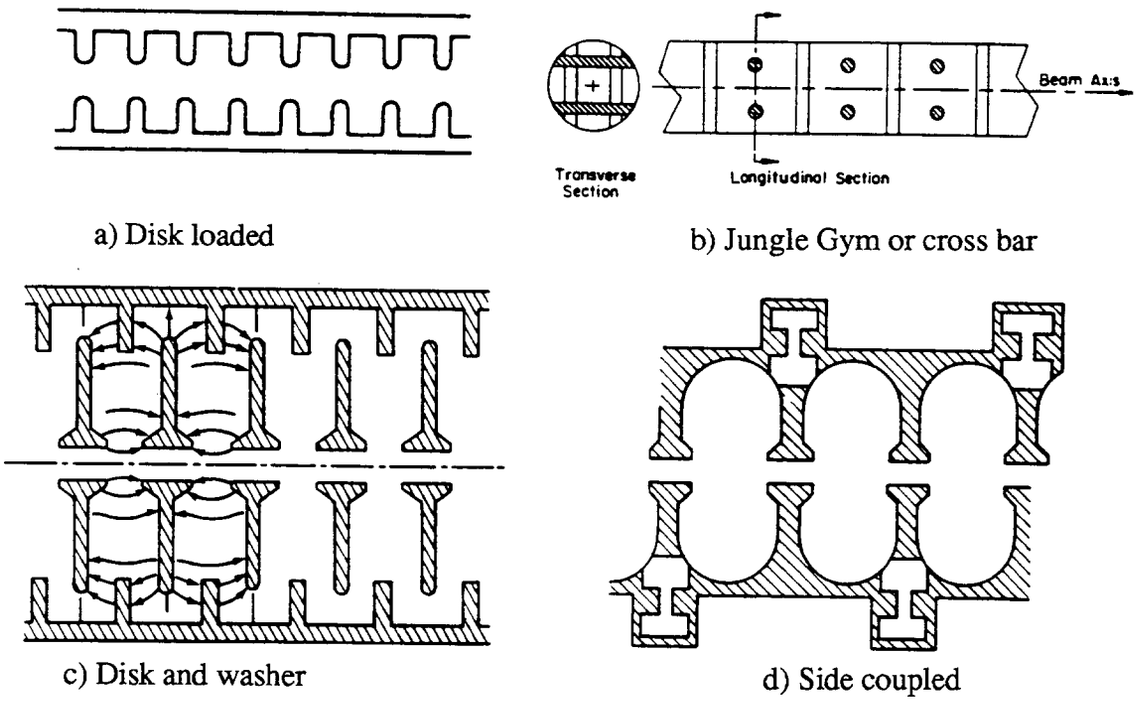
Since the capture mechanism needs higher field at higher operating frequencies, it looks more interesting to operate at high gradients with X-band structures rather than S-band ones. Corresponding dark current studies have been performed on 11.4 GHz disk-loaded structures at SLAC, USA [15] and at KEK, Japan [16] which confirm the previously observed characteristics.

5. A SURVEY OF ACCELERATING STRUCTURES

Previous experiments tell us that accelerating gradients of the order of 100 MV/m can be achieved with conventional disk-loaded structures, but this will need very high peak power and correspondingly high average power to fit the luminosity requirements in a linear collider.

Efforts have already been made to improve the efficiency of accelerating structures and at least four types, either operating in L-band or in S-band, have been developed for the acceleration of electrons (Fig. 15). One can make the following remarks:

- the disk-loaded structure is very well known since it has been used for a long time in linac design. It has a relatively low shunt impedance but a very good ratio E_{max}/E_{acc}
- The jungle-gym structure was first developed for low frequency. Since it has no revolution symmetry it is hard to study this structure with computer codes, and hence it needs more prototype work. However it is expected that this structure will give an improved shunt impedance with a high group velocity.
- The disk-and-washer structure is an open structure, as the previous one, which makes the wall losses smaller and correspondingly leads to a higher shunt impedance. It has also a higher Q but not a higher r/Q.



a) Disk loaded

b) Jungle Gym or cross bar

c) Disk and washer

d) Side coupled

Fig. 15 Different types of accelerating structures

- The side-coupled structure has a very high shunt impedance but a very bad ratio E_{max}/E_{acc}.

The last two structures are quite complicated to build, and up to now they have been mostly considered in the SW mode according to their high shunt impedance.

From the power consumption point of view it is well recognized that for a given type of structure, operation in the SW mode is less efficient (although not obvious when considering small linear accelerators) than operation in the TW mode if correct matching of the source is made in both cases [17]. Hence it is still preferable to consider TW structures for very high energy linacs and in that case the parameters of real importance are r/Q, E_{max}/E_{acc} and v_g/c. For these reasons it is believed that the jungle-gym structure may become a good possibility but still needs more development. In the meantime the old disk-loaded structure will remain a good candidate.

Another advantage of the TW accelerating structure comes from the fact that it can be used in the SLED mode. Moreover, if the group velocity is high the klystron pulse can be made very short and correspondingly the peak power can be increased which is the right direction to follow in the non-SLED case. Table 4 compares the performances of several structures in the TW mode, at different frequencies [18]. Clearly, it is still worthwhile developing short disk loaded structures in the frame of improved power sources.

6. RF POWER SOURCES

Since, for a given accelerating structure, the required RF power must increase as the square of the accelerating gradient, it is an unavoidable necessity to develop very high peak power sources. In principle a very high energy linear collider does not need very high repetition rates provided the luminosity can be achieved by adjusting other parameters (for instance the beam emittance and the final focus); hence one can aim at very high peak power

Table 4
Comparison of structures for a collider

	r (MΩ/m)	Q	v_g/c	L (m)	τ_a (μs)
	2856 MHz				
Disk-loaded (a = 1.16 cm)	56	13,300	0.012	3	0.83
Disk-loaded (a = 1.50 cm)	46	13,000	0.035	6	0.57
Disk and washer	76	32,000	-	-	-
Jungle gym (π/2)	51	9,000	0.20	6	0.10
Jungle gym (π/3)	60	9,000	0.10	6	0.20
	4040 MHz				
Jungle gym (π/2)	61	7,500	0.20	6	0.10
Jungle gym (π/3)	71	7,500	0.10	6	0.20
	5712 MHz				
Jungle gym (π/2)	72	6,500	0.20	6	0.10
Jungle gym (π/3)	85	6,500	0.10	6	0.20

without increasing dramatically the average power. But there are many specific phenomena directly connected to high peak power which make the R&D work tedious and slow. Additionally, high energy accelerators together with high gradients lead to large power consumption which obliges us to concentrate also on improving the efficiency of power sources to limit the operating cost of the machine.

Today the conventional klystron still remains the most interesting RF source for very high peak power, at both S-band and X-band frequencies. Corresponding developments look more like technological improvements rather than breakthroughs. In the following we shall review briefly the present status of pulsed klystrons and summarize the technical problems one has to face.

During the past 10 years new types of RF sources have been considered and prototypes built, such as gyro-klystrons and lasertrons, but they have not been very successfull up to now. The lasertron was a totally new concept, very simple in its principle, but which had to face specific technological short-pulse laser developments. It is hard to know whether the studies on prototypes, which have been stopped everywhere, will be revived in the future. However, as an example of a totally new concept, it will also be presented in this chapter.

6.1 Conventional klystron sources

A klystron is a tube in which a large cathode area produces a high current (many hundred amperes) which is accelerated to a few hundred kilovolts in the anode-cathode space (gun region). For very high power tubes this voltage is produced from a high voltage modulator during a short time (a few microseconds) and at repetition rates up to a few hundred Hertz. A fraction of this beam power ($\geq$ 50%) will be transformed into RF power by modulating spacially the current with the help of a series of bunching cavities, either active or passive, tuned to the desired RF frequency. The last cavity is coupled to a rectangular waveguide to extract the RF power contained in the fundamental frequency. The remaining beam power on the higher harmonics is then dissipated into a collector. The klystron vacuum is isolated from the outside user vacuum by means of a ceramic window located just after the output of the last cavity. Clearly high RF power needs high beam power, $P = IV$, and this has both fundamental and technological impacts on the klystron design, which can be summarized as follows:

1. The maximum current that can be extracted from the cathode is limited by the space-charge effect, and depends on the gun voltage:

$$I = kV^{3/2}$$

where k is called the perveance. Most of the high performance klystrons have high perveance, hence use very high current rather than very high voltage. But the new generation of very high peak power klystrons needs to continuously increase both the current and the voltage. Thermionic cathodes have their own limitations in terms of extracted current density (~ 7 A/cm^2) and are sensitive to vacuum pollution.

2. The modulated high current beam at the RF frequency contains higher harmonics which can interact with the klystron tube and cavities and lead to beam break up (BBU) effects and possibly high field breakdowns on the walls.

3. Output windows have to sustain high peak and average powers. They are subject to breakdown problems. Present solutions consist of relaxing on the electromagnetic field at the window location either by improved designs which minimize the local reflections and the generation of standing wave patterns, or by multiplying the number of output windows. The latter implies recombining the RF power later on.

4. Since a large fraction of the beam power is lost in the collector, increasing this power can lead to cooling problems and to vacuum contamination due to gas desorption. Special designs are required to reduce the incident power density and trap reflected electrons.

SLAC, Stanford, is the most experienced designer and manufacturer of very high peak power klystrons. They recently succeeded in producing a 150 MW, 3 μs, 60 Hz klystron [19], for the DESY S-band collider test facility in Hamburg. The parameters of this klystron, which met the design specifications, are shown in Table 5.

Table 5
Design parameters for the 150 MW klystron

Beam voltage	535 kV
Beam current	700 A
RF pulse width @ rep rate	3 μs @ 60 Hz
Cathode loading	2:1 (6 A/cm^2 max)
Cathode convergence	40:1 (5.25" dia.)
RF output power	150 MW
Cavity gradients	< 360 kV/cm
Saturated gain	~ 55 dB
Efficiency	≥ 40%
Operating frequency	2998 MHz
Solenoidal focusing field	2100 gauss (3*Br)

SLAC is also developing, for its own collider test facility, a high power X-band klystron [20]. Aiming at 100 MW, 1.5 μs at 11.4 GHz, they have already reached 80 MW, 0.8 μs. A special effort is being made to test different types of output cavities (standing-wave multigaps, travelling-wave structures).

At KEK, Japan, a considerable effort is also being made on X-band klystrons and they have already achieved performances which are not so far from those at SLAC.

Finally it should be pointed out that in Protvino, Russia, they are also developing an X-band klystron, with unusually low perveance (1 MV, 300 A), and very high gain (80 dB).

Aiming at 150 MW, 0.7 μs, they already reached 60 MW at the design pulse length and seem to be limited by dipole modes that develop in the long tube with many gain cavities.

6.2 A new RF power source concept: the Lasertron

High beam power either requires a higher beam current, which would create strong space-charge effects, or a higher accelerating voltage, which would reduce the bunching efficiency, hence the amount of power contained in the fundamental and the extraction efficiency.

To overcome these difficulties, although not proven to be fundamental limitations, a new microwave RF power tube has been proposed [21,22] in which a photocathode, illuminated by a modulated laser, emits short, dense current pulses which, after being accelerated, traverse an output cavity where the RF beam modulation is extracted (Fig. 16). Here, a high accelerating voltage is necessary to compensate for the space-charge forces which otherwise would distort the emitted short bunches and reduce the extraction efficiency. Since in principle the laser can provide a train of such bunches with a given repetition rate, the accelerating voltage can be d.c.

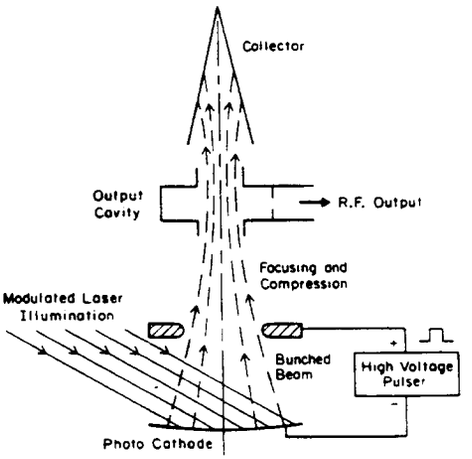


Fig. 16 Schematic of a photocathode microwave power source

Considering the fast pulsed photo-emission it is believed that the maximum charge which can be extracted per pulse from the photocathode is equal to the superficial charge [23]:

$$Q = \varepsilon_0 E_\perp S = \frac{\varepsilon_0 S}{d} V = CV$$

where $E_\perp$ is the accelerating field at the photocahtode, S the useful area of the photocathode, d the distance between the cathode and the anode, C the gun capacitance and V the accelerating voltage.

This maximum charge is half the space-charge limit, showing that the limitation of such a tube is very different from that of a klystron. As a matter of fact, if f_{RF} is the repetition frequency of the laser pulses, the average current per laser burst is:

$$I_0 = f_{RF} CV$$

and the beam power:

$$P_b = f_{RF} CV^2$$

while in a klystron:

$$P_b = k\ V^{5/2}\ .$$

The main parameters of the SLAC prototype [24] are given in Table 6. The power level is comparable to the peak power of klystrons, and this is a first step to check the lasertron principle before envisaging much higher peak power. Prototypes were also under consideration in Japan and in France but all work has now been abandoned or postponed.

Table 6
Parameters of the SLAC prototype lasertron

Peak RF output power	35 MW
Beam power	50 MW
Efficiency	70%
Voltage	400 kV dc
Peak pulse current	735 A
Cathode diameter	3 cm
Average pulse current (~ Peak/6)	126 A
Optical pulse length	60 ps FWHM
Optical pulse separation for a 2856 MHz rate	350 ps
Microwave pulse length or optical pulse train (comb) length	1 μs
Average power (power supply limited)	< 4 kW
Peak electric field in gun region	< 15 MV/m
Electric field on planar cathode	10 MV/m
Maximum magnetic focusing field	0.2 T

Among the difficulties encountered in designing a lasertron it is worthwhile mentioning the high current photocathode. Remembering the poor efficiency of lasers, the photocathode must have a very good quantum efficiency. Unfortunately it so happens that efficient cathodes, such as AsGa for instance, have a poor lifetime. Conversely, metallic cathodes are robust but have a poor quantum efficiency. Modulated lasers at S-band or C-band frequencies, with long pulse trains and high repetition rate have to be developed also, with optical frequencies either in the visible (green) or in the VUV. Figure 17 taken from reference [24] gives a good idea of the lasertron geometry as well as the technologies involved.

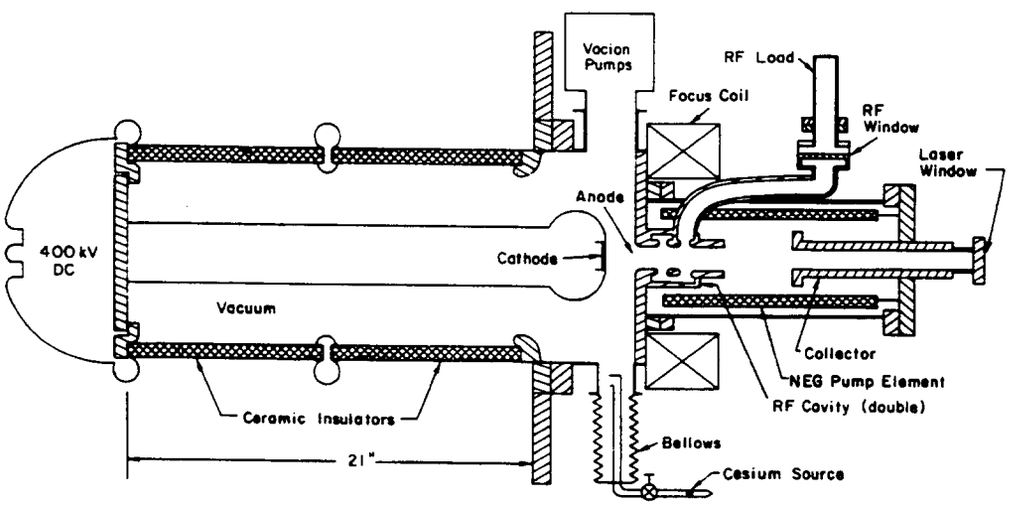


Fig. 17 Geometry of the SLAC prototype lasertron

REFERENCES

[1] SLAC Linear Collider (SLC): Conceptual design report, SLAC-Report 229 (1980).

[2] Z.D. Farkas et al., SLED: A method of doubling SLAC's energy, Proc. 9th Int. Conf. on High Energy Accelerators, Stanford, 1974, (SLAC, Stanford, 1974).

[3] G. Caryotakis, High power microwave tubes in the laboratory and on-line, SLAC-PUB-6361, Sept. 93.

[4] J. Le Duff, LAL/RT/85-03, Orsay (1985).

[5] F. Bulos et al., Physics with linear colliders in the TeV CM energy region, SLAC-PUB3002. Also, contribution to the Proceedings of the Summer Study on Elementary Particle Physics and Future Facilities, 1982, Snowmass, C0.

[6] The LEP Injector Study Group, LEP design report, Vol. 1: the LEP injector chain, CERN-LEP/TH/83-29; LAL/RT/83-09 (1983).

[7] J. Le Duff, Optimization of TW accelerating structures for SLED type modes of operation, LAL/RT/84-01, LAL-Orsay (1984).

[8] R. Belbeoch et al., Rapport d'etudes sur le projet des linacs injecteurs de LEP (LIL), LAL/RT/82-01, Orsay (1982).

[9] W.D. Kilpatrick, Criterion for vacuum sparking designed to include both RF and DC, UCRL-2321, (1953).

[10] E. Tanabe, Voltage breakdown in S band linear accelerator cavities, Proc. 1983 Particle Accelerator Conference, Santa Fe, 1983, IEEE Trans. Nucl. Sci., NS-30, (1983).

[11] H.A. Hogg et al., Experiments with very high power RF pulses at SLAC, IEEE Trans. Nucl. Sci., NS-30, (1983).

[12] J.W. Wang and G.A. Loew, Measurements of ultimate accelerating gradients in the SLAC disk-loaded structure, Proc. 1985 Particle Accelerator Conference, Vancouver, 1985, IEEE Trans. Nucl. Sci., NS-32, (1985).

[13] G. Bienvenu and P. Brunet, Operational limits of high accelerator gradient, Proc. 2nd EPAC, Nice, June 12-16, 1990, p. 934-936, Ed. Frontières.

[14] G. Bienvenu, P. Fernandes and R. Paridi, An investigation of the field emitted electrons in travelling wave accelerating sections, NIM, A320 (1992)-I-8, North Holland.

[15] J.W. Wang et al., High gradient studies on 11.4 GHz copper accelerator structures, Proc. 1992 Linac Conference, Ottawa, Ontario, Canada (August 24-28, 1992), p. 716-718.

[16] T. Higo et al., High gradient experiment for X-band travelling wave structure, Proc. 1992 Linac Conference, Ottawa, Ontario, Canada (August 24-28, 1992), p. 145-147.

[17] P.B. Wilson, IEEE Trans. Nucl. Sci., NS-26, 3255 (1979).

[18] P.B. Wilson, IEEE Trans. Nucl. Sci., NS-28, 2742 (1981).

[19] D. Sprehn, R.M. Phillips and G. Caryotakis, The design and performance of 150 MW S-band klystrons, SLAC-PUB-6677, Sept. 1994.

[20] R.H. Miller, New developments in RF power sources, SLAC-PUB-6557, June 1994. Also as invited paper in the proceedings of the 4th EPAC June 1994, London, UK, World Scientific.

[21] M.T. Wilson, P.J. Tallerico, US patent n°4,313,072-1/26/1982.

[22] M. Yoshioka et al., Laser-Triggered RF sources for Linacs in TeV region, Proc. 1984 Linac Conference, Seeheim, R.F.A., 1984.

[23] H. Nishimura, Particle simulation code for non relativistic electron bunch in lasertron. Proceedings of the 1984 Linac Conference, Seeheim (R.F.A.) 1984.

[24] E.L. Garwin et al., An experimental program to build up a multimegawatt Lasertron for super linear colliders, Proceedings IEEE, NS-32, (1985).

INSTABILITIES IN LINACS

A. Mosnier
CEA Saclay, 91191 Gif/Yvette, France

Abstract
One of the major problems with a linac accelerating intense bunches is the instability driven by the so-called transverse wakefields. This instability, generated by off-axis beam trajectories, can develop within a single-bunch or along a finite or infinite train of bunches and always leads to a dilution of the emittance. We approach this problem by first reviewing the basic mechanisms of single and multibunch instabilities. Analytical formulae allowing an estimation of the emittance growth, and results of computer simulations are then compared in both cases, with emphasis on the design parameters of a linear collider. The cures for the instabilities which are contemplated with the aim to suppress or reduce the effects are examined as well throughout this study.

1. INTRODUCTION

The collective effects which appear in a linac, where the bunches are permanently refreshed, are improperly called instabilities, because there is no exponential growth in time of the excitation but rather a magnification effect which will deform the bunch or the train of bunches. This is unlikely in a circular accelerator where the same bunches undergo essentially an infinite number of turns. A real instability, however, can take place in a single accelerating section [1] due to the energy exchange between the beam and a backward-wave component of the fields, or in a recirculating linac where the same beam circulates several times in a linear accelerator [2]. In both cases there is closure of a feedback loop and the exponential growth of the fields occurs above a threshold current. The former effect is known as regenerative beam breakup and the latter as multipass beam breakup. We will focus hereafter on the more elementary single-pass "instability", which is a crucial problem for the future linear colliders.

In a linac we want to accelerate a beam to its final energy while preserving its emittance; that is, keeping the phase-space volume occupied by the beam as small as possible. If we consider a linear collider, for example, the luminosity of two identical unperturbed beams with Gaussian distributions colliding head-on is given by

$$\mathcal{L} = \frac{N^2 f_c}{4\pi \sigma_x \sigma_y}$$

where N is the number of particles in each bunch,
f_c is the collision frequency,
$\sigma_{x,y} = \sqrt{\beta_{x,y}\,\varepsilon_{x,y}}$ gives the transverse beam sizes,
with ε the emittance and β the beta function.

We see that high bunch populations with low emittances are desired for obtaining a maximum luminosity. On the other hand, to increase the collision rate while using the available rf energy more efficiently, a train of a few – or many – bunches spaced apart by a few wavelengths of the fundamental mode instead of a single-bunch on each rf fill, is considered. As a high charge bunch travels down a large number of accelerating structures, strong wakefields are excited which will act back on the bunch itself (the so-called short-

range wakefield effects). With a train of intense bunches, the wakefieds excited by each bunch will act on the subsequent bunches (the so-called long-range wakefield effects).

The longitudinal wakefields, where the forces are parallel to the beam axis, affect the energy of the particles and induce an energy spread along the bunch.

With off-axis beam trajectories the transverse wakefields, where the forces are perpendicular to the beam axis, are excited and tend to deflect the beam further away from the axis.

1.1 The basic mechanisms

We assume a long linac with focusing forces provided by quadrupole magnets. Consider for example a coherent betatron oscillation of the beam induced by an offset of the whole bunch at the injection point. As the beam moves down, the head undergoes an unperturbed transverse motion, which is a simple harmonic motion of betatron frequency $\omega\beta$. The tail, on the other hand, experiences deflection due to the excitation of preceding particles. Since the wakefield force is proportional to the offset of the leading particles and has therefore a time dependence of frequency $\omega\beta$, the motion of the tail is now the one of a harmonic oscillator driven on resonance. The amplitude of the oscillation will then increase with distance. This causes the tail of the bunch to distort, in the manner represented on Fig. 1, and become "banana shaped", in a process called single-bunch beam break-up. The same phenomenon can be observed when the structures are misaligned with respect to the center beamline, while the beam is injected on-axis, as illustrated on Fig. 2.

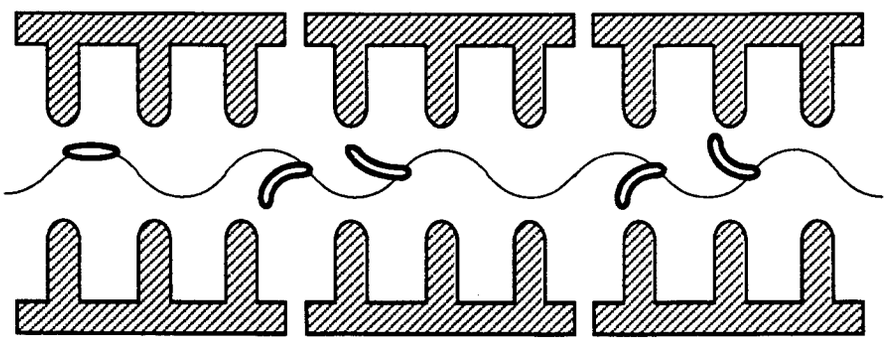


Fig. 1 Single-bunch beam break-up originating from an injection error

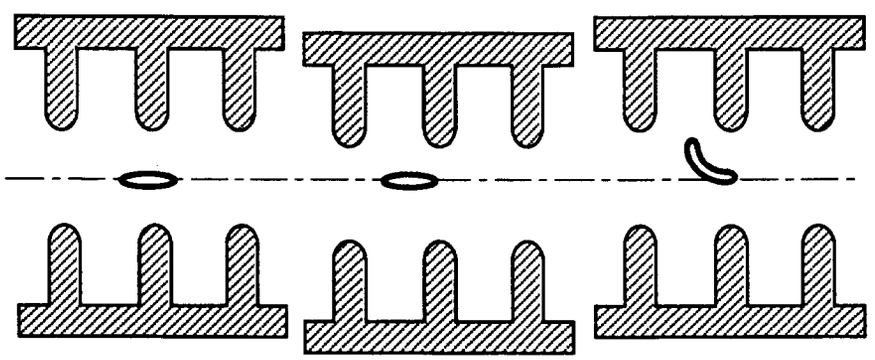


Fig. 2 Single-bunch beam break-up originating from structure misalignment

Similarly, any beam offset or structure misalignment will excite long-range wakefields, which will then cause subsequent bunches to be deflected. On the left part of the Fig. 3 are drawn the electric and magnetic fields of a ringing transverse mode of the accelerating

section. In this example, the TM11-like mode can be recognized, which generally gives the dominant effect in these types of disk-loaded structures. This deflected beam will excite a similar deflecting mode in the downstream structures, which will, in turn, deflect even more the following bunches of the train, and so on. Due to the long-range wakefields, there is now a coupling in the motion of the bunches, which are more and more deflected as they proceed down the linac in a process called multibunch or cumulative beam break-up. Even when particles are not lost by scraping the accelerator walls, the beam emittance can be greatly increased and lead to a significant luminosity reduction.

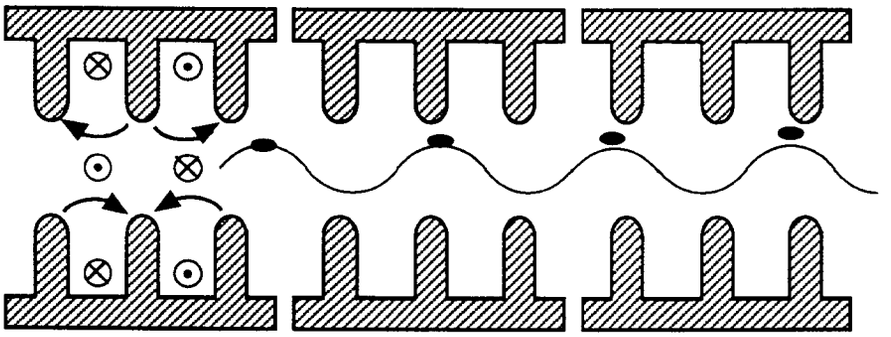


Fig. 3 Cumulative beam break-up driven by a dominant deflecting mode

Off-axis beam trajectories arise due to a variety of errors, departing from an ideal machine, but the causes of main concern are

- a beam offset at injection

- the misalignment of focusing magnets

- the misalignment of accelerating sections

It is worthwhile to note that the final emittance, in addition to the collective effects, will be also affected by single-particle effects, originating from these same errors. The dispersive effects, in particular, due to the finite beam energy spread can dramatically dilute the beam emittance but their study is beyond the scope of this lecture.

1.2 Typical wakes in a linac

Since a linac is normally composed mostly of numerous accelerating structures, these are the main sources of wakefields. Other additional sources such as position monitors, bellows or transitions can contribute to a lesser extent. The accelerating sections are generally cylindrically symmetric structures and the electromagnetic fields can be expanded in terms of multipoles with azimuthal variation proportional to $\cos(m\theta)$. Since however the beam remains near the axis and the beam sizes are tiny compared to the hole radius, we retain only the monopole and dipole wakes which dominate, respectively, the longitudinal and transverse wakefields.

Normal conducting travelling wave structures for relativistic particles are generally long disk-loaded waveguides with period $p = \lambda_{rf}/3$ while standing wave structures with period $p = \lambda_{rf}/2$ are used for superconducting rf cavities. The frequency of the accelerating mode varies on a large scale, from the order of 1 GHz to a few tens of GHz. The scaling of the point-like wake potentials with the rf frequency is the following

$$W_{z(z)} \propto \omega^2 \quad \text{and} \quad W_{\perp(z)} \propto \omega^3$$

The higher the operating frequency, the stronger the wakefield effect on the beam will be! The definition of the point-like – or Green's function – wake potentials can be found in Ref. [3] or [4] and are recalled in Appendix 1.

Fig. 4 shows the short range wakes which drive the single-bunch effects – for a typical disk-loaded structure with a rf frequency of 10 GHz. We note that the longitudinal wake is finite at the origin, whereas the dipole wake is approximately linear on a very short time scale.

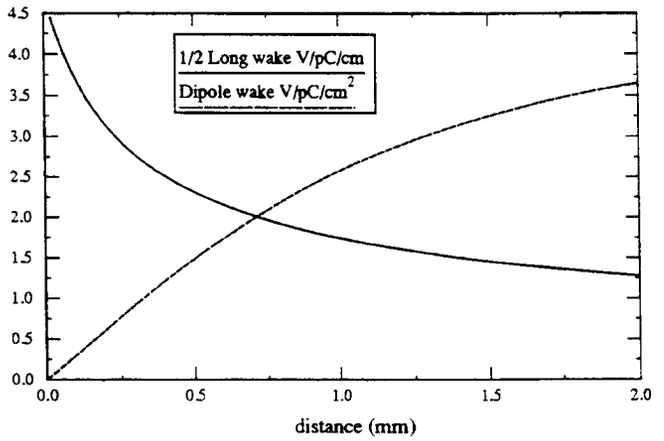


Fig. 4 Short-range wakes for a typical 10 GHz disk-loaded structure

The long range wakes, which drive the multibunch effects, are characterized by the ringing of a few dominant modes and can be analysed in terms of the normal modes of a cavity. In this case, the dipole wake of a particular mode is given by

$$W_{\perp(\tau)} = 2K\, e^{-\tau / T_F} \sin \omega_r\, \tau$$

where ω_r is the frequency of the mode,
$T_F = 2Q/\omega_r$ is the decay time and Q is the quality factor,
K is the kick factor of the mode.

The time τ has been used instead of the distance $s = c\tau$ between the exciting and test particles. The kick factor K can be written in terms of the shunt impedance per unit length divided the quality factor $2K = \omega_r R_\perp / Q$.

The wake and the impedance of a high Q resonant mode are simply related by a Fourier transform (see for instance Ref.[3])

$$Z_{\perp(\omega)} = \frac{\omega_r}{\omega} \frac{R_\perp}{1 + jQ(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega})} = j \int_{-\infty}^{+\infty} d\omega\, W_{\perp(\tau)}\, e^{j\omega\tau}$$

1.3 Longitudinal wakefield effects

The energy spread of the beam must be kept as small as possible, in order to minimize especially chromatic emittance dilution. As a short intense bunch travels through the accelerator, it will gain energy from the externally driven accelerating mode and will lose energy due to the longitudinal wakefields.

The voltage gain of a particle in the bunch depends on its longitudinal position z with respect to the bunch center

$$\Delta V_{(z)} = E_a \cos(\phi_{rf} + \frac{2\pi}{\lambda_{rf}} z)$$

where E_a is the accelerating gradient,
and ϕ_{rf} is the phase of the bunch center with respect to the rf wave crest

Similarly, the voltage loss, also called the bunch potential, of a particle in the bunch depends on its longitudinal position. The bunch potential $\mathcal{W}_z$ is the convolution of the point charge wake W_z and the charge distribution ρ of the bunch

$$\Delta V_{(z)} = Q_b \mathcal{W}_{z\,(z)} = Q_b \int_0^\infty dz' \, \rho_{(z-z')} W_{z\,(z')}$$

The bunch potential can be directly computed by using a time domain cavity code such as TBCI, Ref. [5], or ABCI, Ref. [6], which integrates the Maxwell's equations in the time domain. The net voltage gain will be the superposition of the sinusoidal rf accelerating potential and the bunch potential. Since both the induced wake and the slope of the rf wave look quite linear, at least over a certain range, the resulting energy spread can be reduced by running the bunch off the crest of the accelerating wave, such that the rf wave slope tends to cancel the average slope of the beam induced potential. Figure 5 shows the energy profile within a bunch of charge 3.2 nC tavelling through a 10 GHz disk-loaded structure, when the bunch sits on the crest of the rf wave. The accelerating field is 100 MV/m and the bunch charge distribution is assumed Gaussian with a standard deviation of 0.25 mm. The energy variation due to the rf wave and due to the longitudinal wake are given by the dotted lines. The solid line, which is the sum of both effects, shows the resulting energy variation within the bunch. The final energy spread is about 0.5 %. Figure 6 shows the same plot after minimization of the resulting energy spread (the rf phase is such that the tail is more accelerated than the head) which has been reduced by a factor 3 in this example.

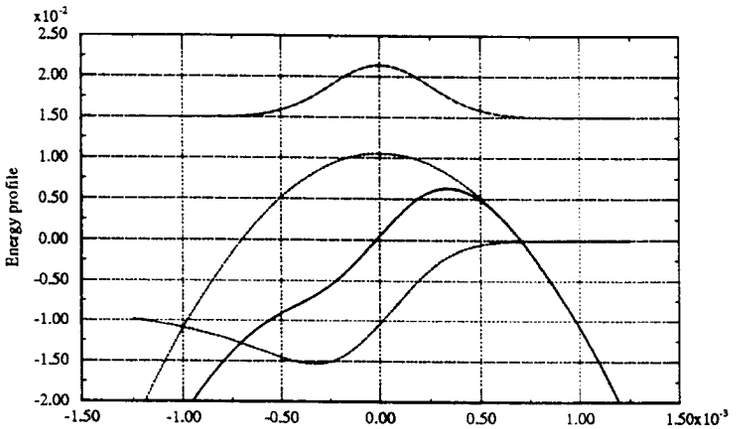


Fig. 5 Energy profile within the bunch sitting on the crest of the rf wave

With a train of bunches, there will be in addition a bunch-to-bunch energy spread because the energy of the particles in a bunch is affected by the wakefields of previous bunches. The fundamental mode and the higher-order longitudinal modes contribute to this bunch to bunch energy variation. Various compensation schemes have been investigated to reduce the large energy variation between bunches due to the fundamental mode beam loading (see for instance Ref. [7] or [8]).

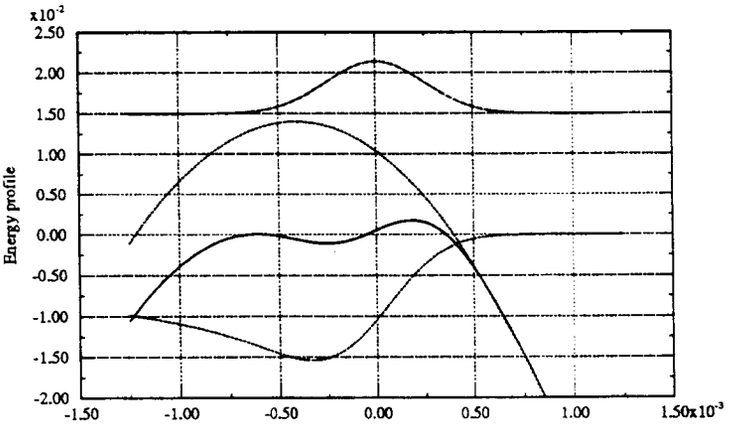


Fig. 6 Energy profile within the bunch after optimization of the rf phase

1.4 The general transverse equation of motion

If we consider only the forces coming from any linear focusing lattice, the particles exhibit pure betatron oscillations around a central trajectory, as it must be in any ideal transport line. Taking acceleration into account, the equation of motion for the transverse offset at beamline location s is

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}x_{(s)}\right]+K_{(s)}x_{(s)}=0$$

where $x_{(s)}$ is the transverse displacement of a particle at position s along the accelerator,

$\gamma_{(s)}$ is the usual Lorentz factor,

$K_{(s)}=\frac{1}{B\rho}\frac{\partial B_y}{\partial x}$ is the focusing function.

Once given the initial conditions, x_0 and x'_0, the linear betatron motion can be described by the usual transport matrix

$$\begin{pmatrix}x_{(s)}\\x'_{(s)}\end{pmatrix}=\begin{bmatrix}R_{11(s_0,s)} & R_{12(s_0,s)}\\R_{21(s_0,s)} & R_{22(s_0,s)}\end{bmatrix}\begin{pmatrix}x_0\\x'_0\end{pmatrix}$$

where the coefficients can be written in terms of the lattice parameters, for example

$$R_{11(s_0,s)}=\sqrt{\frac{\beta_{(s)}}{\beta_0}}(\cos\Delta\psi+\alpha_0\sin\Delta\psi)\times\sqrt{\frac{\gamma_0}{\gamma_{(s)}}}$$

$$R_{12(s_0,s)}=\sqrt{\beta_{(s)}\beta_0}\sin\Delta\psi\times\sqrt{\frac{\gamma_0}{\gamma_{(s)}}}$$

where β_0 and $\beta_{(s)}$ are the beta funtions at locations s_0 and s,

$\Delta\psi=\psi_{(s)}-\psi_{(s_0)}$ is the phase advance between s_0 and s ,

γ_0 and $\gamma_{(s)}$ are the Lorentz factors at locations s_0 and s .

The accelerated beam solution, under the assumption that the beam energy does not change too much in a betatron wavelength, has simply an additional term, called the adiabatic damping factor. It can be shown from the equation of motion that this approximation of

adiabatic acceleration is justified if the following condition, which is true for the cases of interest to us, is satisfied.

$$\gamma' / \gamma << 2 / \beta \quad \text{or} \quad G / E << 4\pi / \lambda_\beta$$

where G is the accelerating gradient and E is the actual energy,

λ_β is the betatron wavelength.

We consider now the transverse motion of particles perturbed by the wakefield effects. There is one equation and only one, governing the transverse motion, in a linear focusing lattice, of a bunch interacting with itself through the short-range wakefield or interacting with the other bunches of a pulse through the long-range wakefield.

The displacement of one charge $x(\tau,s)$ is a function of the location s along the accelerator and of the time interval $\tau = z/c$, z being the longitudinal position behind the head of the bunch for the single-bunch effects study or the first bunch of the train for the multibunch effect study. The transverse force due to dipole wakefields received by a test particle located at time position τ, depends on the displacement of all the preceding charges, located at time position $\tau' < \tau$ (Fig. 7).

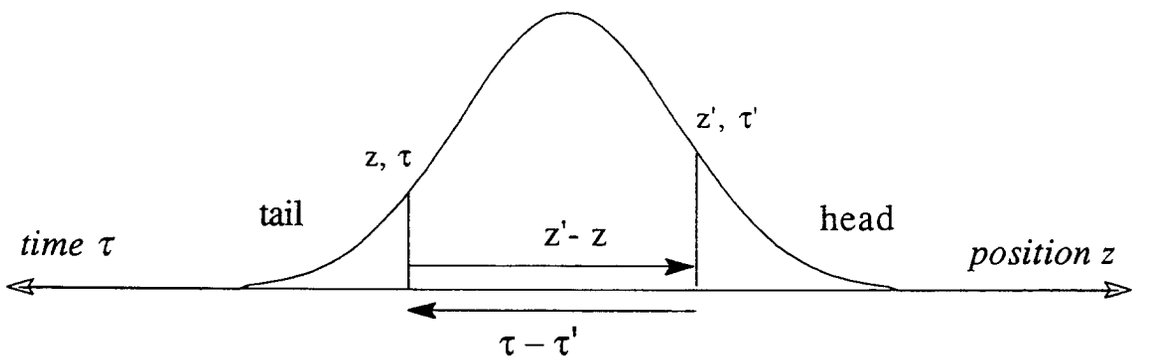


Fig. 7 Longitudinal position of the driving and test particles in a bunch

From the definition of the dipole wake, the elementary wakefield force experienced by the test particle following a driving elementary charge with the time interval τ - τ' is

$$dF_{\perp(\tau,s)} = e\, W_{\perp(\tau-\tau')}\, \rho_{(\tau')} d\tau'\, x_{(\tau',s)}$$

where $W_{\perp(\tau)}$ is the point-like dipole wake per unit length,

$\rho_{(\tau)}$ is the charge distribution of the bunch or of the train of bunches,

with the normalization $\int \rho_{(\tau)} d\tau = Q_b$ = total charge .

The total force experienced by the test particle located at time position τ is then

$$F_{\perp(\tau,s)} = e \int_{-\infty}^{\tau} d\tau'\, W_{\perp(\tau-\tau')} \rho_{(\tau')} x_{(\tau',s)}$$

The general equation of transverse motion for a test particle which feels the external forces of the focusing lattice and of the wakefields is then

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}x_{(\tau,s)}\right]+K_{(s)}x_{(\tau,s)}=\frac{1}{E_{(s)}}F_{\perp(\tau,s)} \qquad (1)$$

In the case of multiple bunches, generally equally spaced apart by a few wavelengths, the bunches are treated as rigid macroparticles. Although the dynamics of a single-bunch and the dynamics of a train of bunches are governed by the same equation, the differences originate mainly from two sources:

i) the shape of the wake: While the short-range wakefield can be approximated by a linearly increasing function, the long-range wakefield is described by means of resonant modes

ii) the charge density: While the charge density of a bunch is generally Gaussian, a pulse consists of bunches of equal charge with uniform spacing between adjacent bunches. This is the reason why the technique of discrete Laplace transforms is more suited for the multiple-bunches case. Taking, however artificially, a uniform distribution for the single-bunch permits us to use directly the powerful Laplace transform technique. Moreover, while the charge density of a single-bunch is continuous and has the feature of a d.c. current, the charge density of a train of bunches is obviously discrete and consists of delta functions. With a train of delta bunches, resonances can then appear when the mode frequency is a multiple of the repetition frequency.

Our purpose is now to study the single-bunch beam break up. But instead of dealing directly with the whole charge distribution, we shall use the well known and very useful model based on just two particles. This model gives a better insight of the single-bunch BBU phenomenon and to find the remedy. Though very simple, it provides results very close to those obtained when using the real charge distribution, as long as the final dilution is not too large.

2. THE SINGLE-BUNCH TRANSVERSE INSTABILITY

The bunch is modelled by two macro-particles, each of charge $Q_b/2$ separated by a longitudinal distance $\Delta z = 2\,\sigma_z$ (Fig. 8).

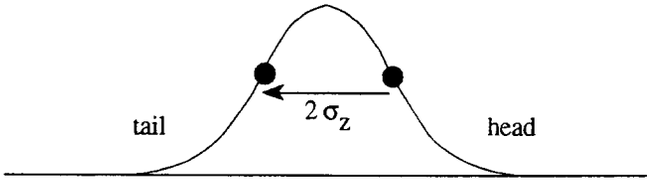


Fig. 8 The two-particle model

In addition, the focusing system, is smoothed in such a way that the unperturbed motion is a pure harmonic oscillator with a betatron wave number given by

$$k = 1/\beta = 2\pi/\lambda_\beta$$

where β is an average beta function which, however, may vary slowly with s

λ_β is the betatron wavelength.

2.1 Effect of a coherent betatron oscillation

We assume that the bunch undergoes a coherent betatron oscillation, induced, for example, by an offset of the beam at injection. Let us consider first the case of a non-accelerated beam. Then, the equations of motion for both particles become

$$\begin{aligned} x_1'' + k_1^2 x_1 &= 0 \\ x_2'' + k_2^2 x_2 &= \frac{1}{E/e} Q_b/2\, W_{\perp(2)}\, x_1 \end{aligned} \tag{2}$$

where k_1 and k_2 are the betatron wave numbers for the head and the tail particle, E is the energy, and primes denote derivatives with respect to location s. $W_{\perp(2)}$ means the transverse wakefield at the distance $\Delta z = 2\,\sigma_z$.

The head particle feels no transverse wakefield and thus undergoes free betatron oscillations. In the smooth focusing approximation and with an initial bunch offset x_0 at $s = 0$, we have

$$x_1 = x_0 \cos k_1 s$$

The tail particle, though, feels the wake force due to the off-axis motion of the leading particle and we can distinguish two cases: the head and the tail particles have the same or have different betatron frequencies.

2.1.1 Same betatron frequency

We assume that both particles have the same betatron wave number $k_1 = k_2 = k$. The equation of motion of particle 2 is that of a harmonic oscillator driven on resonance. The trailing particle oscillates relatively to the driving particle with an amplitude increasing linearly with distance s

$$\Delta x = x_2 - x_1 = \frac{Q_b/2\, W_{\perp(2)}}{2k\, E/e} x_0\, s \sin k s$$

We see that the oscillation of the trajectory difference is in quadrature with the betatron oscillation. The acceleration can be included after noting that the relative displacement of the tail with respect to the head for a length ds of structure is

$$\frac{d}{ds}\left|\frac{\Delta x}{x_1}\right| = \frac{Q_b/2\, W_{\perp(2)}}{2k\, E_{(s)}/e}$$

The energy E is now a function of the distance s along the accelerator. The integration of the previous differential equation up to the distance s, yields

$$\left|\frac{\Delta x}{x_1}\right| = \frac{Q_b/2\, W_{\perp(2)}}{2\,\mathcal{E}_0/e} \int_0^s ds' \frac{\beta_{(s')}}{\gamma_{(s')}} \tag{3}$$

where $\mathcal{E}_0$ is the rest energy of the particle. If we assume a uniform acceleration $E_{(s)} = E_0 + G\,s$, and a constant focusing function, the integral of the expression (3) is readily performed, yielding

$$\int_0^s ds' \frac{\beta_{(s')}}{\gamma_{(s')}} = \beta \frac{\mathcal{E}_0}{G} Ln \frac{\gamma_{(s)}}{\gamma_0}$$

Up to now, we have considered a constant focusing strength from the beginning to the end of the machine, but we can allow the average β function to vary slowly with s. If we consider, for example, a beta function varying as the square root of the energy, which will be justified hereafter,

$$\beta_{(s)} = \beta_0 \left(\frac{\gamma_{(s)}}{\gamma_0} \right)^{1/2}$$

the integral of the expression (3) is approximately proportional to the square root instead of a logarithm of the energy.

$$\int_0^s ds' \frac{\beta_{(s')}}{\gamma_{(s')}} = 2\beta_0 \frac{\mathcal{E}_0}{G} \left(\sqrt{\frac{\gamma_{(s)}}{\gamma_0}} - 1 \right)$$

Having taken into account the acceleration, we can take a step further : we replace the smooth focusing approximation by the more realistic alternating-gradient (AG) focusing, like a FODO lattice, by reintroducing the betatron function. The accelerator is then modeled by a succession of focusing magnets and of accelerating sections, and where the momentum kicks due to the wakefields are concentrated in the middle of the structures (Fig. 9).

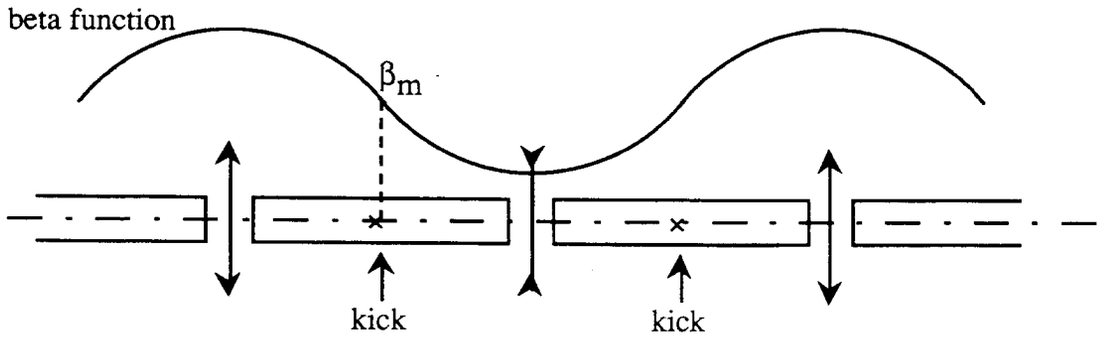


Fig. 9 A FODO lattice

To do this, we have simply to replace the average beta function $\beta = 1/k$ of the smooth-focusing approximation by the β value calculated at the middle of the structure, β_m. The expression of β_m is given in Appendix 2. The amplitude of the relative displacement of the tail with respect to the head thus becomes for the two focusing lattices for the constant β lattice

$$\left| \frac{\Delta x}{x} \right| = \frac{Q_{b/2}\, W_{\perp(2)}}{2\, G/e} \beta_m \, Ln \frac{\gamma_f}{\gamma_0}$$

and for the energy dependent β lattice

$$\left| \frac{\Delta x}{x} \right| = \frac{Q_{b/2}\, W_{\perp(2)}}{2\, G/e} 2\beta_{m0} \left(\sqrt{\frac{\gamma_f}{\gamma_0}} - 1 \right)$$

Figure 10 shows the result of a simulation of a bunch traversal through a linac composed of discrete focusing elements. We check the growing oscillation of the trailing particle along the linac for both focusing types when the two particles are launched with a unit offset. The logarithmic (dotted line) and the squared root (solid line) behaviours are for, respectively, the constant and energy-dependent beta lattices.

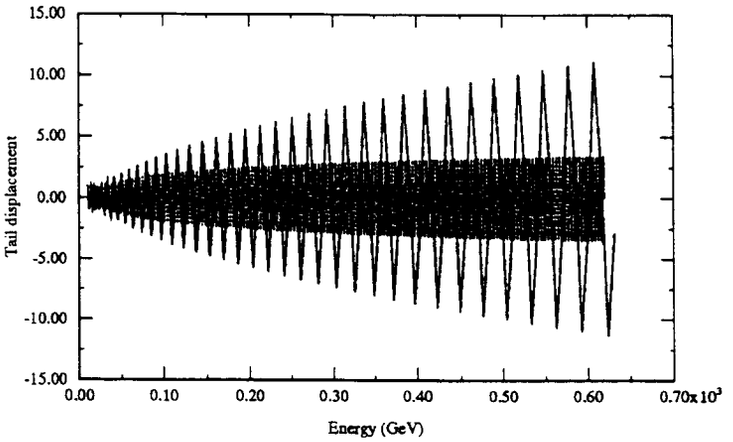


Fig. 10 Normalized displacement of the trailing particle along the linac

The displacement on the Y-axis has been normalized, such that the beta function, the adiabatic damping factor and the initial offset have been taken out.

$$x_n = \frac{x}{\left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0}$$

Table 1 gives an example of a parameter set of a fictitious linear collider. These machine parameters with the dipole wake of Fig. 4 have been used in the previous simulation and will be used throughout this section.

Table 1
Parameter set of a fictitious linear collider

Injection energy	10 GeV
Gradient	100 MV/m
rf frequency	10 GHz
Focusing type	FODO lattice
Phase advance/cell	90 °
Initial average beta	10 m
Bunch length	0.25 mm
Bunch population	$2\ 10^{10}$

For the constant beta lattice, Fig. 11 gives the distortion of the whole bunch versus the longitudinal position z, near the end of the linac, at a location where the tail displacement is maximum. In order to simulate the real bunch charge distribution, the bunch was divided into many macro-particles, evenly spaced with a Gaussian variation of charge. The two particle result is shown for comparison by the two solid circles, sitting at $\pm\ \sigma_z$ with respect to the bunch center.

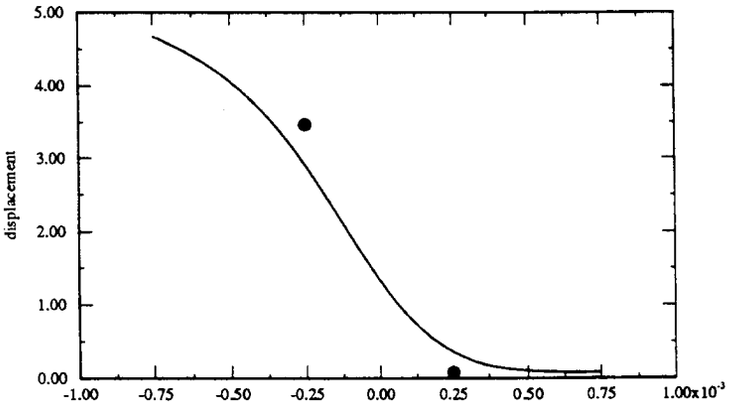


Fig. 11 Distorsion of the bunch vs. the longitudinal position z

The rms emittance in one transverse plane is defined by

$$\varepsilon_x^2 = < x^2 >< x'^2 > - < x\, x' >^2$$

where the coordinates refer to the bunch centroid. This is a "measurement" of the phase space surface occupied by the beam using the second moments of the beam distribution. In a linear focusing structure, the normalized rms emittance $\gamma\varepsilon$ is an invariant. For a small emittance growth caused by the wakefield forces, the tail displacement has to be small compared to the beam size. In the future linear colliders, the normalized emittance $\gamma\varepsilon$ is of the order of 10^{-6} m.rad or even 10^{-8} m.rad. Figure 12 gives the emittance growth for the constant beta lattice, assuming an initial emittance of 10^{-6} m.rad and an offset of 10 μm at injection as the beam moves down the linac with the two-particle model (solid line) and with a real Gaussian bunch distribution (dotted line). The simple model and the simulation with the real bunch are in good agreement. Figure 13 gives the same plot but for the energy-dependent beta lattice, where the beam blow up is more pronounced. In this case, the emittance is growing more than linearly. We can conclude that the two-particle model is only valid in the domain of small beam blow up.

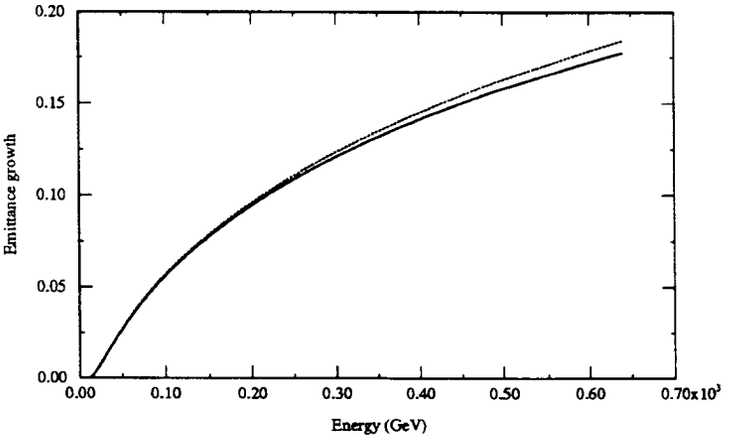


Fig. 12 Emittance growth for the constant beta lattice

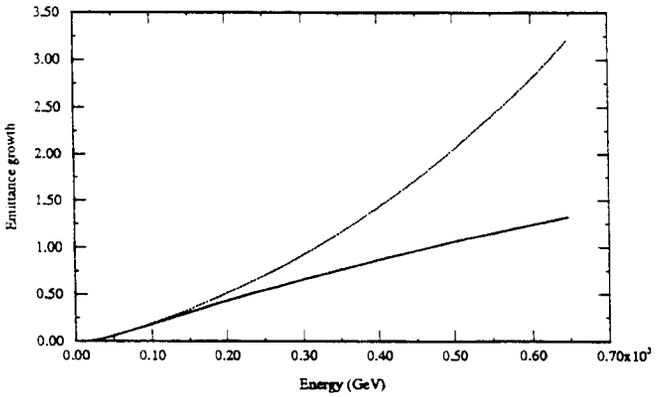


Fig. 13 Emittance growth for the $\sqrt{E}$ beta scaling lattice

For a final energy of 500 GeV and the $\sqrt{E}$ beta scaling, the tolerance on the initial amplitude of the coherent oscillation would have to be reduced to about 1 μm in our machine. Fortunately, this dramatic growth with distance in the amplitude of the oscillation of the tail particle can be suppressed by introducing an energy difference between the head and the tail particles.

2.1.2 Different betatron frequencies

We assume that both particles have different betatron wave numbers $k_1 \neq k_2$.

The head particle has the design momentum p0 while the tail particle has the relative momentum deviation $\delta = (p - p_0)/p_0$. The two particles will have then two different betatron wave numbers k_1 and $k_2 = k_1 + \Delta k$ with Δk given by $\Delta k = k\,\xi\,\delta$. The chromaticity ξ of the lattice describes the energy dependence of the phase advance and is defined by

$$\frac{\Delta\psi}{\psi} = \xi \frac{\Delta p}{p}$$

The chromaticity depends on the focusing system and is

$\xi = -1/2$ for the smooth focusing approximation, and

$\xi = -\dfrac{\tan\mu/2}{\mu/2}$ for a FODO lattice, where μ is the phase advance per cell.

For this case, the equation of motion of particle 2 is that of a harmonic oscillator driven off resonance. The solution of equation (2) is now

$$\frac{\Delta x}{x_0} = \left(\frac{Q_{b/2}\, W_{\perp(2)}}{2k\,E/e} - \Delta k \right) \sin\left[(k_0 + \Delta k/2)s\right] \frac{\sin(\Delta k/2\,s)}{\Delta k/2} \qquad (4)$$

where $\Delta k = k_2 - k_1$ is the difference in wave numbers.

There is now a beating of the relative displacement of the tail instead of a linear growth. We see that, by a proper difference in betatron frequencies, the resonant growth of the tail oscillation can be suppressed through two distinct effects: the amplitude of the beating of the two sine waves can be set to zero; or a minimum of the beating can be made to coincide with the end of the linac. Both cures are called BNS damping, referring to the authors who first introduced the idea [9]. The first cure, which is a local cancellation, is often denoted as "autophasing".

The beating is zero at the end of the linac if the argument of the sine in (4) is a multiple of π. This occurs for a relative energy deviation

$$\delta = \frac{2n\pi}{k\xi L} = \frac{n\lambda_\beta}{\xi L} = \frac{n}{\xi N_\lambda}$$

where n is an integer $\neq 0$, L is the linac length, and
N_λ is the total number of betatron oscillations.

Note that the energy deviation can be positive or negative.

For the autophasing correction technique, the wakefield force is locally cancelled by the difference in betatron frequencies Δk, giving the required energy deviation. This cancellation occurs for

$$\Delta k = \frac{Q_b/2\, W_{\perp(2)}}{2k\, E/e} \quad \Rightarrow \quad \delta = \frac{Q_b/2\, W_{\perp(2)}}{2k^2\, \xi\, E/e} \tag{5}$$

Since the chromaticity is negative, the trailing particle should have lower energy than the driving particle. We note that if the particles have a constant fractional energy deviation δ from the beginning to the end of the linac, a beta function varying as the square root of the energy would be optimal, because the cancellation always occurs

$$k_{(s)} = k_0\sqrt{\frac{\gamma_0}{\gamma_{(s)}}} \Leftrightarrow \beta_{(s)} = \beta_0\sqrt{\frac{\gamma_{(s)}}{\gamma_0}}$$

Figure 14 illustrates the compensation method in the normalized transverse phase space. The coordinates have been normalized such that the trajectory of the unperturbed motion, a pure betatron oscillation, is a circle

$$\begin{pmatrix} x_n \\ x_n' \end{pmatrix} = \sqrt{\frac{\gamma}{\gamma_0}}\sqrt{\frac{\beta_0}{\beta}}\begin{bmatrix} 1 & 0 \\ \alpha & \beta \end{bmatrix}\begin{pmatrix} x \\ x' \end{pmatrix}$$

The trailing particle moves onto this circle (solid line) when the wake and the energy deviation are vanishing. When the wake is switched on but without energy deviation, the trajectory becomes deformed (dotted line) and a phase lag appears. Conversely, when the wake is switched off but with some BNS energy deviation, there is now a phase advance (dashed line). This extra chromatic phase advance of the second particle due to the energy deviation, compensates the phase lag due to the wakefield and the trajectory is again a circle.

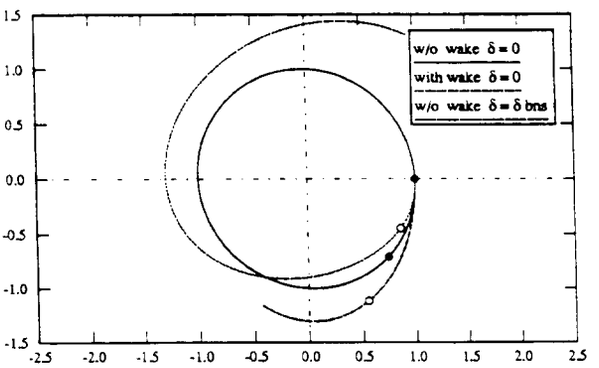


Fig. 14 Illustration of the BNS damping in the normalized phase space

In the smooth-focusing approximation, the cancellation can be permanently perfect and the trailing particle exactly follows the betatron motion of the leading particle. There is in this case no emittance growth at all. In a realistic AG focusing structure, however, the focusing strength, instead of being a smooth function of s, changes abruptly with alternate signs with the distance s due to the periodic array of discrete focusing magnets. The energy difference is then, at best, adjusted such that the cancellation occurs in average over a cell of the focusing lattice. Replacing in (5) the chromaticity term of a FODO cell and the average beta function by the beta value where the wakefield kick occurs, we find the BNS damping condition for a FODO lattice

$$\delta_{(s)} = -\frac{Q_b/2}{2\,E_{(s)}/e}\frac{W_{\perp(2)}}{\tan(\mu/2)}\frac{\mu/2}{}\beta^2_{m(s)} \tag{6}$$

In Fig. 15, the trajectories of the trailing particle are plotted in the normalized phase space for different energy deviations. For $\delta = 0$ (solid circles), the particle is quickly ejected. In the smooth-focusing approximation, the particle follows exactly the circle of the leading particle if the autophasing condition (5) is satisfied (dashed line). With the AG focusing, the local cancellation is not perfect and there is some residual beating even if the BNS damping condition (6) is satisfied (empty cicles). If the energy deviation is slightly decreased (empty triangles) or increased (solid triangles), the beating is more apparent but the motion is stable.

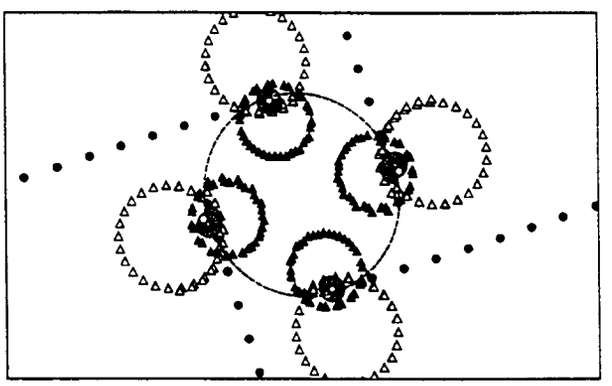


Fig. 15 Trajectories of the tail particle for different energy deviations for AG focusing

2.1.3 Multiparticle analysis

We would like now to extend the BNS criteria of the two-particle model to a continuous charge distribution. We start from the general equation (1) of transverse motion by taking however the smooth-focusing approximation, resulting in

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}x_{(\tau,s)}\right] + k_\beta^2\, x_{(\tau,s)} = \frac{1}{E_{(s)}}F_{\perp(\tau,s)}$$

We assume an energy spread $\delta_{(z)}$, which depends on the longitudinal position within the bunch. The wave number of a particle located at $z = \tau c$ is given by

$$k^2_{(z)} = k_0^2 + \Delta k^2 = k_0^2 + 2k_0^2\,\xi\,\delta_{(z)}$$

where ξ is the chromaticity of the lattice. The previous equation can then be written

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}x_{(\tau,s)}\right] + k_0^2\, x_{(\tau,s)} = \frac{1}{E_{(s)}}F_{\perp(\tau,s)} - \Delta k^2 x_{(\tau,s)}$$

The solution of the homogeneous equation is the unperturbed betatron motion of the "first" particle. All other particles of the bunch exactly track the motion of this particle if the driving term vanishes

$$\Delta k^2 x_{(\tau,s)} = \frac{1}{E_{(s)}/e}\int_{-\infty}^{\tau} d\tau' \, W_{\perp(\tau-\tau')}\rho_{(\tau')}x_{(\tau',s)}$$

Since all particles must have the same displacement at the distance s

$$\Delta k^2 = \frac{1}{E_{(s)}/e}\int_{-\infty}^{\tau} d\tau' \, W_{\perp(\tau-\tau')}\rho_{(\tau')}$$

The integral is just the definition of the bunch potential (times the bunch charge). The intra-bunch energy spread required for BNS damping of a bunch with any charge distribution is hence

$$\delta_{(z)} = \frac{Q_b W_{\perp(z)}}{2k_{(s)}^2 \xi \, E_{(s)}/e} \tag{7}$$

The two-particle model result can be deduced directly from (7) if we take only two particles. Once again, the $\sqrt{E}$ beta scaling permits a constant energy spread, independent of the energy, all along the linac. To obtain the BNS condition in the AG focusing case, it remains only to replace the chromaticity and the beta function by the correct values

$$\delta_{(z)} = -\frac{Q_b W_{\perp(z)}}{2\,E_{(s)}/e}\beta_{m(s)}^2 \frac{\mu/2}{\tan\mu/2} \tag{8}$$

Like the two-particle model, the cancellation is perfect for the entire bunch in the smooth-focusing approximation, but some residual beating remains with a real AG focusing because of an average compensation.

On Fig. 16 is plotted the ideal energy profile which fulfills the autophasing condition (8) for a Gaussian charge distribution with the FODO lattice case.

Figure 17 shows the resulting bunch coordinates at the end of the linac in the transverse phase space. Instead of a large bunch distortion when $\delta = 0$ all particles (solid circles), the beam wraps itself around a point when the previous energy spread distribution is introduced (empty circles).

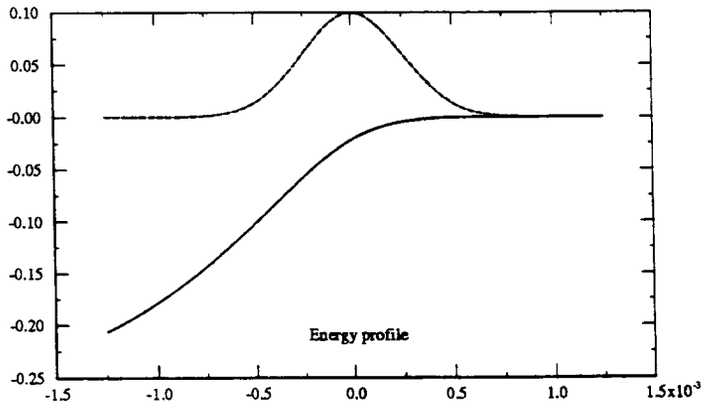


Fig. 16 Ideal energy profile for autophasing condition

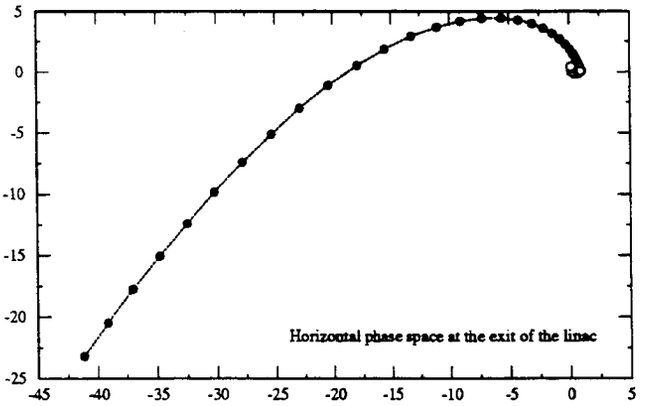


Fig. 17 Bunch coordinates at the end of the linac

But how can we create this ideal energy spread distribution? We now switch on the longitudinal wakefieds and the phase of the bunch with respect to the rf wave is optimized to obtain the minimum dilution. Figure 18 shows the development of the emittance growth in arbitrary units as the beam progresses down the linac with $\delta = 0$ (dotted line), with the ideal bunch energy spread (dashed line) and with the energy spread resulting from the longitudinal wakes and the curvature of the rf wave after optimization of the rf phase (solid line). The bunch is now delayed such that the tail is less accelerated than the head, conversely to the case of a minimum energy spread (section 1.3). Although the real energy deviation along the bunch does not coincide exactly with the ideal one, the final emittance growth is not much larger, because the end particles of the Gaussian bunch located beyong a few standard deviations do not contribute very much to the rms emittance. The situation could be further improved by shaping the bunch with a slightly different charge distribution from the Gaussian distribution, Ref. [10].

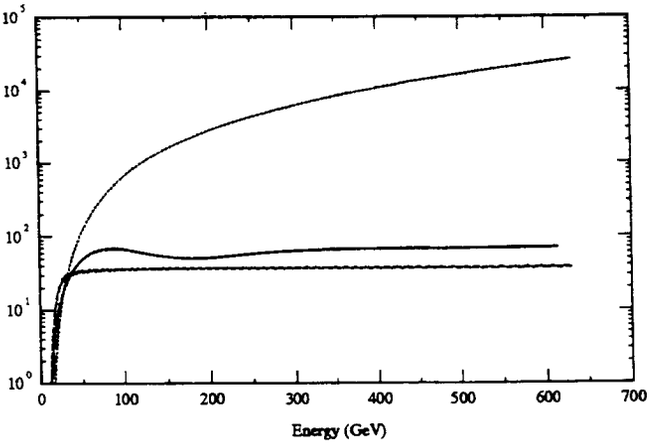


Fig. 18 Emittance growth for a real bunch charge distribution with autophasing

2.2 Effect of misalignment of sections

In the previous study, the beam was assumed to enter off axis into an accelerator with no errors. Now consider the beam injected on axis but with the structures no longer perfectly aligned. Wakefields will be again excited and will distort the core and the back of the bunch, leaving the head of the bunch undamaged (Fig. 19).

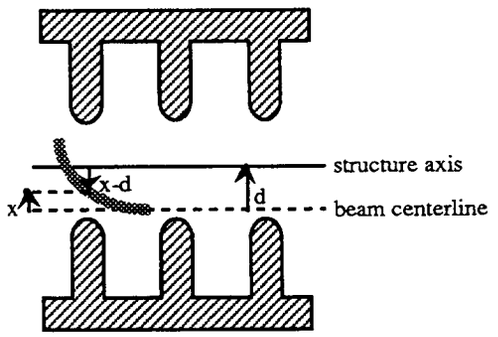


Fig. 19 Misalignment of a structure

If d is the displacement of the structure with respect to the beam centerline and x is the displacement of a particle of the bunch at position s, the offset of the particle with respect to the cavity axis is then $x-d$. The driving term of the general equation of motion (1) can be written as

$$\frac{F_{\perp(\tau,s)}}{E_{(s)}} = \frac{1}{E_{(s)}/e}\int_{-\infty}^{\tau} d\tau'\, W_{\perp(\tau-\tau')}\rho_{(\tau')}\left[x_{(\tau',s)} - d_{(s)}\right]$$

In the two-particle model, the first particle travels down the linac in a straight line since the beam is assumed to be injected without error, while the second particle experiences random kicks due to wakefields excited by the leading particle travelling through randomly displaced structures. At every traversal of the structure j, the trailing particle receives a transverse kick

$$\Delta x'_j = -\frac{Q_{b/2}\; W_{\perp(2)}}{E_j/e} d_j L_j$$

with d_j and L_j the misalignment and the length of the struture j.

The beam displacement at distance s along the linac is the sum of the betatron oscillations excited upstream by the transverse kicks

$$x_{(s)} = \sum_{j}^{s_j<s} R_{12(s_j,s)}\,\Delta x'_j$$

where $R_{12(s_j,s)}$ transforms the angular kicks given at s_j to the displacement at position s

$$R_{12(s_j,s)} = \sqrt{\beta_{(s)}\beta_j}\,\sin\left(\psi_{(s)} - \psi_j\right) \times \sqrt{\frac{\gamma_j}{\gamma_{(s)}}}$$

Here, β_j refers to the beta function at the center of structure j and $\left(\psi_{(s)} - \psi_j\right)$ is the phase advance from the center of structure j to the position s. After substitution, the beam displacement can be written

$$x_{(s)} = -\frac{Q_{b/2}\; W_{\perp(2)}}{\mathcal{E}_0/e}\left(\frac{\beta_{(s)}}{\gamma_{(s)}}\right)^{1/2}\sum_{j=1}^{N_s}\left(\frac{\beta_j}{\gamma_j}\right)^{1/2}\sin\left(\psi_{(s)} - \psi_j\right) d_j L_j$$

where $\mathcal{E}_0$ is the rest energy of the particle.

If we assume an uncorrelated sequence of misalignment errors, the expected value of the squared displacement is

$$< x_{(s)}^2 >= \left(\frac{Q_{b/2}\, W_{\perp(2)}}{\mathcal{E}_0 / e} \right)^2 \frac{\beta_{(s)}}{\gamma_{(s)}} \sum_{j=1}^{N_s} \frac{\beta_j}{\gamma_j} L_j^2 \frac{< d^2 >}{2} \tag{9}$$

where $< d^2 >^{1/2}$ is the rms value of the misalignment.

Both types of focusing can be considered.

i) constant beta lattice

The beta function and the structure length between two quadrupoles are constant and we are left in equation (9) with a sum over the inverse of Lorentz factor that we can approximate by an integral:

$$\sum_{j=1}^{N_s} \frac{1}{\gamma_j} \approx \frac{1}{\Delta\gamma} Ln \frac{\gamma_{(s)}}{\gamma_0} \quad \text{with} \quad \Delta\gamma = \frac{G L_s}{\mathcal{E}_0}$$

where G is the accelerating gradient and L_s is the length of a section

$$< x_{(s)}^2 >= \frac{\beta_{(s)}}{\gamma_{(s)}} \left(\frac{Q_{b/2}\, W_{\perp(2)}}{\mathcal{E}_0 / e} \right)^2 \frac{\mathcal{E}_0}{2G} \beta_m L_s\, Ln \frac{\gamma_{(s)}}{\gamma_0} < d^2 >$$

ii) energy-dependent beta lattice

The beta function and the structure length between two quadrupoles scale like

$$\beta_j = \beta_0 \sqrt{\frac{\gamma_{(s)}}{\gamma_0}} \quad \text{and} \quad L_j = L_{s0} \sqrt{\frac{L_{s(s)}}{L_{s0}}}$$

We are left now in equation (9) with simply a sum over all the structure lengths; that is, approximatly the total length of the accelerator. The expected value of the squared displacement can be expressed as

$$< x_{(s)}^2 >= \frac{\beta_{(s)}}{\gamma_{(s)}} \left(\frac{Q_{b/2}\, W_{\perp(2)}}{\mathcal{E}_0 / e} \right)^2 \frac{\mathcal{E}_0}{2G} \beta_{m0} L_{s0} \left(\frac{\gamma_{(s)}}{\gamma_0} - 1 \right) < d^2 >$$

In both cases, the expected value of the tail displacement has to be small compared to the beam size. We note that if the initial beta function of both focusing schemes are identical, the emittance dilution due to the misalignment effects will be larger for the energy-dependent beta lattice, with the final emittance ratio

$$\frac{\varepsilon_{(\beta \propto \sqrt{E})}}{\varepsilon_{(\beta = const)}} = \frac{\gamma_f / \gamma_0 - 1}{Ln(\gamma_f / \gamma_0)}$$

For example, in a 500 GeV linac with an injection energy of 10 GeV, the emittances ratio would be 12.5.

When simulations are performed, statistical fluctuations are found in the results because the displacements of the structures are set by random numbers with a Gaussian distribution, and though the rms misalignment is the same, the spatial distribution is different and varies from one simulation to the next. A few tens of different machines are then simulated and the average of the results is computed. We now take a real bunch of Gaussian charge distribution and divide it into many macroparticles, also called slices. With the help of a tracking code, the bunch shape and the resulting rms emittance can be computed as the beam progresses down the linac. Figure 20 shows the development of the average emittance for the constant beta focusing. The initial normalized emittance is 10-6 m.rad and the rms misalignment of the sections is 20 μm. The multi-particle result (dotted line) deviates slightly from the two-particle model (solid line). For the $\sqrt{E}$ beta scaling of the lattice, the single-bunch beam blowup is more pronounced and the multiparticle result departs even more from the two-particle model. In this case, the emittance is growing more then linearly, as we can see on the Fig. 21.

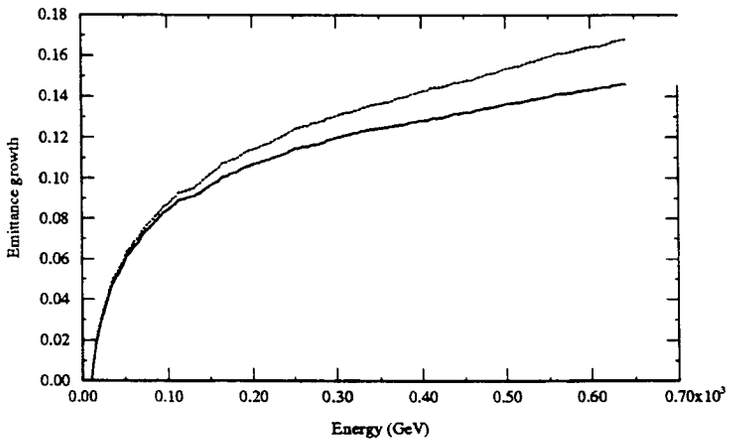


Fig. 20 Emittance growth with structure alignment errors (constant beta lattice)

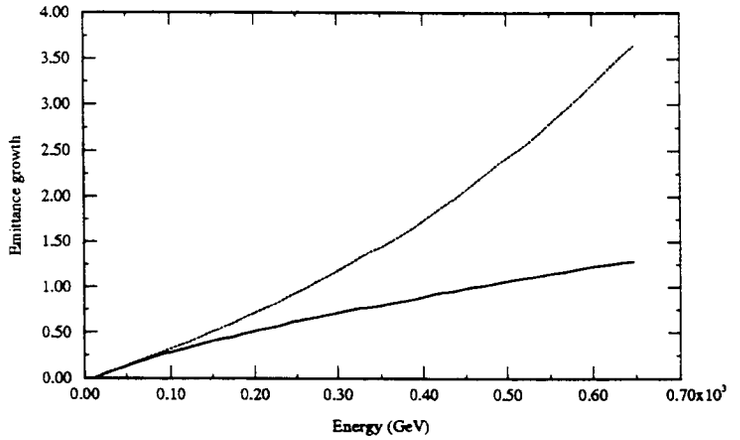


Fig. 21 Emittance growth with structure alignment errors ($\sqrt{E}$ beta scaling)

The discrepancy would have been greater with a stronger wakefield force. We conclude like for the injection error (section 2.1.1), that the two-particle model is only valid in the domain of small beam blowup. In addition, for the alignment errors of sections, the second particle only experiences random kicks and is not resonantly driven since the first particle travels on axis all along the linac. With a real bunch however, which can be considered as a collection of many macroparticles, as soon as the head particles start to deviate from the centerline, they also drive resonantly the trailing particles, like the case for an injection error, if no betatron frequency spread is present.

2.3 A more rigorous method

Before attempting to solve the general equation of transverse motion (1) for the multiple-bunches case, we study again the single-bunch case in a more rigorous fashion than in the previous section.

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}x_{(\tau,s)}\right]+K_{(s)}x_{(\tau,s)}=\frac{1}{E_{(s)}}F_{\perp(\tau,s)}$$

with

$$F_{\perp(\tau,s)}=e\int_{-\infty}^{\tau}d\tau'\,W_{\perp(\tau-\tau')}\rho_{(\tau')}x_{(\tau',s)}$$

This problem has been solved in time domain by using a perturbation method by A. Chao, B. Richter and C. Yao in Ref. [11] and an analytical expression of the asymptotic solution was found for a rectangular distribution of the bunch charge. We give hereafter a different formalism, similar to that one used by K. Yokoya in Ref. [12], which is valid for a uniform charge distribution and involves the Laplace transformation.

Since the driving second member of the previous equation is the convolution of two continuous functions, it is natural to take the Laplace transform. Upon introducing the Laplace transforms of the displacement and of the wakefield force with the Laplace variable p

$$\tilde{x}_{(p,s)}=\int_0^{\infty}d\tau\,x_{(\tau,s)}e^{-p\tau}\quad\text{and}\quad\tilde{F}_{\perp(p,s)}=\int_0^{\infty}d\tau\,F_{\perp(\tau,s)}\,e^{-p\tau}$$

the equation of motion becomes

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}\tilde{x}_{(p,s)}\right]+K_{(s)}\tilde{x}_{(p,s)}=\frac{1}{E_{(s)}}\tilde{F}_{\perp(p,s)}\tag{10}$$

Using the most useful property of the Laplace transform, usually called the convolution theorem, the wakefield force term can be expressed as a simple product

$$\tilde{F}_{\perp(p,s)}/e=\tilde{W}_{\perp(p)}\,\tilde{\rho x}_{(p,s)}$$

where $\tilde{\rho x}_{(p,s)}$ is the Laplace transform of the product of the displacement with the bunch charge distribution.

We would not gain much, except for the case of a uniform charge distribution, because the Laplace transform of the displacement and not of the product then appears explicitly in the expression of the wakefield force. For a rectangular bunch of duration T_p, and of total charge $Q_b=I_0\,T_p$,

$$\tilde{\rho x}_{(p,s)}=\frac{Q_b}{T_p}\tilde{x}_{(p,s)}=I_0\,\tilde{x}_{(p,s)}$$

In this case, the equation is written explicitly in terms of the wake function and the displacement

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}\tilde{x}_{(p,s)}\right]+K_{(s)}\tilde{x}_{(p,s)}=\frac{I_0}{E_{(s)}/e}\tilde{W}_{\perp(p)}\,\tilde{x}_{(p,s)}\tag{11}$$

To solve this differential equation in $\tilde{x}_{(p,s)}$, we have the choice between two different approaches:

- either we smooth the focusing system to obtain a second order differential equation and then use the WKB approximation;
- or we use a perturbation method in which the solution is expanded into a series of the driving wakefield force

Both methods lead to the same result. We arbitrarily choose the WKB solution for the continuous single-bunch and the perturbation method for the multibunch case for completness in this study. The general WKB solution of the equation

$$\frac{d}{ds}\left[A_{(s)}\frac{dx_{(s)}}{ds}\right]+B_{(s)}x_{(s)}=0$$

for moderate s-dependence of A and B is given by

$$x^{\pm}_{(s)}=\frac{1}{(AB)^{1/4}}e^{\pm i\int_0^s (B/A)^{1/2}ds'}$$

The solution is the linear combination of x^+ and x^- that satisfies the initial conditions $x_{(0)}$ and $x'_{(0)}$. Detailed explanations about the WKB approximation can be read in physics books [12]. If we assume further that the wakefield effect is small during one betatron wavelength, or more explicitly

$$k^2_{(s)} >> \frac{I_0}{E_{(s)}}\tilde{\mathbf{W}}_{\perp(p)}$$

Then the general WKB solution of (11) is

$$\tilde{\mathbf{x}}^{\pm}_{(p,s)}=\left(\frac{\beta_{(s)}}{\gamma_{(s)}}\right)^{1/2}e^{i\psi_{(s)}}\,e^{\pm i a_{(s)}\tilde{W}_{\perp(p)}}$$

where the parameter a, for a constant focusing strength, takes the value

$$a_{(s)}=\frac{I_0\beta_m}{2G/e}Ln\frac{\gamma_{(s)}}{\gamma_0}$$

We assume, for example, for the initial condition an offset of the bunch at the injection point

$$x_{(t,0)}=x_0\,\eta_{(t)} \quad \text{where} \quad \eta_{(t)} \text{ is the Heaviside step function}$$

whose Laplace transform is simply $\tilde{\mathbf{x}}_{(p,0)}=x_0/p$. The WKB solution for this initial condition is therefore

$$\tilde{\mathbf{x}}_{(p,s)}=\left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2}e^{i\psi_{(s)}}\,e^{-ia_{(s)}\tilde{W}_{\perp(p)}}\quad\frac{x_0}{p} \tag{12}$$

where the real part of the complex exponential is understood.

We see that, in the Laplace domain, and if we approximate the bunch charge by a rectangular distribution, the bunch displacement at the exit ot the linac is simply the product of the bunch displacement at the entrance by a transfer function, which depends, in particular, on the BBU parameter and on the wake function. This method is valid for any bunch displacement at the entrance and a transverse beam modulation, for instance, could be studied instead of an offset of the whole bunch.

The solution in time domain, which we are interested in, is found by taking the inverse Laplace transform of (12)

$$x_{(\tau,s)} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} dp\ \tilde{\mathbf{x}}_{(p,s)}\, e^{p\tau}$$

where all singularities must be to the left of the line $x = c$.

The solution cannot generally be expressed in closed form. We are, however, most often interested in the asymptotic behaviour of the solution for which the beam blow up starts to be significant. It can be shown that the displacement, by using asymptotic techniques, takes the analytic form

$$x_{(\tau,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} \frac{\eta^{-1/6}}{\sqrt{6\pi}} e^{\frac{3\sqrt{3}}{4}\eta^{1/3}}\ x_0 \cos\left(\psi_{(s)} + \frac{\pi}{12} - \frac{3}{4}\eta^{1/3}\right) \qquad (13)$$

where the BBU strength η, assumed not too small, is defined by

$$\eta_{(\tau,s)} = 2\,a_{(s)}\,W_\tau'\,\tau^2 = \frac{\beta_m I_0}{G/e} Ln\frac{\gamma_{(s)}}{\gamma_0} W_\tau'\,\tau^2$$

The wake function has been approximated by its linear part $W_{\perp(\tau)} = W_\tau'\,\tau$, and the front of the bunch is at $\tau = 0$.

We note that contrary to the two-particle model, which gave a tail displacement growing linearly with the distance s for a non-accelerated beam (or with the logarithm of the energy for an accelerated beam), the growth now contains an exponential factor. This result explains the discrepancy (sections 2.1 and 2.2) between the simulations and the two-particle model predictions, when the emittance growth due to the single-bunch beam breakup is large.

3. THE MULTIBUNCH TRANSVERSE INSTABILITY

The basic mechanism of the cumulative BBU is essentially the same as the one of the single-bunch BBU: as a bunch in a beam pulse (like a "slice" in a bunch) is displaced from the axis of the axi-symmetric accelerating structures, transverse deflecting modes (instead of a short-range transverse wakefield) are excited in the cavities. The trailing bunches are then deflected by the wakefield forces whether they are on-axis or not. The angular deflections transform into displacements through the transfer matrices of the focusing system and these displaced bunches will themselves create wakefields in the downstream cavities of the linac. The subsequent bunches will be further deflected leading to a beam blow up. This phenomenon, called multibunch beam breakup, has been observed at SLAC in 1966, Ref. [14], and since then, numerous papers [15-19] have been published on this subject, attempting to give a quantitative treatment of this problem.

Since all the bunches normally contribute to the process on a very different time scale in comparison with the single-bunch BBU, a simplified model, in the style of the two-particle

model, is no longer adequate to describe the cumulative BBU quantitatively. In some cases the wakefield magnitude has been sufficiently reduced after a time comparable to the bunch separation time by some means, which will be reviewed hereafter. The wakefield generated by each bunch is then felt only by its immediate downstream neighbor and it is a good approximation to consider only the nearest neighbor coupling, Ref. [20]. In the following sections, however, we treat the problem of multibunch BBU in the general case and consider the coupling of the motion of all the bunches.

3.1 The mathematical tools

Let's start again from the general equation of motion (1)

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}x_{(\tau,s)}\right]+K_{(s)}x_{(\tau,s)}=\frac{1}{E_{(s)}}F_{\perp(\tau,s)}$$

with

$$F_{\perp(\tau,s)}=e\int_{-\infty}^{\tau}d\tau'\,W_{\perp(\tau-\tau')}\rho_{(\tau')}x_{(\tau',s)}$$

The current is now composed of a train of bunches with identical charge Q_b and evenly spaced by period T, which is an integral number of rf periods of the accelerating mode. The bunches are considered to be rigid macroparticles, like delta-functions, separated by period T (Fig. 22).

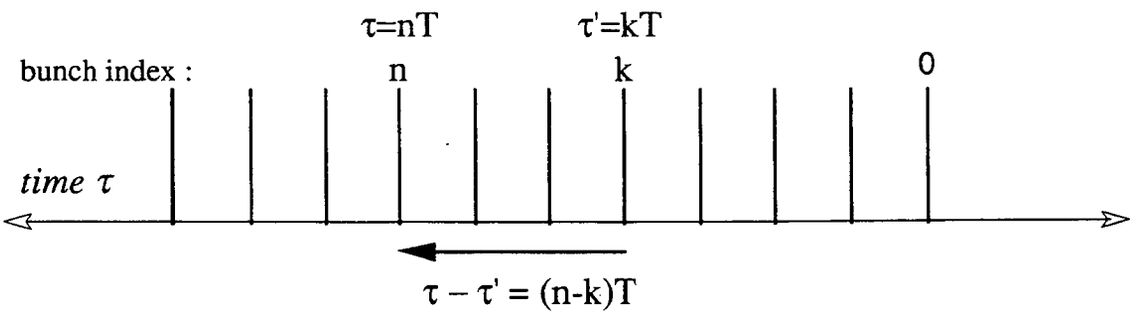


Fig. 22 Train of bunches evenly spaced by period T

The transverse wakefield force experienced by bunch n, spaced $\tau = nT$ from the first bunch, is then

$$F_{\perp(\tau,s)}=eQ_b\sum_{k=0}^{n}W_{\perp[(n-k)T]}x_{(kT,s)} \qquad (14)$$

where the integral has been replaced by a sum over the preceding bunches. We identify the wakefield force (14) as the convolution of two discrete functions of time instead of two continuous-time functions for the single-bunch case (section 3). The study of such discrete systems may be favourably carried through by using the **Z-transform** method. The definition and the properties of the Z-transform can be found in books dealing with the theory of discrete signals (see bibliography).

The displacement, which was a continuous-time function $x(\tau)$ for the single-bunch study, is now considered at discrete values of time and produces the impulse sequence

$$x^*_{(\tau)} = \sum_{n=0}^{\infty} x_{(\tau)}\ \delta_{(\tau - nT)}$$

where $\delta_{(\tau-nT)}$ is the unit impulse occuring at $\tau = nT$.

The Laplace transform of this sampled signal is then

$$\mathcal{L}_{\left[x^*_{(\tau)}\right]} = X^*_{(p)} = \sum_{n=0}^{\infty} x_{(nT)}\ e^{-nTp}$$

The Z-transform of $x_{(\tau)}$ is then defined after the change of variable $z = e^{Tp}$

$$Z_{\left[x^*_{(\tau)}\right]} = \tilde{x}_{(z)} = \sum_{n=0}^{\infty} x_{(nT)}\ z^{-n} \tag{15}$$

The Z-transform of the wakefield force is simply, from the convolution theorem, the product of the Z-transform of the displacement and the wake function

$$\tilde{F}_{\perp(z,s)} = eQ_b\ \tilde{W}_{\perp(z)}\tilde{x}_{(z,s)} \tag{16}$$

We are from now on concerned with the long-range transverse wakes and we select a dominant deflecting mode (cf section 1)

$$W_{\perp(\tau)} = 2K\, w_{\perp(\tau)}$$

with the amplitude $$2K = \omega r_\perp / Q$$

and a wake function of the form $$w_{\perp(\tau)} = e^{-\tau/T_f} \sin \omega\, \tau$$

where ω, T_f and $r_\perp/Q$ are the angular resonance frequency, the filling time (decay time) and the geometric shunt impedance of the selected resonant mode.

The simplest *cure* we can imagine to cumulative BBU is to make the decay time T_f as small as possible by damping the modes or/and lengthening the bunch spacing T, in such a way that the wakefield excited by a bunch has significantly decayed before the following bunch arrives. This is, however, not so easy to implement in reality, because to reduce the Q's significantly, extra waveguides must couple strongly to the higher-order modes without perturbing the accelerating mode.

The Z-transform of the wake function is readily obtained

$$\tilde{w}_{\perp(z)} = \frac{\sin \omega T}{z\, e^{\varsigma} - 2\cos \omega T + z^{-1} e^{-\varsigma}}$$

where we defined the parameter ς , as the ratio of the bunch spacing T over the filling time $T_f = 2Q/\omega$.

This expression can be written in different forms

$$\tilde{w}_{\perp(z)} = \frac{z\,e^{-\varsigma}\sin\omega T}{(z-z_1)(z-z_2)} = \frac{1}{2i}\left(\frac{z}{z-z_1} - \frac{z}{z-z_2}\right) \tag{17}$$

with the poles, which are complex-conjugate, given by $z_{1,2} = e^{-\varsigma}\, e^{\pm i\omega T}$.

After all these definitions, we can now take the Z-transform of the general equation of motion (1)

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}\tilde{x}_{(z,s)}\right] + K_{(s)}\tilde{x}_{(z,s)} = \frac{1}{E_{(s)}}\tilde{F}_{\perp(z,s)} \tag{18}$$

with $$\tilde{F}_{\perp(z,s)} = eQ_b\,\tilde{W}_{\perp(z)}\tilde{x}_{(z,s)}\,.$$

We are left with a similar equation to the single-bunch case, except the Laplace transforms of the continuous-time functions have been replaced by the Z-transforms of the discrete-time functions. To solve this differential equation, we choose now the perturbation method instead of the WKB approximation.

The solution is expanded into a series of the driving wakefield force

$$\tilde{x}_{(z,s)} = \sum_{n=0}^{\infty}\tilde{x}^{(n)}_{(z,s)} \quad \text{with} \quad \tilde{F}_{\perp(z,s)} = \sum_{n=0}^{\infty}\tilde{F}^{(n)}_{\perp(z,s)}$$

The equation (18) is written in terms of the series expansion

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}\sum_n\tilde{x}^{(n)}_{(z,s)}\right] + K_{(s)}\sum_n\tilde{x}^{(n)}_{(z,s)} = \frac{1}{E_{(s)}}\sum_n\tilde{F}^{(n)}_{\perp(z,s)}$$

Since the 0-order solution is given for a vanishing driving force,

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}\tilde{x}^{(0)}_{(z,s)}\right] + K_{(s)}\tilde{x}^{(0)}_{(z,s)} = 0$$

the solution of order n is driven by the wakefield excited by the solution of order n-1

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}\tilde{x}^{(n)}_{(z,s)}\right] + K_{(s)}\tilde{x}^{(n)}_{(z,s)} = \frac{1}{E_{(s)}}\tilde{F}^{(n-1)}_{\perp(z,s)} \tag{19}$$

This differential equation can be solved by means of the Green's function $G_{(s,s')}$ of the focusing system, defined by

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}G_{(s,s')}\right] + K_{(s)}G_{(s,s')} = \delta_{(s-s')}$$

The Green's function is the response of the system to a unit impulse deflection. This is just the R_{12} transport element (section 1.4) of the accelerator lattice.

The solution of (19) is then the superposition of the system's response to each differential deflection

$$\tilde{x}^{(n)}_{(z,s)} = \int_0^s ds'\, G_{(s,s')} \frac{1}{E_{(s')}} \tilde{F}^{(n-1)}_{\perp(z,s')}$$

Replacing the wakefield force by the expression (16), the solution of order n is finally obtained from an integral over the solution of order n-1

$$\tilde{x}^{(n)}_{(z,s)} = \frac{Q_b}{\mathcal{E}_0/e} \tilde{W}_{\perp(z)} \int_0^s ds' \frac{G_{(s,s')}}{\gamma_{(s')}} \tilde{x}^{(n-1)}_{(z,s')} \tag{20}$$

with the Green's function of the focusing system (do not forget the adiabatic damping!)

$$G_{(s,s')} = R_{12(s,s')} = \sqrt{\beta_{(s)}\beta_{(s')}} \sin\left(\psi_{(s)} - \psi_{(s')}\right) \times \sqrt{\frac{\gamma_{(s')}}{\gamma_{(s)}}}$$

Similarly to the single-bunch BBU, we distinguish a coherent betatron oscillation induced by an offset of all bunches at injection and the misalignment of structures, as the main sources of off-axis trajectories.

3.2 Effect of a coherent betatron oscillation

We assume that all of the bunches are launched with the initial offset x_0 at injection. To know the solution of order n, we will have to calculate the preceding orders.

0th order solution

The solution of order 0 is of course the unperturbed motion, a pure betatron oscillation. The beam is launched at $\tau = 0$, so that the displacements of the bunches are given for discrete times of the function

$$x^{(0)}_{(\tau,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2} \left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0 \cos\psi_{(s)}\ \eta_{(\tau)}$$

where $\eta_{(\tau)}$ is the usual Heaviside step function, with the Z-transform $\frac{z}{z-1}$

The Z-transform of the 0th-order solution is hence

$$\tilde{x}^{(0)}_{(z,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2} \left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0 \cos\psi_{(s)} \frac{z}{z-1}$$

1st-order solution

In accordance with (20), the 1st-order solution is calculated from the free betatron oscillation solution

$$\tilde{x}^{(1)}_{(z,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2} \left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0 \frac{z}{z-1} \frac{Q_b}{\mathcal{E}_0/e} \tilde{W}_{\perp(z)} \times \int_0^s ds' \frac{\beta_{(s')}}{\gamma_{(s')}} \sin\left(\psi_{(s)} - \psi_{(s')}\right) \cos\psi_{(s')}$$

We use the trigonometric identity

$$2\sin\left(\psi_{(s)}-\psi_{(s')}\right)\cos\psi_{(s')}=\sin\psi_{(s)}+\sin\left(\psi_{(s)}-2\psi_{(s')}\right)$$

Assuming many betatron oscillations in the linac, we drop the quickly oscillating terms, which will average to zero

$$\tilde{x}^{(1)}_{(z,s)}=\left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2}x_0\sin\psi_{(s)}\frac{Q_b}{2\mathcal{E}_0/e}\tilde{W}_{\perp(z)}\frac{z}{z-1}\int_0^s ds'\frac{\beta_{(s')}}{\gamma_{(s')}}$$

We note that the 1st-order solution oscillates with the distance s in quadrature with the 0th order solution. The integral appearing in the solution depends on the focusing lattice and has been already estimated in Section 1 for both focalisation types,

$$\int_0^s ds'\frac{\beta_{(s')}}{\gamma_{(s')}}=\beta_m\frac{\mathcal{E}_0}{G}Ln\frac{\gamma_{(s)}}{\gamma_0}\qquad\text{for the constant beta}$$

$$\int_0^s ds'\frac{\beta_{(s')}}{\gamma_{(s')}}=2\beta_{m0}\frac{\mathcal{E}_0}{G}\left(\sqrt{\frac{\gamma_{(s)}}{\gamma_0}}-1\right)\qquad\text{for the energy-dependent beta}$$

The 1st-order solution can be written in a more compact form

$$\tilde{x}^{(1)}_{(z,s)}=\left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2}x_0\sin\psi_{(s)}a_{(s)}\,\tilde{w}_{\perp(z)}\frac{z}{z-1}$$

where we defined the dimensionless parameter $a_{(s)}$ characterizing the "BBU strength" as

$$a_{(s)}=\frac{Q_b\beta_m}{2\,G/e}\omega r_\perp/Q\,Ln\frac{\gamma_{(s)}}{\gamma_0}\qquad\text{for the constant beta}$$

$$a_{(s)}=\frac{Q_b\beta_{m0}}{G/e}\omega r_\perp/Q\left(\sqrt{\frac{\gamma_{(s)}}{\gamma_0}}-1\right)\qquad\text{for the energy-dependant beta}$$

For $a<<1$, the series expansion may be stopped at the first order term. The solution is the response of a second-order system to a step function

$$\tilde{x}^{(1)}_{(z,s)}=H_{(z,s)}\frac{z}{z-1}$$

with the transfer-function

$$H_{(z,s)}=\left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2}x_0\sin\psi_{(s)}\,a_{(s)}\,\tilde{w}_{\perp(z)}$$

The transfer-function points out that the system behaves like a single cavity filling the whole accelerator for the first-order solution.

n^{th}-order solution

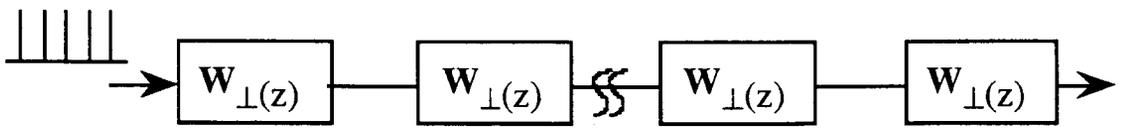
We now carry the calculation of the integral (20) to higher orders. With the assumption that the BBU is moderate during one betatron wavelength, we can again drop the rapidly oscillating terms. We find the Z-transform of the bunch train displacement for order n as the nth-power of the Z-transform of the wake function

$$\tilde{x}^{(n)}_{(z,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0 e^{i\psi_{(s)}} \frac{a^n_{(s)}}{i^n n!} \tilde{G}_{n(z)} \tag{21}$$

where the real part is understood and with the z-dependent function

$$\tilde{G}_{n(z)} = \tilde{w}^n_{\perp(z)} \frac{z}{z-1} \tag{22}$$

The solution of order n is similar to the response to a step function of a system, which consists of a cascade of n cavities.



Finally, we recognize in the n^{th} order solution the n^{th} term of the expansion of the exponential function, so that the complete solution can be written simply

$$\tilde{x}_{(z,s)} = \sum_{n=0}^{\infty} \tilde{x}^{(n)}_{(z,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0 e^{i\psi_{(s)}} \tilde{H}_{(z)} \tag{23}$$

The z-dependent function is

$$\tilde{H}_{(z)} = e^{-i a_{(s)} \tilde{w}_{\perp(z)}} \frac{z}{z-1} \tag{24}$$

In order to get the beam displacement in time domain, it remains to evaluate the inverse Z-transform. The evaluation of the inverse Z-transform of the expression (23), in general, is not possible in closed analytical form. We can however circumvent the difficulty by using the asymptotic techniques, valid once the blowup is significant, or by performing the inverse transform upon the n^{th}-order solution (21). It is found that the cumulative BBU can be characterized by three regime: the first corresponds to an exponential increase of bunch displacement with time, the second occurs when the displacement is maximum, while the third is the steady-state regime.

3.2.1 Steady-state regime

The occurrence of the steady-state regime depends on the decay time of the wakefields and is reached when long trains of bunches are accelerated. The steady-state solution is readily obtained from the complete solution (23). The bunch displacement for an infinite time is given by the final-value theorem of the Z-transform

$$\lim_{t\to\infty} x^*_{(t,s)} = \lim_{z\to 1} (z-1)\tilde{x}_{(z,s)}$$

Taking the real part of (23) with the previous limit, we finally obtain for the steady-state solution

$$x_{(\infty,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0 \cos\left(\psi_{(s)} - a_{(s)} p_r\right)$$

where the function p_r is defined as $p_r = \tilde{w}_{\perp(1)} = \dfrac{1}{2}\dfrac{\sin \omega T}{\cosh \varsigma - \cos \omega T}$

Since the cosine is bounded, the steady-state is mathematically stable, meaning that the solution will oscillate with the distance s along the linac. This result is a consequence of the previous assumption of a moderate BBU during one betatron wavelength, which is true for most of the practical cases, especially for a collider linac. Conversely if the focusing strength were very poor in comparison with the BBU strength, an exponential growth of the steady-state solution with the distance s could be obtained, Ref [17].

Due to the discrete nature of the beam current, the function p_r exhibits resonances when the beam bunching is coherent with deflecting mode frequency. This resonant function is of main concern, both for the steady-state and the transient BBU and is plotted on Fig. 23 for Q = 100 and Q = 1000.

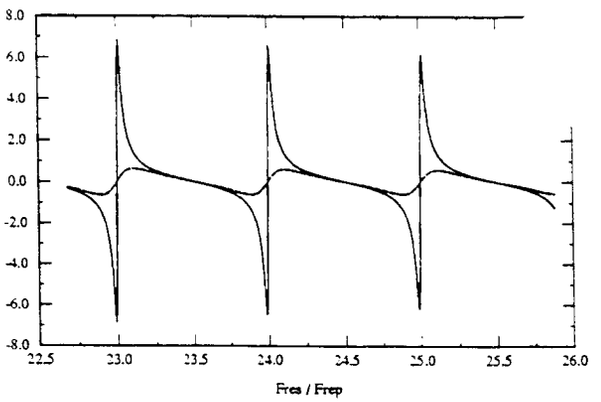


Fig. 23 Function p_r vs. mode and bunch-frequencies ratio

This function reaches its peak values $p_r = \pm\dfrac{Q}{2n\pi}$ at $\omega = n\omega_b\left(1 \pm \dfrac{1}{2Q}\right)$ and is vanishing at the zero-crossings $\omega = n\omega_b$ and $\omega = \left(n + \dfrac{1}{2}\right)\omega_b$

The *second cure*, suggested in Ref. [18], lies in placing the bunches near zero-crossings of the wake function. If the frequency of the deflecting mode were perfectly tuned to the bunch repetition frequency, the displacement of all the bunches would be zero. With a small tuning error however, the BBU effect is expected to be larger when the frequency ratio is an integer than when it is a half integer, because the contributions to the excitation of the dipole wake by the bunches have the same signs in the first case and alternate signs in the latter case. This method, however, is very sensitive to a slight mode detuning and assumes that the dipole wake is strongly dominated by only one resonant mode.

The following goal is to determine the transverse displacement of the bunches, before the steady state is reached. We can obtain two different representations of the complete solution. Either we look for the exact solution by calculating the inverse Z-transform for each order of the infinite sum or we look directly for an approximated form of the complete solution by using the asymptotic techniques.

3.2.2 Transient regime: the exact solution

From the expression of the n^{th} order solution (21), we deduce the displacement of bunch k in time domain

$$x^{(n)}_{(kT,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0\, e^{i\psi_{(s)}}\,\frac{a^n_{(s)}}{i^n n!}\,G_{n(kT)}$$

where $G_{n(kT)}$ is the inverse Z-transform of $\tilde{G}_{n(z)}$

This time-dependent function can be evaluted by using, for example, the partial-fraction expansion procedure. After some tedious algebra, we obtain finally for the bunch $k \geq n$ (see Appendix 3)

$$G_{n(kT)} = p_r^n + e^{-k\tau}\sum_{p=0}^{n-1}\sum_{l=0}^{p}(-1)^p\, C_{n-p-1}^{k-p-1}\, C_l^{l+n-1} \times e^{-\tau}\,\frac{\cos\left[(k-l+1)\omega T-(n+l)\pi/2+(p-l+1)\phi\right]}{2^{n+l-1}\,A^{p-l+1}(\sin\omega T)^l} \tag{25}$$

where the amplitude A and the phase ϕ are defined by the relationships

$$A e^{i\phi} = z_2 - 1 = e^{-\varsigma}e^{-i\omega T} - 1 \quad \text{and} \quad A = \left(2e^{-\varsigma}\right)^{1/2}\left(\cosh\varsigma - \cos\omega T\right)^{1/2}$$

The first term of (25) is the n^{th} power of the resonant function and provides the steady-state, while the second term represents the transient displacements, and decays with the time constant of the deflecting mode. The complete solution is the sum over all the different orders

$$x_{(kT,s)} = \sum_{n=0}^{\infty} x^{(n)}_{(kT,s)}$$

After noting the n^{th} order solution affects only the bunches of index larger than n, the summation can be stopped at the M^{th} term for a pulse containing M bunches. Furthermore, when the BBU strength parameter a is moderate, it is sufficient to keep only a few terms. Figure 24 shows the normalized displacements of the bunches at the exit of a 500 GeV linac for a strong BBU strength parameter (a = 2.7), but a heavily damped structure (Q = 100). The bunch spacing is four times the rf wavelength (bunch frequency of 2.5 GHz) and the parameters of the deflecting mode are the typical ones for a dominant TM11-like mode of a 10 GHz accelerating structure

$$F_r = 15.7\ Ghz \quad \text{and} \quad r_\perp/Q = 4.5\,10^5\ \Omega/m^2$$

The result of numerical simulation is plotted with empty circles. After a growing displacement of the subsequent bunches, the amplitude of the oscillation decays towards the steady-state. The solution given by this analytic formula is shown for different truncations of the sum: up to the orders 2 (solid line), 6 (dashed line) and 12 (solid circles). The more higher-order terms that are taken into account, the more the back of the pulse is recovered. Simulation and analytic results agree perfectly after the 10^{th} order.

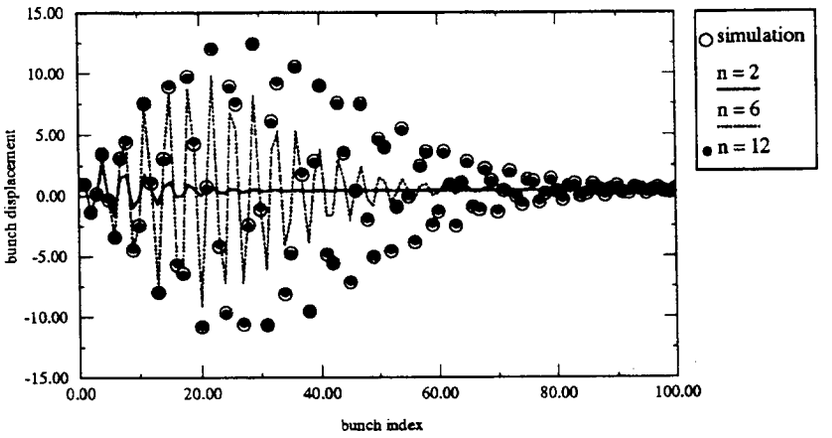


Fig. 24 Normalized displacements of the bunches (simulation and analytical results)

3.2.3 Transient regime: the asymptotic solution

The displacement of a bunch with index k is found from the complete solution (23) by taking the inverse Z-transform, with the complex integral representation

$$x_{(kT,s)} = \frac{1}{2\pi i} \oint_{\Gamma} dz\, \tilde{x}_{(z,s)}\, z^{k-1}$$

where the integration contour Γ encloses all singularities of $\tilde{x}_{(z,s)}$

The solution cannot then be expressed simply in closed form, but the asymptotic behaviour, valid when the blow up is significant, can be obtained by an asymptotic technique such as the method of "steepest descent". A complete description of this method can be found in physics books, see for example the bibliography, and will be briefly recalled below.

The inverse Z-transform of the expression (23) can be written

$$x_{(kT,s)} = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2} \left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0\, e^{i\psi_{(s)}}\, H_{(kT,s)} \tag{26}$$

where the time-dependent function is the inverse Z-transform of (24) and can be written as follows

$$H_{(kT,s)} = \frac{1}{2\pi i} \oint_{\Gamma} dz\, e^{-i a_{(s)} \tilde{w}_{\perp(z)}} \frac{z^k}{z-1} = \oint_{\Gamma} dz\, g_{(z)}\, e^{f_{(z)}} \tag{27}$$

with

$$g_{(z)} = \frac{1}{2\pi i} \frac{1}{z-1}$$

$$f_{(z)} = k\, Ln(z) - i a_{(s)} \tilde{w}_{\perp(z)} = k\, Ln(z) - \frac{a_{(s)}}{2} \left(\frac{z}{z-z_1} - \frac{z}{z-z_2} \right)$$

When BBU is pronounced, the entire contribution of the integrand to the previous integral of (27) comes from the region in which the real part of the exponent $f_{(z)}$ takes on a positive maximum value. Away from this maximum, the integrand becomes negligibly small.

Figure 25, drawn from realistic parameters of a linac, illustrates the method of steepest descent. It is a 3-dimensional plot giving the real part of $f_{(z)}$ when the complex number, $z = x + iy$, is swept. The points at which the maxima occur are the zeros of the derivative and are called the saddle points, because of the shape of the figure. They can be a maximum for one contour, the path of "steepest descent", and a minimum for another. Since the integral is dominated by the region close to the saddle points, the function $f_{(z)}$ can be expanded in a Taylor series. Performing the integration, the integral in (27) can then be approximated by

$$H_{(kT,s)} = \oint_{\Gamma} dz\, g_{(z)} e^{f_{(z)}} \approx \sqrt{2\pi} \frac{g_{(z_s)} e^{f_{(z_s)}}}{\left| f''_{(z_s)} \right|^{1/2}} e^{i\alpha} \tag{28}$$

where z_s is the saddle point

$f''_{(z_s)}$ is the second derivative of the function

α is the phase of the contour for the steepest descent

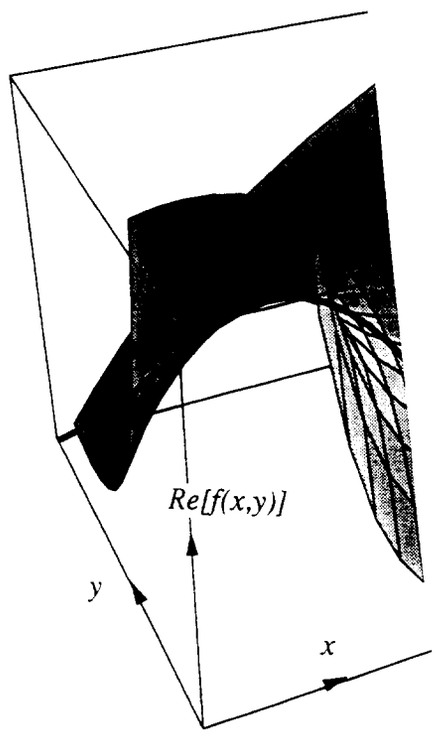


Fig. 25 Illustration of the method of "steepest descent"

The evaluation of $f(z)$ at the saddle points results in exponential growth of the bunches displacements. Only the saddle point generating the largest growth is kept, because the other saddle points give generally negligible growths in comparison.

The saddle points are solutions of the equation

$$\frac{df}{dz} = \frac{k}{z} + \frac{a_{(s)}}{2}\left[\frac{z_1}{(z-z_1)^2} - \frac{z_2}{(z-z_2)^2}\right] = 0$$

If poles z_1 and z_2 coincide, which occurs at the zero-crossings, we find the unperturbated 0th order solution. We assume later on that z_1 and z_2 are not too close. The four roots of the previous equation are approximately given by

$$z_s = \left[1 \pm i\left(\frac{a}{2k}\right)^{1/2}\right] z_1 \quad \text{and} \quad z_s = \left[1 \pm \left(\frac{a}{2k}\right)^{1/2}\right] z_2$$

The saddle points are drawn on the Fig. 26 (solid circles).

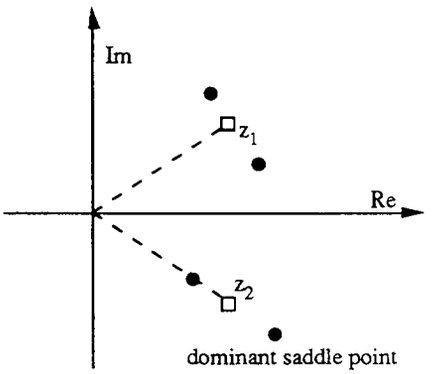


Fig. 26 Saddle points in the z-plane

The dominant saddle point, which generates the largest growth, is of course the one farthest away from the origin and is given by

$$z_s = \left[1 + \left(\frac{a}{2k}\right)^{1/2}\right] z_2$$

Evaluating each of the terms of the integral approximation (28) at the location of the dominant saddle point, we get

$$f_{(z_s)} = -k\varsigma - ik\omega T + (2ak)^{1/2}$$

$$\left|f''_{(z_s)}\right|^{1/2} = \frac{2ke^{\varsigma}}{(2ak)^{1/4}}$$

$$g_{(z_s)} = \frac{1}{2\pi i}\frac{1}{z_2 - 1} = \frac{1}{2\pi i}\frac{1}{Ae^{i\phi}}$$

$$\alpha = \pi/2 - \omega T$$

where the amplitude A and the phase ϕ have been already defined for the calculation of the exact n^{th}-order solution (section 3.2.2)

$$Ae^{i\phi} = z_2 - 1 = e^{-\varsigma}e^{-i\omega T} - 1 \quad \text{and} \quad A = \left(2e^{-\varsigma}\right)^{1/2}(\cosh\varsigma - \cos\omega T)^{1/2}$$

The real part of the exponent $f_{(z)}$ gives the exponential growth of the multibunch BBU, excited by a single deflecting mode

$$\Gamma = \Re e\left\{f_{(z_s)}\right\} = (2ak)^{1/2} - k\varsigma = (2ak)^{1/2} - kT/T_f \tag{29}$$

The asymptotic form of the integral (28) is finally

$$\oint_\Gamma dz\, g_{(z)} e^{f_{(z)}} \approx \frac{1}{\sqrt{2\pi}} \frac{e^{-k\varsigma}}{\left(2e^{\varsigma}\right)^{1/2}(\cosh\varsigma - \cos\omega T)^{1/2}} \frac{(2ak)^{1/4}}{2k} e^{(2ak)^{1/2}} e^{-i\,[(k+1)\omega T + \phi]}$$

The function exhibits rapid oscillations in time with a slower varying amplitude. Since we are especially interested in the beam envelope, which we want to be small, we express the result in terms of the amplitude only. But upon evaluating the expression (27) by the steepest descent method, we omitted the contribution of the pole $z = 1$. Applying the residue theorem, we find this is nothing else but the steady-state solution already calculated previously. The beam envelope, centered around the steady-state displacement is finally

$$\left|x_{(kT,s)} - x_{(\infty,s)}\right| = \left(\frac{\beta_{(s)}}{\beta_0}\right)^{1/2}\left(\frac{\gamma_0}{\gamma_{(s)}}\right)^{1/2} x_0 \times \frac{1}{\sqrt{2\pi}} \frac{e^{-kT/T_f}}{\left(2e^{\varsigma}\right)^{1/2}(\cosh\varsigma - \cos\omega T)^{1/2}} \frac{(2ak)^{1/4}}{2k} e^{(2ak)^{1/2}} \tag{30}$$

The cumulative beam breakup is essentially described by the exponent $(2ak)^{1/2}$, which gives the growth rate of the beam displacement and is directly connected with the BBU strength parameter a. Since this parameter is larger for the constant beta lattice than for the energy-dependent beta lattice, a larger displacement is expected when the focusing strength is progressively relaxed.

We find again the decay term e^{-kT/T_f}, which will make the solution converge towards the steady-state.

We find also again the resonant behaviour when the mode frequency is close to a multiple of the bunch frequency, caused by the function $(\cosh\varsigma - \cos\omega T)$ at the denominator.

Figure 27 shows the normalized displacements of the bunches at the exit of the 500 GeV linac with the same parameter list as Fig. 24. The circles are the results of tracking and the solid line reproduces the asymptotic behaviour of the beam envelope as given by the formula (30). The BBU strength parameter is large (a = 2.7), but the structure is heavily damped (Q = 100), the agreement with the simulation is very good even at the beginning of the pulse.

Figure 28 shows the same plot, except that the BBU strength parameter has been doubled (bunch population of 2 10^{10} instead of 10^{10}). The steady-state has been delayed while the maximum amplitude of the envelope has been amplified by a factor 500.

Figure 29 shows another plot of the simulation and of the asymptotic formula for a = 2.1, but the structure is not damped (Q = 104). The beam envelope is growing still more, but the maximum and hence the steady-state have not been reached, because the bunch train was assumed to be short (80 ns).

The parameters of the fictitious linear collider and of the single deflecting mode used for the last figure are listed in Table 2 and will be used from now on.

Table 2
Parameter set of a fictitious linear collider

Injection energy	10 GeV	bunch population	$2\ 10^9$
Gradient	100 MV/m	bunch rep rate	0.625 GHz
rf frequency	10 GHz	Mode frequency	15.175 GHz
Phase adv./cell	90 °	$r_\perp/Q$	$4.5\ 10^5\ \Omega/m^2$
Constant beta	10 m	Q	104

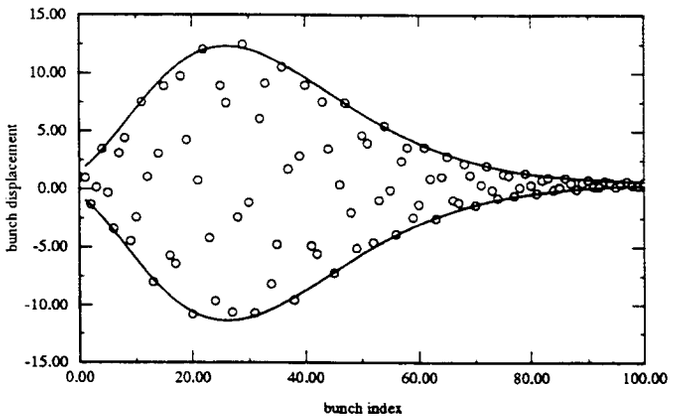


Fig. 27 Asymptotic solution of the bunch displacement (a = 2.7, Q = 100)

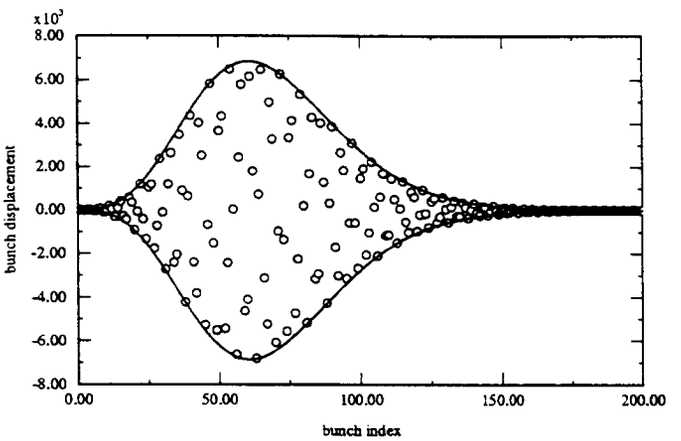


Fig. 28 Asymptotic solution of the bunch displacement (a = 5.4, Q = 100)

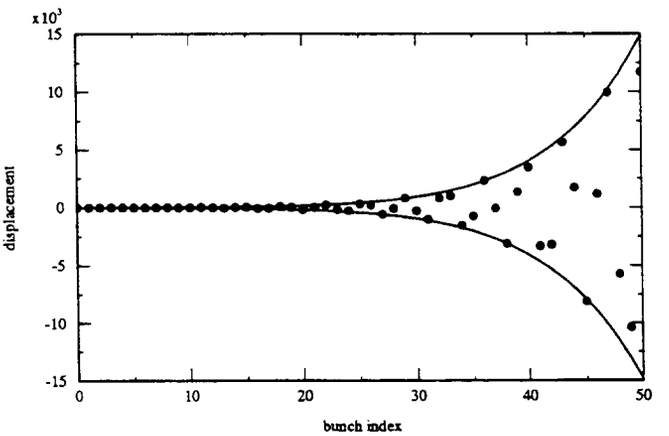


Fig. 29 Asymptotic solution of the bunch displacement (a = 2.1, Q =1 04)

For a small emittance growth caused by the multibunch beam breakup, the maximum bunch displacement has to be small compared to the beam size. With a design normalized emittance $\gamma\varepsilon = 10^{-6}$, the injection error must be lower than 1 nanometer if the structures are not damped.

3.3 The detuning technique

In any case, the BBU driven either by a coherent oscillation or by the misalignment of structures, will lead to a rapid growth of the transverse oscillations of the bunches as they proceed through many accelerating sections. A strong damping is needed to retain the multibunch beam emittance. We have, however, considered so far only a single dipole mode, which oscillates always at the same frequency. If we now introduce a spread in the resonance frequency, the deflecting mode will no longer be excited coherently by the bunched beam. This efficient cure, known as the detuning technique, is contemplated in most designs of future linear colliders because of its simplicity. The costly and somewhat cumbersome waveguide couplers could then be spared. There are two variants of this technique.

1) The frequencies of the dominant modes change from section to section over the whole length of the linac, Ref. [12]. After a bunch has traversed the structures, the resonant wakefields, of different frequency in every section, start to phase-shift. Minimum blow up is obtained when the phases are completely distributed when the next bunch arrives.

2) The detuning takes place inside the structure from cell to cell ("detuned structure") so that the average deflection over the structure effectively damps due to the decoherence of the various cell wakefields. It has been shown [21] that a Gaussian distribution of the cell frequencies, which provides a rapid drop in the wakefield for a given total frequency spread, would be optimal. The physical dimensions of successive cells are varied, such that the synchronous modes of the dominant passbands follow the Gaussian distribution, while keeping the accelerating mode constant. Since the variation from cell to cell is very small and the number of cells in a structure is very large, field calculations with computer cavity codes are difficult, and most of the calculations on such detuned structures have been carried out by using equivalent-circuit models, including or not including the cell-to-cell coupling and the mixing of two adjacent modes [21].

In addition, to enlarge the recoherence time, due to the finite number of frequencies in the distribution, in case of a long bunch train, both variants can be combined by interleaving several detuned structures with slightly different mean frequencies, Ref.[22]. Figure 30 reproduces by courtesy of the authors of Ref. [22] the envelope of the dipole wakefield averaged on a single detuned X-band structure (upper curve), and on four structures with interleaved frequency distributions.

We want to evaluate the effect of α the spread in resonance frequencies on the cumulative BBU process. For that purpose, we assume the BBU effect is small over a sufficiently large number of cavities in case of a detuning from section to section. The beam encounters therefore many distributed frequencies over a distance, where the BBU displacement can be neglected. In case of a detuning from cell to cell of the same structure, this assumption amounts to neglecting the recoherence of the wakefields after approximately a time $\tau = 1/\delta\nu$, where $\delta\nu$ is the mode frequency separation in the distribution. For a distribution of deflecting mode frequencies, the dipole wake function now depends on the distance s.

$$w_{\perp(\tau,s)} = e^{-\omega_{(s)}\tau/2Q} \sin \omega_{(s)}\tau$$

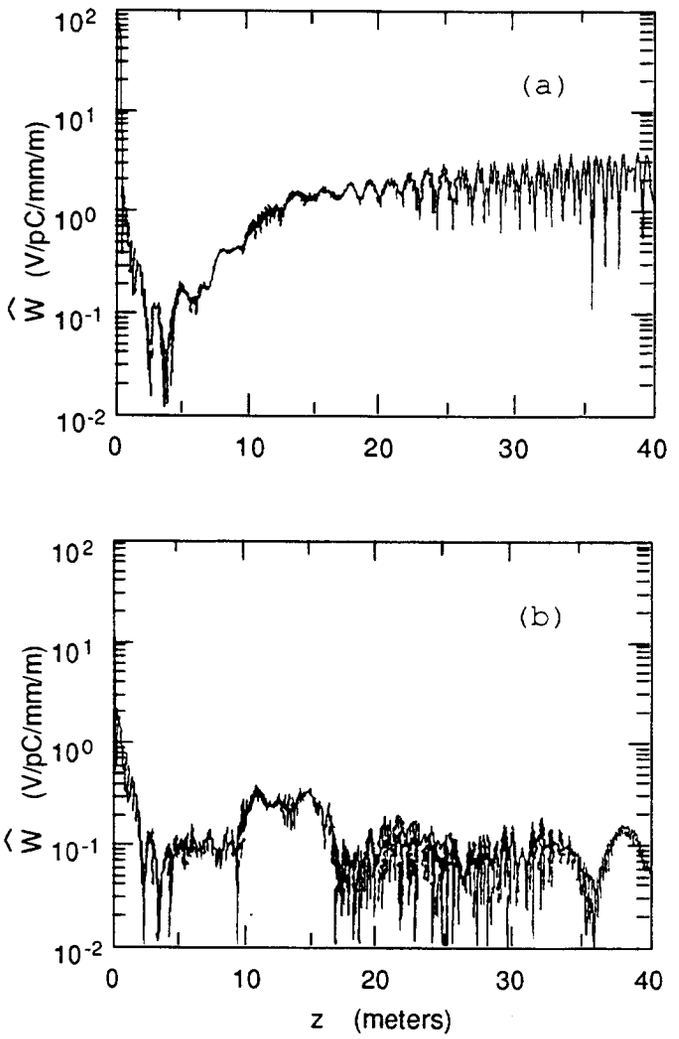


Fig. 30 Envelope of the single-particle wake function $W(z)$ (a) for single structure type, (b) for four structure types with interleaved frequency distributions (reproduced by courtesy of the authors of Ref. [22])

We approximate the driving force induced by the train of bunches by averaging the dipole wake over a distance Δs. This distance is chosen small enough that the bunch displacement is approximately constant, but is large enough that the mode frequency takes all possible values. The expected value of the wake function between s and $s + \Delta s$ depends in that way only on the probability density of the mode frequencies $p_{(\omega)}$

$$\left\langle w_{\perp(\tau)} \right\rangle = \frac{1}{\Delta s}\int_{s}^{s+\Delta s} ds'\, w_{\perp(\tau,s')} = \int_{-\infty}^{+\infty} d\omega\, w_{\perp(\tau,\omega)} p_{(\omega)}$$

After substitution of the resonant dipole wake function, we obtain

$$\left\langle w_{\perp(\tau)} \right\rangle = \int_{-\infty}^{+\infty} d\omega\, e^{-\omega\tau/2Q} \sin\omega\tau\; p_{(\omega)}$$

We consider a narrow probability density peaked around the mean frequency ω_0. After a change of variable, the wake function can be written

$$\left\langle w_{\perp(\tau)}\right\rangle = e^{-\omega_0 \tau / 2Q} \int_{-\infty}^{+\infty} d\omega \, \sin\left[(\omega+\omega_0)\tau\right] \, p_{(\omega+\omega_0)}$$

Since the probability density is an even function around ω_0

$$\left\langle w_{\perp(\tau)}\right\rangle = e^{-\omega_0 \tau / 2Q} \sin\omega_0\tau \int_{-\infty}^{+\infty} d\omega \;\; p_{(\omega+\omega_0)} e^{i\omega\tau}$$

We find that the new wake function is the product of the original wake and the inverse Fourier transform of the probability density, after having been centered about zero.

$$\left\langle w_{\perp(\tau)}\right\rangle = w_{\perp(\tau)} \times 2\pi \, \mathcal{F}^{-1}_{[p(\omega+\omega_0)]} \tag{31}$$

We expect, therefore, a strong dependence of both steady-state and transient regimes on the actual probability density function. We will consider two particular probability densities of mode frequencies; the Lorentzian distribution, because this distribution leads to very simple results and the Gaussian distribution because this is the natural distribution given by the fabrication tolerances and because it provides the stronger roll off of the wake.

The Lorentzian distribution of mean frequency ω_0 and half width at half maximum $\Delta\omega$ has the form

$$p_{(\omega)} = \frac{1}{\pi} \frac{\Delta\omega}{(\omega-\omega_0)^2 + \Delta\omega^2}$$

The inverse Fourier transform is rather simple in this case

$$2\pi \, \mathcal{F}^{-1}_{[p(\omega+\omega_0)]} = e^{-\Delta\omega\,\tau}$$

and the expected wake is identical to the original wake but with a shorter decay time

$$\left\langle w_{\perp(\tau)}\right\rangle = e^{-\omega\tau\left(\frac{1}{2Q}+\frac{\Delta\omega}{\omega}\right)} \sin\omega\,\tau$$

The effect of a spread in deflecting mode frequencies with a Lorentzian distribution is to broaden the real bandwidth of the cavity mode by the width of the distribution. We can then substitute directly an effective Q in the previous results obtained without frequency spread.

$$\frac{1}{Q_{eff}} = \frac{1}{Q} + \frac{2\Delta\omega}{\omega_0}$$

In particular, assuming an infinite real Q, the exponential growth of the bunch displacement, given by the expression (29), has the reduced exponent

$$\Gamma = (2ak)^{1/2} - kT/T_f = (2ak)^{1/2} - k\,\Delta\omega T$$

The blow up is totally suppressed if the excited wakefield has vanished before the next bunch arrives. The bunches do not "talk" to each other if the decay time, of the deflecting

mode is smaller than the bunch spacing. If we neglect the real Q of the mode, the required frequency spread for the suppression of BBU is

$$\frac{2Q_{add}}{\omega_0} < T \quad \Rightarrow \quad \Delta\omega > 1/T \tag{32}$$

For a Gaussian distribution of the spread in the mode frequencies, the result is not as simple as for the Lorentzian distribution and the calculation is more tedious. The inverse Fourier transform of the Gaussian distribution with mean frequency ω_0 and standard deviation σ_ω

$$p_{(\omega)} = \frac{1}{\sigma_\omega \sqrt{2\pi}} e^{-\frac{1}{2\sigma_\omega^2}(\omega-\omega_0)^2}$$

is also of Gaussian type

$$2\pi \, \mathcal{F}^{-1}_{[p(\omega+\omega_0)]} = e^{-\frac{1}{2}\sigma_\omega^2 \tau^2}$$

The expected value of the dipole wake is then

$$\left\langle w_{\perp(\tau)} \right\rangle = e^{-\tau/T_f} \sin\omega\tau \, e^{-\frac{1}{2}\sigma_\omega^2 \tau^2}$$

The result being not so simple as for the Lorentzian distribution, we have to repeat the procedure developed for the case of a single deflecting mode with a constant resonance frequency, except that the wake function has to be replaced by the expected wake in the presence of frequency spread. In particular, we learned that the asymptotic behaviour of the solution is entirely determined by the behaviour of the integrand (27) of the complex integral representation in the vicinity of the dominant saddle point.

$$x_{(kT,s)} = \frac{1}{2\pi i} \oint_\Gamma dz \, \tilde{x}_{(z,s)} z^{k-1}$$

We have then to know the Z-transform of the modified wake function only near the dominant saddle point z_2. It can be shown (Appendix 4) that this Z-transform can finally be written

$$\left\langle \tilde{w}_{\perp(z)} \right\rangle = -\frac{1}{2i} \frac{\sqrt{\pi}}{\sigma_\omega T \sqrt{2}} \operatorname{werf}(iu) \quad \text{with the variable} \quad u = \frac{1}{\sigma_\omega T \sqrt{2}} \frac{z - z_2}{z_2}$$

where the function werf is related to the complex error function $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du$

$$\operatorname{werf}(iz) = e^{z^2} \operatorname{erfc}(z) = e^{z^2} (1 - \operatorname{erf}(z))$$

The exponent of the integral representation is given by

$$f_{(z)} = k \, Ln(z) - i\,a \left\langle \tilde{w}_{\perp(z)} \right\rangle = k \, Ln(z) + \frac{a}{2} \frac{\sqrt{\pi}}{\sigma_\omega T \sqrt{2}} \operatorname{werf}(iu) \tag{33}$$

where we replaced the original wake function by the wake function modified by the presence of frequency spread. The dominant saddle point, which is found by solving $df/dz = 0$, is $z_s = (1 + \sigma_\omega T\sqrt{2}\,u)z_2$, where u is the root of the transcendental equation

$$\sqrt{\pi}\,u\,\mathrm{werf}(iu) = 1 - \frac{2k}{a}(\sigma_\omega T)^2 \tag{34}$$

We do not get the solution in a close form this time because there is no explicit expression for the saddle point. For large values of the variable u, however, we can keep only the first term of the asymptotic expansion of the werf function. The saddle point is then given by

$$\sqrt{\pi}\,u\,\mathrm{werf}(iu) \approx 1 - \frac{1}{2u^2} = 1 - \frac{2k}{a}(\sigma_\omega T)^2$$

We find again the same value calculated without frequency spread

$$z_s = \left[1 + \left(\frac{a}{2k}\right)^{1/2}\right] z_2$$

The growth rate, which is the real part of (33) is then

$$\Gamma = \Re\mathrm{e}\left\{f_{(z_s)}\right\} = -kT/T_f + (2ak)^{1/2} - \frac{a}{2}(\sigma_\omega T)^2\left(\frac{2k}{a}\right)^{3/2}$$

The first two terms represent the solution without frequency spread while the last term due to the frequency spread points out a fast decay to the steady-state. Unfortunately, this approximated expression of the growth rate, also obtained in Ref. [23] by using another formalism, gives too optimistic results and cannot generally be used, because it is only valid for small values of the frequency spread.

$$u >> 1 \;\Rightarrow\; (\sigma_\omega T)^2 << \frac{a}{4k} \tag{35}$$

Figures 31, 32 and 33 show the results of simulation for respectively a weak ($\sigma_\omega T = 5\,\%$), moderate ($\sigma_\omega T = 10\,\%$) and strong ($\sigma_\omega T = 15\,\%$) spread in frequency with a Gaussian distribution. Since the effect of the density probability depends on its repartition along the linac, the results of tracking will show statistical fluctuations, depending on the seed of the random generator. The dotted lines are again the exponential growths without any frequency spread. The symbols "+" are the tracking results of 20 different machines, corresponding to 20 different seeds and the solid circles are the average values of the 20 simulations. The solid lines are the asymptotic solutions (calculated with the exact saddle points without approximation), in agreement with the simulations. In the last case, the BBU has been almost totally suppressed.

3.4 Effect of misalignment of sections

We consider now the misalignment of the structures as starting condition for the multibunch transverse instability. The beam is injected at $\tau = 0$ on-axis while the cavities are displaced $d_{(s)}$ from the beam centerline. The general equation of motion (1) now has the driving force

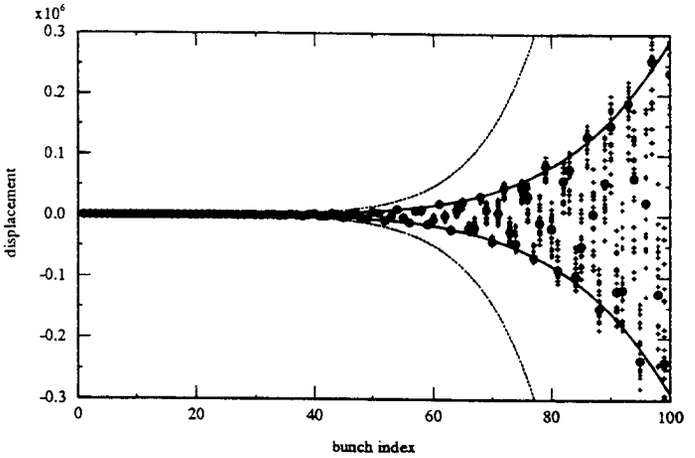


Fig. 31 Norm. bunch displacements with a weak frequency spread ($\sigma_f = 5$ MHz)

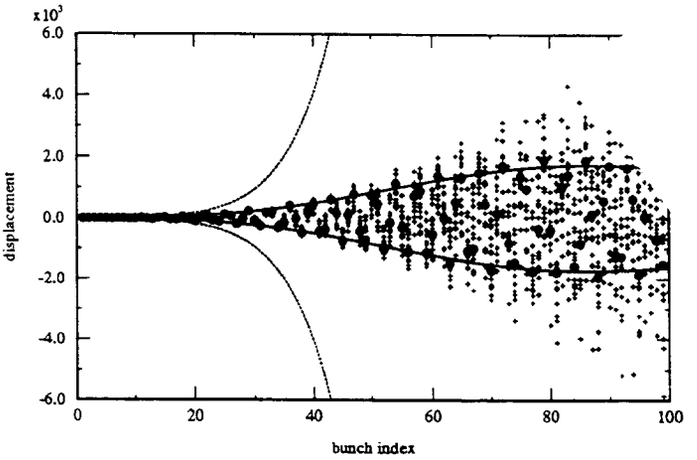


Fig. 32 Norm. bunch displacements with a moderate frequency spread ($\sigma_f = 10$ MHz)

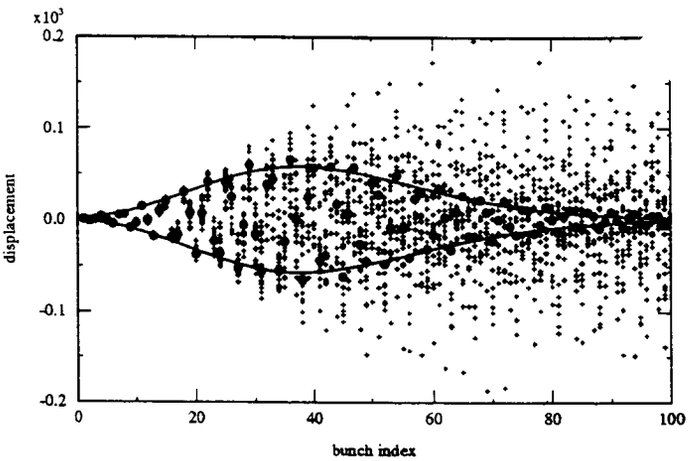


Fig. 33 Norm. bunch displacements with a strong frequency spread ($\sigma_f = 15$ MHz)

$$F_{\perp(\tau,s)} = eQ_b \sum_{k=0}^{n} W_{\perp[(n-k)T]}\left[x_{(kT,s)} - d_{(s)}\,\eta_{(kT)}\right]$$

where $\eta_{(\tau)}$ is the usual Heaviside step function.

We repeat the procedure previously developed for the case of a coherent betatron oscillation (section 3.2). The details of the calculations are given in Appendix 5. The four steps of the procedure are summarized below.

1) Take the Z-transform of the equation of motion

$$\frac{1}{\gamma_{(s)}}\frac{d}{ds}\left[\gamma_{(s)}\frac{d}{ds}\tilde{x}_{(z,s)}\right]+K_{(s)}\tilde{x}_{(z,s)}=\frac{1}{E_{(s)}}\tilde{F}_{\perp(z,s)}$$

with

$$\tilde{F}_{\perp(z,s)}=eQ_b\,\tilde{W}_{\perp(z)}\left[\tilde{x}_{(z,s)}-d_{(s)}\frac{z}{z-1}\right]$$

2) Use the perturbation method to solve the equation

The complete solution, in the Z-domain, is the infinite sum of the different orders

$$\tilde{x}_{(z,s)}=\frac{iQ_b}{\mathcal{E}_0/e}\omega r_\perp/Q\sum_j\sqrt{\beta_{(s)}\beta_{(s_j)}}\frac{e^{i\left[\psi_{(s)}-\psi_{(s_j)}\right]}}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}}\,d_jL_s\,\tilde{H}_{(z)} \tag{36}$$

where d_j is the misalignment of the structure with index j

L_s is the length of one structure

$\beta_{(s_j)}$ has to be evaluated at the center of the structure j located at s_j

with the z-dependent function

$$\tilde{H}_{(z)}=\tilde{w}_{\perp(z)}e^{-i\,a_{(s,s_j)}\tilde{w}_{\perp(z)}}\frac{z}{z-1} \tag{37}$$

3) Take the inverse Z-transform

Coming back to the time domain, the displacement of bunch with index k is found from the complete solution (36) by taking the inverse Z-transform, with the complex integral representation

$$x_{(kT,s)}=\frac{1}{2\pi i}\oint_\Gamma dz\,\tilde{x}_{(z,s)}\,z^{k-1}$$

We substitute the expression (36)

$$x_{(kT,s)}=\frac{iQ_b}{\mathcal{E}_0/e}\omega r_\perp/Q\sum_j\sqrt{\beta_{(s)}\beta_{(s_j)}}\frac{e^{i\left[\psi_{(s)}-\psi_{(s_j)}\right]}}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}}\,d_jL_s\,H_{(kT,s_j)}$$

with

$$H_{(kT,s)}=\frac{1}{2\pi i}\oint_\Gamma dz\,\tilde{H}_{(z,s)}\,z^{k-1}=\oint_\Gamma dz\,g_{(z)}e^{f_{(z)}}$$

4) Apply the method of "steepest descent"

The asymptotic solution can be generated by the method of "steepest descent". We notice that the z-dependent function (37) has the same form as the one (24), already worked out for the injection error, with some minor differences. The function has to be multiplied by $\tilde{w}_{\perp(z)}$; and the BBU strength parameter $a_{(s)}$ has to be replaced by $a_{(s,s_j)}$, which depends on the location of the structure j (written as $a_{(s,s_j)} = a_{(j)}$ for simplicity). We can therefore take the injection error results, taking into account these minor changes.

3.4.1 Transient regime: the asymptotic solution without frequency spread

We first assume a single deflecting mode with a constant resonance frequency.

The dominant saddle point is obviously the same $z_s = \left[1+\left(\frac{a}{2k}\right)^{1/2}\right] z_2$, but a is now function of the location of the misaligned structures. The inverse Z-transform is obtained by taking the injection error result (section 3.2.3) and multiplying it by the wake function at the saddle point

$$\tilde{w}_{\perp(z_s)} \approx \frac{i}{2}\left(\frac{2k}{a_{(j)}}\right)^{1/2}$$

$$\oint_\Gamma dz\, g_{(z)} e^{f_{(z)}} \approx \frac{i}{2\sqrt{2\pi}} \frac{e^{-k\varsigma}}{\left(2e^{\varsigma}\right)^{1/2}\left(\cosh\varsigma - \cos\omega T\right)^{1/2}} \frac{e^{(2ak)^{1/2}}}{(2ak)^{1/4}} e^{-i\left[(k+1)\omega T+\phi\right]}$$

After having reintroduced the contribution of the pole $z = 1$, which is just the steady-state, and taking the real part, the displacement of the bunch k is finally

$$x_{(kT,s)} - x_{(\infty,s)} = -\frac{\omega r_\perp/Q}{2\sqrt{2\pi}} \frac{Q_b}{\mathcal{E}_0/e} \frac{e^{-kT/T_f}}{\left(2e^{\varsigma}\right)^{1/2}\left(\cosh\varsigma - \cos\omega T\right)^{1/2}}$$
$$\times \sum_j \sqrt{\beta_{(s)}\beta_{(s_j)}} \frac{\cos\theta_{(s,s_j)}}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}} \frac{e^{(2a_{(j)}k)^{1/2}}}{\left(2a_{(j)}k\right)^{1/4}} d_j L_s$$

with the angle

$$\theta_{(s,s_j)} = \psi_{(s)} - \psi_{(s_j)} - (k+1)\omega T - \phi$$

Assuming no correlation between the alignment errors, the variance of the bunch displacements around the steady-state is then

$$\left\langle X^2_{(kT,s)} \right\rangle = \frac{1}{8\pi}\frac{\beta_{(s)}}{\gamma_{(s)}}\left(\frac{Q_b\,\omega r_\perp/Q L_s}{\mathcal{E}_0/e}\right)^2 \frac{e^{-2kT/T_f}}{2e^{\varsigma}\left(\cosh\varsigma-\cos\omega T\right)}$$
$$\times\frac{\left\langle d^2\right\rangle}{2}\sum_j \frac{\beta_{(s_j)}}{\gamma_{(s_j)}}\frac{e^{2(2a_{(j)}k)^{1/2}}}{\left(2a_{(j)}k\right)^{1/2}}$$

where $$X_{(kT,s)} = x_{(kT,s)} - x_{(\infty,s)}$$

$a_{(j)} = \dfrac{Q_b\beta_m}{2\,G/e}\omega r_\perp/Q\,Ln\dfrac{\gamma_{(s)}}{\gamma_{(s_j)}}$ for a constant beta lattice

$\left\langle d^2\right\rangle^{1/2}$ is the rms value of the alignment error.

Since the quantity $a_{(j)}$ takes its larger value at the beginning of the linac, the sum over the misaligned structures shows that the errors near the injection point will have much stronger effects than the errors near the linac exit. This is not true for the energy-dependant beta lattice, when the beta function is progressively relaxed with the distance, most of the misaligned structures, from the beginning to nearly the end of the linac, will contribute to the dilution. For a large number of sections, we can approximate the sum by an integral. For a constant β lattice, we find

$$\sum_j \frac{\beta_{(s_j)}}{\gamma_{(s_j)}}\frac{e^{2(2a_{(j)}k)^{1/2}}}{\left(2a_{(j)}k\right)^{1/2}} \approx \frac{\mathcal{E}_0/e}{Q_b\,\omega r_\perp/Q L_s}\frac{1}{k}\left(e^{2(2ak)^{1/2}}-1\right)$$

where $a = a_{(s)}$ is the BBU strength parameter, independant of the index j

We normalize the displacement squared of the bunches by taking out the final beta and Lorentz factor values, such that this quantity has the dimension of an emittance

$$\varepsilon_{n(k)} = \frac{\gamma_{(s)}}{\beta_{(s)}} X^2_{(kT,s)}$$

The expected value of this "bunch emittance" is finally

$$\left\langle \varepsilon_{n(k)}\right\rangle = \frac{1}{8\pi}\frac{Q_b\,\omega r_\perp/Q L_s}{\mathcal{E}_0/e}\frac{e^{-2kT/T_f}}{2e^{\varsigma}\left(\cosh\varsigma-\cos\omega T\right)}\frac{e^{2(2ak)^{1/2}}-1}{2k}\left\langle d^2\right\rangle \qquad (38)$$

We notice the same exponential growth of the bunch displacement squared as for the injection offset. If we compare the effects due to the structure misalignment and the effects due to the injection error, we find

$$\frac{\left\langle \varepsilon_{n(k)}\right\rangle_{align\ error}}{\varepsilon_{n(k)\ inject\ error}} = \left[\frac{L_s}{2\ \mathcal{E}_0/e}\left(\frac{Q_b\,\omega r_\perp/Q\,G/e}{\beta_m\,Ln\,\gamma_{(s)}/\gamma_0}\right)^{1/2}\right]\ \sqrt{k}\ \frac{\left\langle d^2\right\rangle}{x_0^2}$$

The factor in square brackets is generally larger than 1 and of the order of a few tens for typical linear collider designs. For the same blow up, the cavity alignment tolerance will be therefore more stringent than the injection offset tolerance.

Figure 34 gives the "bunch emittance" at the end of the machine with the parameter list of Table 2 with a Gaussian distribution of the alignment errors of unit standard deviation. Since the individual effect depends on the location of the cavity, the numerical result depends on the repartition of the random numbers along the linac. The symbols "+" are the tracking results of 20 different machines, corresponding to 20 different seeds of the random generator, and the circles are the average value of the 20 simulations. The solid line is the plot of the analytical formula (38), in agreement with the simulations. Figure 35 is the same plot but with more heavily damped cavities (Q = 500 instead of 10^4). The bunch displacement stops growing before the pulse end.

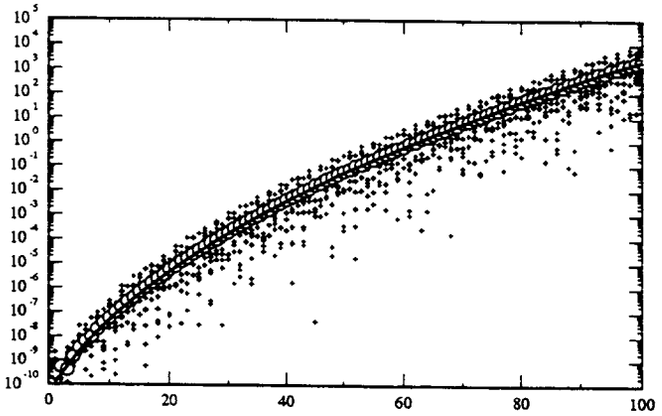


Fig. 34 Bunch emittance for $Q = 10^4$

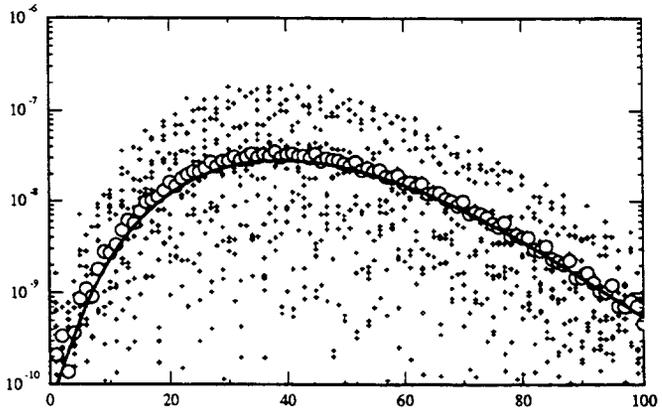


Fig. 35 Bunch emittance for Q = 500

3.4.2 Transient regime: the asymptotic solution with frequency spread

We introduce now some spread in the resonance frequencies. If the probability density is the Lorentzian distribution, we learned that the real bandwidth of the cavity mode is broadened by the width of the distribution. We can substitute directly the effective Q in the expression (38) obtained without frequency spread. If the probability density is the Gaussian distribution, the result is not so simple. The sequence of the bunch displacements is given by

$$x_{(kT,s)} = \frac{iQ_b}{\mathcal{E}_0/e}\omega r_\perp/Q \sum_j \sqrt{\beta_{(s)}\beta_{(s_j)}}\frac{e^{i\left[\psi_{(s)}-\psi_{(s_j)}\right]}}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}}\, d_j L_s\, H_{(kT,s_j)}$$

with $$H_{(kT,s)} = \frac{1}{2\pi i}\oint_\Gamma dz\, \tilde{H}_{(z,s)}\, z^{k-1} = \oint_\Gamma dz\, g_{(z)} e^{f_{(z)}}$$

Since the BBU parameter $a_{(s,s_j)} = a_{(j)}$, relative to the structure j , depends on its location and varies from $a_{(s)}$ for the structure located at the beginning to zero for the structure located to the end, the condition (35)

$$(\sigma_\omega T)^2 << \frac{a_{(j)}}{4k}$$

cannot be satisfied all along the linac. The asymptotic behaviour must then be evaluated with the exact saddle points, roots of the equation (34).

Figure 36 gives the results of simulation and the asymptotic solution for a strong frequency spread ($\sigma_\omega T = 15\,\%$). Two Gaussian distributions, one for the misalignment of structures and one for the deflecting mode frequencies, are generated. We observe therefore statistical fluctuations and several trackings have to be performed with different seeds of the random generator. The symbols "+" give the growth of the subsequent "bunch emittances" at the exit of the linac for 20 different machines. The circles are the average of these 20 simulations, while the solid line is the asymptotic solution of the formula (38), in agreement with the simulations. If we compare with the results obtained without any frequency spread (Fig. 34) , we note that the long-range wakefield effects have been practically suppressed.

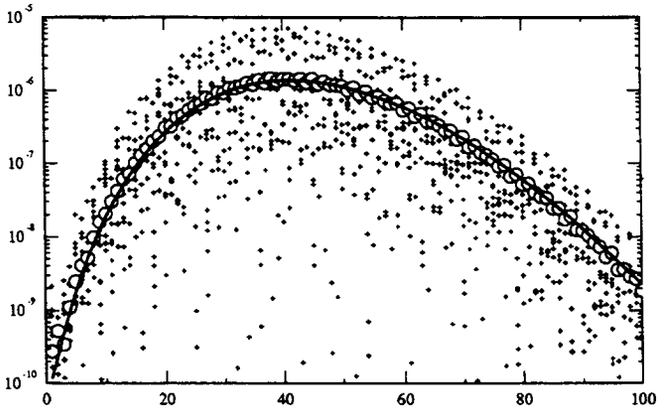


Fig. 36 Bunch emittance for Q = 10^4 with a strong frequency spread (σ_f = 15 MHz)

4. SUMMARY AND CONCLUSIONS

After a brief examination of the longitudinal wakefield effects, we have studied the effects of the transverse wakefields on the emittance growth of an intense bunch and on a train of bunches in a linear accelerator. We have looked especially at the beam instabilities caused by coherent betatron oscillations and by misalignments of the acceleration sections. The famous two-particle model was reviewed first because it permits a good understanding of the single-bunch beam breakup phenomenon, both in the smooth approximation and for real AG focusing. The stabilization of the beam through BNS damping against changes in

injection conditions into the linac has been demonstrated for the two-particle model and for a real bunch charge distribution. A more rigorous method, using a Laplace transform and the WKB approximation, gave the asymptotic expression of the bunch shape for cases with strong wakefields. The multibunch beam breakup, for which no simplified model is available, was treated by means of a discrete Laplace transform, the Z-transform, and a perturbation method. For the case of isolated deflecting modes, the exact solution was derived from the n^{th}-order solution, while the asymptotic solution was calculated by the method of "steepest descent", giving the exponential growth factor. In any case, it appeared that the cumulative beam breakup, driven either by a coherent oscillation or by the misalignment of structures, leads to a rapid growth of the transverse oscillations of the bunches as they proceed through many accelerating sections. Finally, methods for suppressing the dramatic multibunch emittance growth were reviewed in details: damping the resonant transverse modes; tuning the frequency of the dominant dipole mode so that the bunches are located near wake zero-crossings; and last but not least, introducing a distribution of mode frequencies.

A complete list of references would have been longer than this entire lecture and I apologize, in advance, for having omitted many contributions on these subjects. The interested reader who requires details on any of the topics discussed here, should pursue the study, beginning for example with the splendid thesis from T.O. Raubenheimer given in the bibliography.

REFERENCES

[1] R.H. Helm and G.A. Loew, in Linear Accelerators (North-Holland, Amsterdam) 1970, p. 173

[2] A. Mosnier, Nucl. Inst. and Meth. A257 (1987), p.81

[3] P.B. Wilson, AIP Conf. Proc. 184 (AIP, New York, 1989), p.525. Also in SLAC-PUB-4547 (1989)

[4] L. Palumbo, Wakefields, these proceedings

[5] T. Weiland, Nucl. Inst. and Meth. 216 (1983), p.329

[6] Y.H. Chin, "Advances and Applications of ABCI", to be published in Proc. IEEE Part. Acc. Conf., Washington D.C., May 17-20, 1993

[7] R.D. Ruth, SLAC-PUB-4541 (1988)

[8] K.A. Thompson and R.D. Ruth, to be published in Proc. IEEE Part. Acc. Conf., Washington D.C., May 17-20, 1993. Also in SLAC-PUB-6154 (1993)

[9] V.E. Balakin, A.V. Novokhatsky and V.P. Smirnov, Proc. 12th Int. Conf. on High Energy Accelerators, Batavia, 1983 (Fermilab, Batavia, 1984), p.119

[10] V.E. Balakin and V.A. Dolgashev, Proc. Workshop on Emittance Preservation in Linear Colliders, Tsukuba, 1993 (KEK Proceedings 93-13), p.443

[11] A. Chao, B. Richter, C. Yao, Nucl. Instr. Meth. 178 (1980), p.1

[12] K. Yokoya, Cumulative Beam Breakup in Large-scale Linacs, DESY 86-084

[13] P.M. Morse and H. Feshbach, Methods of Theoretical Physics (NY, Mc Graw-Hill book company) 1953

[14] R.B. Neal and W.K.H. Panofsky, Science 152, 1966, p. 1353

[15] W.K.H. Panofsky and M. Bander, The Review of Scientific Instruments, Vol. 39, Number 2, (1968), p. 206

[16] V.K. Neil, L.S. Hall and R.K. Cooper, Part. Accel., 1979, Vol.9, p. 213-222

[17] R.L. Gluckstern, R.K. Cooper and P.J. Channell, Part. Accel., 1985, Vol. 16, p.1 25-153

[18] K.A. Thompson and R.D. Ruth, Physical Review D, Vol. 41 (1990), p. 964

[19] C.L. Bohn and J.R. Delayen, Physical Review A, 45, (1992), p. 5964

[20] C. Adolphsen, SLAC-PUB-5942 (1992)

[21] K.L.F. Bane and R.L. Gluckstern, SLAC-PUB-5783 (1992).

[22] K.A. Thompson, C. Adolphsen and K.L.F. Bane, "Multibunch Beam Break-Up in Detuned Structures", to be published in Proc. IEEE Part. Acc. Conf., Washington D.C., May 17-20, 1993. Also in SLAC-PUB-6153 (1993)

[23] C.L. Bohn and J.R. Delayen, Proc. Linear Acc. Conf, Ottawa, August 24-28, 1992 (AECL-10728)

BIBLIOGRAPHY

Single-bunch instability

T.O. Raubenheimer, "Generation and Acceleration of Low Emittance Flat Beams for Future Linear Colliders", Ph.D. Thesis, Stanford University, SLAC-387 (1991).

K. Bane, "Wakefield Effects in a Linear Collider", Proc. of the U.S. Summer School on High Energy Particle Accelerators, Batavia, Aug. 13-31,1984. Also in SLAC-PUB-4169 (1986).

Multibunch instability

K.A. Thompson and R.D. Ruth, Physical Review D, Vol. 41 (1990), p. 964

K. Yokoya, "Cumulative Beam Breakup in Large-scale Linacs", DESY 86-084, 1986

C.L. Bohn and J.R. Delayen, "Cumulative beam breakup in linear accelerators with periodic beam current", Physical Review A, 45, (1992), p. 5964

Z-Transform techniques

E.I. Jury, Theory and Application of the Z-Transform Method (NY, J. Wiley & Sons) 1964

J.A. Cadzow, Discrete Time Systems (Englewood Cliffs) 1973

The method of steepest descent

G. Arfken, "Mathematical methods for physicists", Orlando, Academic Press, 1985, p. 428

APPENDIX 1

POINT-LIKE WAKE POTENTIALS DEFINITIONS

Consider an ultra-relativistic particle of charge Q_b travelling through an accelerating structure of length L. The longitudinal wake per unit length is defined as the total voltage lost by a test particle following at distance s, divided by the structure length and the charge Q_b

$$W_{z(s)} = -\frac{1}{Q_b L}\int_0^L dz\; E_{z(z,t=(z+s)/c)}$$

with the usual units of V/pC/m.

Similarly, the dipole wake is defined as the transverse momentum kick imparted to a test charge by a driving particle travelling with an offset x. It is a common usage to multiply by the normalization factor $c/(q \times Q_b L)$, c being the light speed, q and Q_b the charges of the test and exciting particles, L the structure length and x the offset of the driving charge

$$W_{\perp(s)} = \frac{c}{q\,x\,Q_b\,L}\Delta p_{\perp(s)} = \frac{1}{x\,Q_b\,L}\int_0^L dz\,\left[E_\perp + (v\times B)_\perp\right]_{t=(z+s)/c}$$

with the usual units of V/pC/m^2.

We note in particular that the transverse wake is divided by x, in order to be independent of the offset of the driving charge.

APPENDIX 2

BETA VALUE AT THE FODO CELL CENTER

For a FODO lattice, the average β and the β value at the center of a half cell are given by

$$\beta_{av} = \frac{1}{2}(\beta_F + \beta_D) = \frac{L_c}{\sin\mu}$$

$$\beta_m = \beta_{av}\frac{1}{1+\sin\mu/2}\left(\frac{\sin\mu}{4\sin\psi_m}\right)^2$$

where β_F and β_D are the beta functions at the focusing and defocusing quadrupoles,
L_c is the cell length,
μ is the phase advance of the cell

ψ_m is the phase advance at the center and is given by

$$\tan\psi_m = \frac{\sin\mu}{4(1+\sin\mu/2)(1-\frac{\sin\mu/2}{2})}$$

APPENDIX 3

INVERSE Z-TRANSFORM OF THE EXACT SOLUTION (MULTIBUNCH INSTABILITY)

To evaluate the inverse Z-transform of the n^{th} solution, we choose the partial-fraction expansion procedure.

The z-dependent part of the n^{th} order solution is

$$\tilde{G}_{n(z)} = \tilde{w}^{n}_{\perp(z)} \frac{z}{z-1} = \left(e^{-\varsigma} \sin \omega T\right)^{n} \frac{z^{n}}{(z-z_1)^{n}(z-z_2)^{n}} \frac{z}{z-1}$$

We make a partial-fraction expansion of $\tilde{H}_{(z)} = z^{n} \tilde{G}_{n(z)}$.

Since z_1 and z_2 are multiple poles, it will be of the form

$$\tilde{H}_{(z)} = \gamma \frac{z}{z-1} + \sum_{p=0}^{n-1} \alpha_{n-p} \left(\frac{z}{z-z_1}\right)^{n-p} + \beta_{n-p} \left(\frac{z}{z-z_2}\right)^{n-p}$$

Knowing the inverse Z-transform of the elementary function

$$\left(\frac{z}{z-a}\right)^{p} \xrightarrow{Z^{-1}} C^{k+p-1}_{p-1} a^{k}$$

where $C^{n}_{m} = \dfrac{n!}{m!(n-m)!}$ is the binomial coefficient

The inverse Z-transform is then

$$H_{(kT)} = \gamma \, \eta_{(kT)} + \sum_{p=0}^{n-1} C^{k+n-p-1}_{n-p-1} \left(\alpha_{n-p} z_1^{k} + \beta_{n-p} z_2^{k}\right) \eta_{(kT)}$$

Finally, using the right-shifting theorem of the Z-transform

$$\tilde{G}_{n(z)} = z^{-n} \tilde{H}_{(z)} \xrightarrow{Z^{-1}} H_{[(k-n)T]} \, \eta_{[(k-n)T]}$$

we obtain for the inverse Z-transform

$$G_{n(kT)} = \gamma \, \eta_{[(k-n)T]} + \sum_{p=0}^{n-1} C^{k-p-1}_{n-p-1} \left(\alpha_{n-p} z_1^{k-n} + \beta_{n-p} z_2^{k-n}\right) \eta_{[(k-n)T]}$$

substituting the values of z_1 and z_2

$$G_{n(kT)} = \gamma \, \eta_{[(k-n)T]} + e^{-kT/T_f} \sum_{p=0}^{n-1} C^{k-p-1}_{n-p-1} \left(\frac{\alpha_{n-p}}{z_1^{n}} e^{i k \omega T} + \frac{\beta_{n-p}}{z_2^{n}} e^{-i k \omega T}\right) \eta_{[(k-n)T]}$$

We note that for $k < n$, the solution is vanishing. The n^{th} order solution affects only the bunches of index larger than n. The first term represents the steady-state and the sum represents the transient displacements, which then decay with the decay time of the resonant mode. The most difficult aspect of the inversion procedure is the evaluation of the coefficients in the expansion.

For the simple pole $$\gamma = \frac{z-1}{z} G_{(z)}\Big|_{z=1}$$

For the multiple poles $$\alpha_{n-p} = \frac{1}{z_1^p\, p!} D^p \left[\left(\frac{z-z_1}{z} \right)^n G_{(z)} \right] \Bigg|_{z=z_1}$$

$$\beta_{n-p} = \frac{1}{z_2^p\, p!} D^p \left[\left(\frac{z-z_2}{z} \right)^n G_{(z)} \right] \Bigg|_{z=z_2}$$

where the operator is defined by $D = z^2 \frac{d}{dz}$

Since the poles of the wake function z_1 and z_2 are complex-conjugate, the coefficients of the expansion are also complex-conjugate. After some algebra, we find

$$\gamma = \frac{1}{(1-z_1)^n (1-z_2)^n}$$

$$\alpha_{n-p} = \left(\frac{z_1}{z_1 - z_2} \right)^n \frac{z_1}{z_1 - 1} (-1)^p \sum_{l=0}^{p} C_l^{n+l-1} \left(\frac{z_2}{z_1 - z_2} \right)^l \left(\frac{1}{z_1 - 1} \right)^{p-l}$$

$$\beta_{n-p} = \left(\frac{z_2}{z_2 - z_1} \right)^n \frac{z_2}{z_2 - 1} (-1)^p \sum_{l=0}^{p} C_l^{n+l-1} \left(\frac{z_1}{z_2 - z_1} \right)^l \left(\frac{1}{z_2 - 1} \right)^{p-l}$$

The solution of order n is finally

$$G_{n(kT)} = \gamma\, \eta_{[(k-n)T]} + e^{kT/T_f} \sum_{p=0}^{n-1} C_{n-p-1}^{k-p-1} \quad 2\, \Re e \left\{ \frac{\alpha_{n-p}}{z_1^n} e^{i k \omega T} \right\} \eta_{[(k-n)T]}$$

which is identical to the expression of section 3.2.2 after substitution of the values of z_1 and z_2.

APPENDIX 4

Z-TRANSFORM OF THE MODIFIED WAKE FUNCTION (MULTIBUNCH INSTABILITY)

The modified wake function in the presence of frequency spread is

$$\left\langle w_{\perp(\tau)}\right\rangle = e^{-\tau/T_f}\ \sin\omega\tau\, e^{-\frac{1}{2}\sigma_\omega^2\tau^2}$$

After expanding the exponential function and knowing the Z-transform of the elementary function

$$\tau^{2n} \xrightarrow{Z} T^{2n} D^{2n}\left(\frac{z}{z-1}\right) \quad \text{with the operator} \quad D = z\frac{d}{dz}$$

we find the Z-transform of the averaged transverse wake function

$$\left\langle \tilde{w}_{\perp(z)}\right\rangle = \frac{1}{2i}\sum_{n=0}^{\infty}\frac{A_{(n)}}{(2n)!}(\sigma_\omega T)^{2n}\left[D^{2n}\left(\frac{z}{z-z_1}\right) - D^{2n}\left(\frac{z}{z-z_2}\right)\right]$$

where z_1 and z_2 are the poles of the original wake, already defined $z_{1,2} = e^{-\varsigma}\, e^{\pm i\omega T}$

and $A_{(n)} = (-1)^n 1.3\ldots(2n+1)$

We check that this expression of the wake for $n = 0$ is identical to the expression (17) when there is no frequency spread. In the vicinity of the dominant sadle point, the previous expression takes the form

$$\left\langle \tilde{w}_{\perp(z)}\right\rangle = -\frac{1}{2i}\sum_{n=0}^{\infty} A_{(n)}(\sigma_\omega T)^{2n}\left(\frac{z_2}{z-z_2}\right)^{2n+1}$$

where we recognize the asymptotic expansion of the werf function

$$\sqrt{\pi}\, u\,\mathrm{werf}(iu) \approx 1 + \sum_{n=1}^{\infty}\frac{A_{(n)}}{\left(2u^2\right)^n}$$

where the function werf is related to the complex error function $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-u^2}\,du$

$$\mathrm{werf}(iz) = e^{z^2}\,\mathrm{erfc}(z) = e^{z^2}\left(1-\mathrm{erf}(z)\right)$$

In the vicinity of the dominant saddle point, the modified wake function can finally be written

$$\left\langle \tilde{w}_{\perp(z)}\right\rangle = -\frac{1}{2i}\frac{\sqrt{\pi}}{\sigma_\omega T\sqrt{2}}\,\mathrm{werf}(iu) \quad \text{with} \quad u = \frac{1}{\sigma_\omega T\sqrt{2}}\frac{z-z_2}{z_2}$$

APPENDIX 5

EFFECT OF STRUCTURE MISALIGNMENTS (MULTIBUNCH INSTABILITY)

Similarly to the case of an injection error, we use the perturbation method to solve this equation of motion. The solution is expanded into a series of the driving wakefield force

$$\tilde{x}_{(z,s)} = \sum_{n=0}^{\infty} \tilde{x}^{(n)}_{(z,s)} \quad \text{with} \quad \tilde{F}_{\perp(z,s)} = \sum_{n=0}^{\infty} \tilde{F}^{(n)}_{\perp(z,s)}$$

The solution of order n is again given by the integral of the product of the Green's function of the focusing system and the wakefield force excited by the solution of preceding order

$$\tilde{x}^{(n)}_{(z,s)} = \int_0^s ds'\, G_{(s,s')} \frac{1}{E_{(s')}} \tilde{F}^{(n-1)}_{\perp(z,s')}$$

but the wakefield force, given by the expression (34), now contains the position error of the structure. The 0^{th}-order solution is the vanishing solution because the beam is assumed to be injected on-axis, without error. This trivial solution describes the motion of the front bunch, which remains on-axis from the beginning to the end of the linac.

$$\tilde{x}^{(0)}_{(z,s)} \equiv 0$$

Since the preceding solution was identically zero, the driving wakefield force of the 1st order solution contains only the position error of the structures

$$\tilde{F}^{(0)}_{\perp(z,s)} = -eQ_b\, \tilde{W}_{\perp(z)} d_{(s)} \frac{z}{z-1}$$

so that the first order-displacement is simply the result of the kicks generated by the randomly displaced cavities, without resonant excitation, conversely to the case of a driving coherent oscillation.

$$\tilde{x}^{(1)}_{(z,s)} = -\frac{\beta^{1/2}_{(s)}}{\gamma^{1/2}_{(s)}} \frac{Q_b}{\mathcal{E}_0/e} \tilde{W}_{\perp(z)} \frac{z}{z-1} \int_0^s ds' \frac{\beta^{1/2}_{(s')}}{\gamma^{1/2}_{(s')}} \sin\left[\psi_{(s)} - \psi_{(s')}\right] d_{(s')}$$

The integral in the previous equation can be replaced for simplicity by a sum over all the structures forming the linac and located before s

$$\tilde{x}^{(1)}_{(z,s)} = -\frac{Q_b}{\mathcal{E}_0/e} \tilde{W}_{\perp(z)} \frac{z}{z-1} \sum_j \sqrt{\beta_{(s)}\beta_{(s_j)}} \frac{\sin\left[\psi_{(s)} - \psi_{(s_j)}\right]}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}} d_j L_s$$

where d_j is the misalignment of the structure with index j

L_s is the length of one structure

$\beta_{(s_j)}$ has to be evaluated at the center of the structure j located at s_j

In order to get the form of the general solution of the multibunch BBU excited by the misalignment of structures, we must go to the second-order solution. Substituting the previous 1st-order expression, the Z-transform of the 2nd-order solution is

$$\tilde{x}^{(2)}_{(z,s)} = -\frac{\beta^{1/2}_{(s)}}{\gamma^{1/2}_{(s)}}\left(\frac{Q_b}{\mathcal{E}_0/e}\right)^2 \tilde{W}^2_{\perp(z)}\frac{z}{z-1}\sum_j d_j L_s \int_{s_j}^{s} ds' \frac{\beta_{(s')}}{\gamma_{(s')}} \sin\left[\psi_{(s)} - \psi_{(s')}\right]\sin\left[\psi_{(s')} - \psi_{(s_j)}\right]$$

We use the trogonometric identity

$$2\sin\left[\psi_{(s)} - \psi_{(s')}\right]\sin\left[\psi_{(s')} - \psi_{(s_j)}\right] = \cos\left[\psi_{(s)} - 2\psi_{(s')} + \psi_{(s_j)}\right] - \cos\left[\psi_{(s)} - \psi_{(s_j)}\right]$$

and we make the same assumption as for the error injection case, e.g. many betatron oscillations so that the rapidly oscillating terms, which average to zero, can be dropped

$$\tilde{x}^{(2)}_{(z,s)} = \frac{Q_b}{\mathcal{E}_0/e}\omega r_\perp/Q\ \tilde{w}^2_{\perp(z)}\frac{z}{z-1}\sum_j \sqrt{\beta_{(s)}\beta_{(s_j)}}\frac{\cos\left[\psi_{(s)} - \psi_{(s_j)}\right]}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}} a_{(s,s_j)} d_j L_s$$

where the BBU strength parameters $a_{(s,s_j)}$ are now starting at the locations of the displaced structures

$$a_{(s,s_j)} = \frac{Q_b\beta_m}{2\,G/e}\omega r_\perp/Q\, Ln\frac{\gamma_{(s)}}{\gamma_{(s_j)}} \qquad \text{for the constant beta lattice}$$

$$a_{(s,s_j)} = \frac{Q_b\beta_{m0}}{G/e}\omega r_\perp/Q\left(\sqrt{\frac{\gamma_{(s)}}{\gamma_0}} - \sqrt{\frac{\gamma_{(s_j)}}{\gamma_0}}\right) \qquad \text{for the } \sqrt{E} \text{ beta scaling}$$

We can imagine, by pursuing the same procedure to higher orders, that the order (n+1) will be given by

$$\tilde{x}^{(n+1)}_{(z,s)} = \frac{iQ_b}{\mathcal{E}_0/e}\omega r_\perp/Q\ \tilde{G}_{n+1(z)}\sum_j \sqrt{\beta_{(s)}\beta_{(s_j)}}\frac{e^{i\left[\psi_{(s)} - \psi_{(s_j)}\right]}}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}}\frac{a^n_{(s,s_j)}}{i^n\, n!} d_j L_s$$

where the real part is understood with the z-dependent function $\tilde{G}_{n+1(z)} = \tilde{w}^{n+1}_{\perp(z)}\dfrac{z}{z-1}$

The complete solution, which is the infinite sum of the different orders, is found again as the response of a sytem of exponential-like transfer function to a step function

$$\tilde{x}_{(z,s)} = \frac{iQ_b}{\mathcal{E}_0/e}\omega r_\perp/Q\ \sum_j \sqrt{\beta_{(s)}\beta_{(s_j)}}\frac{e^{i\left[\psi_{(s)} - \psi_{(s_j)}\right]}}{\sqrt{\gamma_{(s)}\gamma_{(s_j)}}}\, d_j L_s\, \tilde{H}_{(z)}$$

with the z-dependent function $\tilde{H}_{(z)} = \tilde{w}_{\perp(z)}\, e^{-i\, a_{(s,s_j)}\tilde{w}_{\perp(z)}}\dfrac{z}{z-1}$.

HIGH LUMINOSITY

J. Le Duff
Laboratoire de l'Accélérateur Linéaire, Orsay, France

Abstract
The new generation of colliders are required to have high luminosity. Among these colliders are medium energy electron-positron storage rings, called "factories", for which high luminosity is not only useful but absolutely necessary since the energy range has already been covered by existing machines. In their case the designers are aiming at higher luminosities by a factor 10 to 100. After discussing the main luminosity limitations, and the possible ways of increasing these limits, the different factories under consideration, design or construction will be reviewed.

1 INTRODUCTION

Although very high energy accelerators, such as LHC and linear electron-positron colliders, must as well have high luminosities, the present review will be limited to low and medium-energy circular e^+e^- colliders, named "factories".

Factories will operate at C.M. energies already covered by existing storage rings, but with expected luminosities two orders of magnitude higher, in order to produce better statistics to refine the physical properties of known particles. Actually there are three classes of factories under consideration, design or construction. The "PHI" factories have a C.M. energy of 1 GeV and require a luminosity of $10^{32}\text{cm}^{-2}\text{s}^{-1}$, while "Tau-Charm" and "Beauty" factories will have C.M. energies of respectively 4 and 10 GeV and luminosities of 10^{33} to $10^{34}\text{cm}^{-2}\text{s}^{-1}$.

It turns out that increasing the luminosity requires the storing of high circulating currents. Since the beam-beam effect limits the amount of stored particles in a single bunch it is then necessary to increase the number of bunches. Hence, to avoid parasitic crossings there is a real advantage in storing the two beams in separate rings with common straight sections wherever necessary for collisions (detector locations).

However, some PHI factory projects, not yet under construction, escape from this rule and remain based on a single ring, with a single bunch per beam. But, in order to increase the collision rate they minimize the circumference (compact ring), for instance by using high field superconducting dipoles.

2 THE LUMINOSITY AND ITS LIMITATIONS

In circular colliders beams generally consist of equally spaced bunches. When two bunches, having the same dimensions, collide head-on the particles interact and produce physical processes such as scattering, annihilation.

The rate of events for a given process is:

$$\frac{dn}{dt} = k f_r \frac{N_b^+ \, N_b^-}{S} \sigma_T \tag{1}$$

with
$N_b^{\pm}$ = number of particles per bunch for each beam
k = number of bunches in each beam
f_r = revolution frequency
S = $4\pi\sigma_x^*\sigma_y^*$ (Gaussian shapes with r.m.s. σ)
σ_T = total cross section of physical process.

It is usual to write:

$$\dot{n} = \sigma_T L \tag{2}$$

where

$$L = k f_r \frac{N_b^+ N_b^-}{4\pi\sigma_x^*\sigma_y^*} \tag{3}$$

is called the LUMINOSITY which only depends on the machine parameters.

Past experience has proven that the ultimate limitation of e^+e^- colliders is the beam-beam interaction, and that this limit could be described as a constant maximum value, for all machines, of a parameter ξ called the space-charge parameter and which is connected to the amplitude of the space-charge force that one bunch feels from the other during the collision. This parameter, whose maximum value ξ_{max} is of the order of 0.04 ± 0.01, is expressed for both transverse planes of motion as follows:

$$\xi_{x,y} = \frac{N_b\ r_e}{2\pi\ \gamma\ (\sigma_x^* + \sigma_y^*)} \left(\frac{\beta^*}{\sigma^*}\right)_{x,y} \tag{4}$$

having assumed Gaussian transverse beam shapes with r.m.s. beam sizes σ^* at the interaction point. N_b is the number of particles in each bunch, γ the normalized beam energy, r_e the classical electron radius and β^* the betatron envelope function at the crossing point which characterizes the sensitivity of the particle motion to the defect caused by the space-charge force.

The limit on the number of particles per bunch N_{bmax} is reached when ξ reaches ξ_{max} in one plane or another. On most of the existing machines which operate with flat beams ($\sigma_y^* \ll \sigma_x^*$) the limit is first reached in the vertical plane. By properly adjusting the bunch aspect ratio $r = \sigma_y^*/\sigma_x^*$ it is possible to reach the limit in both planes at the same time; in this case $\sigma_y^*/\sigma_x^* = \beta_y^*/\beta_x^*$ (optimum coupling) and corresponds to an optimum luminosity.

As we have seen the beam-beam limit corresponds to a maximum value for the number of particles per bunch, N_{bmax}. Hence using additional bunches will increase the luminosity. If k is the number of bunches per beam and f_r the revolution frequency, the total circulating beam current is:

$$I = k f_r\ e\ N_{bmax} \tag{5}$$

where e is the electron charge and kf_r the collision frequency.

The luminosity, which is the essential machine parameter for physics experiments, can be expressed in a very general form as follows [1]:

$$L_{[\mathrm{cm}^{-2}\mathrm{s}^{-1}]} = 2.17\ 10^{34}\ I_{[\mathrm{A}]} E_{[\mathrm{GeV}]} \left\{ \frac{\xi_x}{\beta_x^*[\mathrm{cm}]} + \frac{\xi_y}{\beta_y^*[\mathrm{cm}]} \right\} \tag{6}$$

The formula remains valid for asymmetric rings provided the terms in the bracket are made the same for the two rings and the beam sizes overlap perfectly at the crossing point. Then the constant product IE can be taken from one ring or another and obviously the higher energy ring will handle less current. For optimum luminosity the $\xi_{x,y}$ values will be taken as ξ_{max}.

From the previous formulae it is seen that the main parameters for optimizing the luminosity are the β^* values and the total circulating current. The first one should be small and the second one high. The total circulating current itself depends on the number of bunches per beam as well as on the maximum number of particles per bunch; the latter depends on the emittances $\varepsilon_{x,y}$ since: $\sigma^* = (\beta^* \varepsilon)^{1/2}$.

Note that from the previous set of formulae one can also extract a practical formula for the maximum number of particles per bunch, in the case of optimum coupling ($r = \sigma_y^*/\sigma_x^* = \beta_y^*/\beta_x^* = \varepsilon_y/\varepsilon_x$):

$$N_{bmax} = 4.36\ 10^{16} E_{[\mathrm{GeV}]} (\sigma_x^* + \sigma_y^*)_{[\mathrm{cm}]} \xi_{max} \left(\frac{\sigma^*}{\beta^*} \right)_{x\ or\ y} \tag{7}$$

How to play with all these parameters is sometimes a matter of feeling but also and essentially a matter of technical and technological limitations.

If for example the β^* are made small and equal in the two transverse planes it is seen from (6) that for the same total current the luminosity can be doubled. This supposes that the emittances are also made equal in both planes which leads to round colliding beams. Recent simulation studies [2] have predicted that higher values for ξ_{max} could be obtained in that case, hence higher luminosity. Though it has never been proven experimentally on existing machines, designers at Novosibirsk have included this fact in their PHI factory project.

The parametrization of the maximum luminosity appearing in the above formulae enable us to define the ingredients for high luminosities:

- Very low β^* optics in the insertion
- Increase of the number of particles per bunch following an increase in emittance
- Increase in total circulating current by having more bunches per beam, but avoiding parasitic crossings.

All these desirable design requirements however lead to design difficulties and technical challenges.

3 MAIN DESIGN APPROACHES

3.1 Low β^* insertion

The aim is to produce very low β values at the crossing points, at least in one plane. A value as low as 1 cm is envisaged in most projects, since it is technically achievable.

In a straight section, starting from a symmetry point, the β's variation is quadratic:

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*} \tag{8}$$

where s is the azimuthal coordinate.

Hence the first quadrupole set, either a doublet or a triplet, needs to be sufficiently close to the collision point to keep the maximum β value low enough.

Following specific user's requirements, the minimum distance of the first quadrupole from the collision point should be of the order of 1 meter which brings the maximum β value to the order of 100 meters in the case of a flat beam ($\beta_y^* = 1$ cm), and 400 meters in the case of a round beam ($\beta_x^* = \beta_y^* = 1$ cm) as seen for example on Fig. 1 where the free space is 0.8 meter.

For the flat beam case such a low β^* insertion design ends with reasonable natural chromaticities, eventually allowing for two insertion regions.

For the round beam case the chromaticities become quite high and are difficult to compensate with sextupole arrangements that could provide good dynamic acceptance. With a single mini-β insertion the operation of the machine is already a problem. Increasing both β^* would certainly facilitate machine operation but at the same time considerably reduce the advantage of the round beam scheme. Notice also that the triplet is necessary in the case of a round beam and corresponding quadrupole lengths have to be reduced to limit β_{max} ; this leads to much higher quadrupole gradients which give no other choice than the use of superconducting magnets.

The increase in luminosity coming from a low β^* value is only valid provided the bunch length is small enough that all collisions benefit from this low value. In other words the β value should not change too much over the collision length. This requirement is generally expressed as:

$$\sigma_z \leq \beta^* \tag{9}$$

where σ_z is the r.m.s bunch length for each colliding bunch.

With standard RF system frequencies (350 to 500 MHz) the required accelerating voltage, to bring the r.m.s bunch length down to 1 cm, becomes relatively high. In addition high luminosities need high bunch currents which lead to large linear bunch densities and necessitate particular technological efforts to lower the ring impedances to avoid any types of instability.

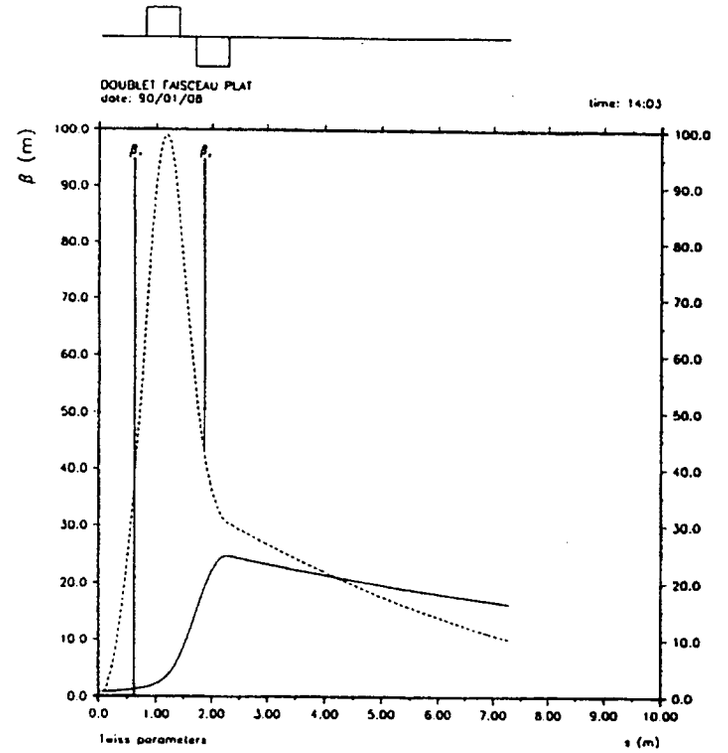
Among other reasons to limit the decrease of β^*, is the necessity to avoid too much background from synchrotron radiation generated by the strong insertion quadrupoles.

3.2 Transverse emittance

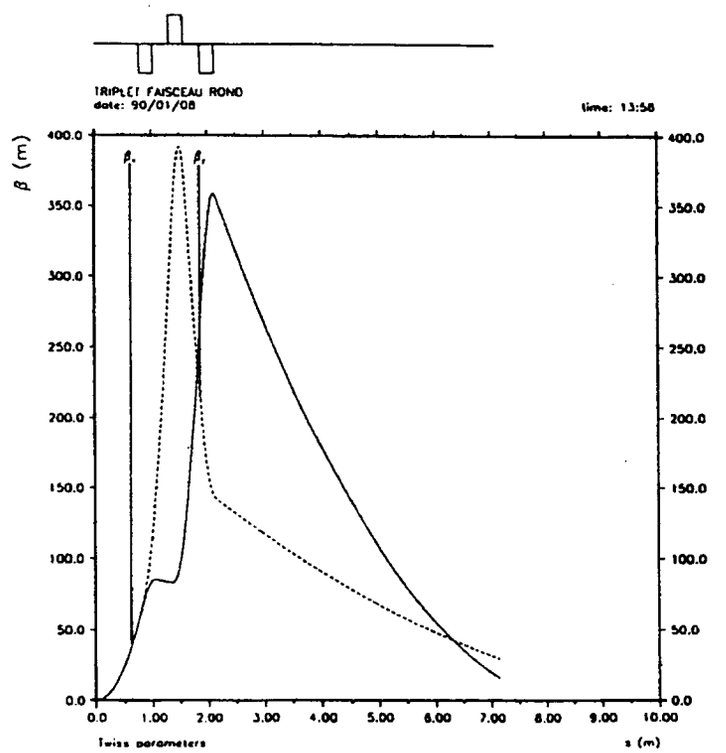
As we saw above, the maximum luminosity, limited by the beam-beam effect, can be increased by increasing the transverse emittance which in turn leads to a larger number of particles per bunch.

In an electron (or positron) machine the emittance, hence the transverse beam sizes, is the result of an equilibrium between the damping effect and the quantum fluctuations effect which are both generated by the synchrotron radiation. Consequently the bending magnets in the arcs are mostly responsible for the horizontal emittance ε_x while vertical bending (generally weak) together with transverse betatron coupling are responsible for the vertical emittance ε_y. In most practical cases $\varepsilon_x \gg \varepsilon_y$, unless the machine is operated on a coupling resonance.

Since high emittances are desirable, arcs made of FODO cells (Fig. 2-a) are suitable,



(a) Doublet scheme for flat beam



(b) Triplet scheme for round beam

Figure 1: Low β^* insertion optics

and well developed on existing machines. The natural horizontal emittance ε_x (neglecting betatron coupling) in that case is given by:

$$\varepsilon_{x_{[mrd]}} \simeq 3.8\ 10^{-13}\gamma^2 \frac{< H_x >_{magn}}{\rho_x} \tag{10}$$

where ρ_x is the bending magnet radius in the horizontal arcs and $< H_x >_{magn}$ is the average, taken in the arc bending magnets, of an optical function which can be approximated with a trigonometric function of the phase advance per FODO cell, 2ψ, and the bending angle θ [3]. Fig. 2-b corresponds to a tabulation of this function with respect to 2ψ and for different values of θ, made in the case of a 2 GeV beam. It is observed that small phase advances per cell and large bending angles (small number of cells) favours large emittances. Since in most cases the phase advance per cell is chosen to simplify the chromaticities correction, the remaining parameter, in a single machine, for adjusting the emittance is the number of elementary cells in the arcs.

However increasing the luminosity by increasing the emittance has its own limitation due to the corresponding increase in the bunch current. As a matter of fact:

- higher circulating current leads to higher beam power loss, due to synchrotron radiation, which is dissipated in the vacum chamber walls with a limit on the tolerable amount per unit length. Beyond this limit it is necessary to increase the bending magnet radius which in turn increases the circumference
- higher peak current leads to single-bunch lengthening, as well as fast transverse instabilities, due to the broad-band impedance of the vacuum pipe (including the RF cavities)
- higher average current leads to multibunch instabilities due to narrow-band impedances of the surroundings (essentially RF cavities).

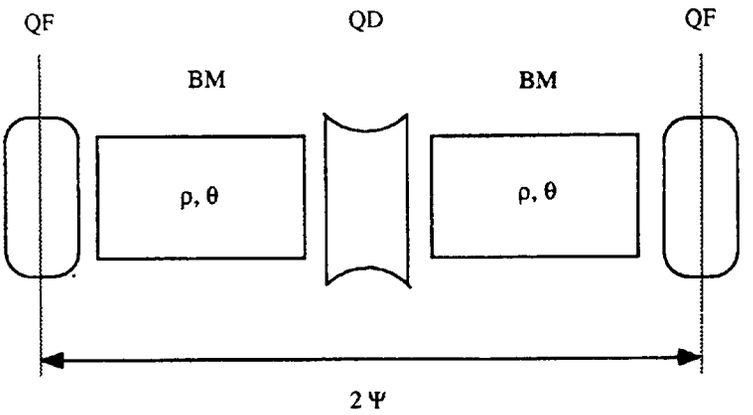
Finally a compromise is necessary between peak and average current which leads to a reasonable choice of the number of circulating bunches in each beam.

However before entering the next section, we should mention that today's sophisticated machines make use of wiggler magnets to control the emittances by fine ajustment of the synchrotron radiation power. Depending on their design and locations around the ring circumference, they can either enhance the damping of the particle motions or change the partition numbers between transverse and longitudinal motions.

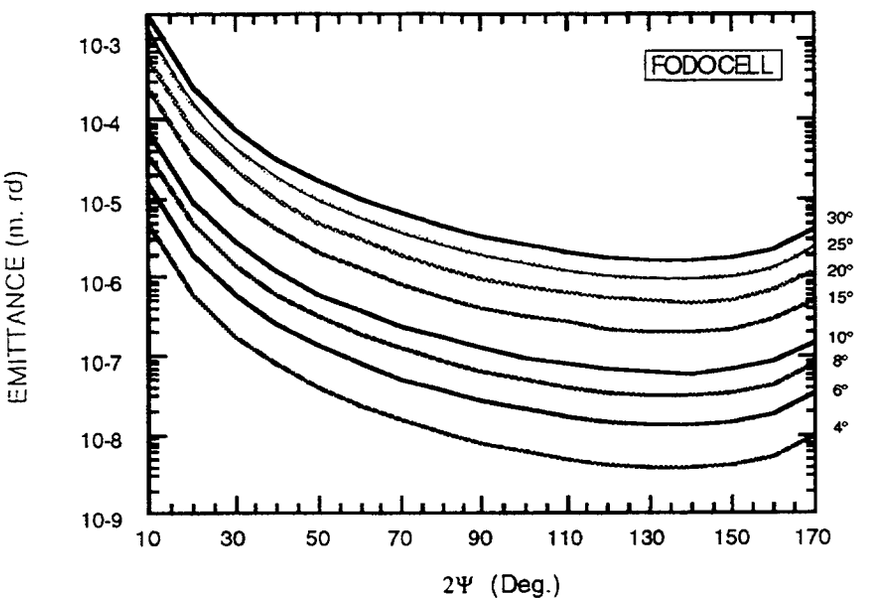
3.3 Collision frequency

When the emittance has been choosen, according to the different constraints previously mentionned together with the pratical limitation that may come from the size of the vacum chamber, the collision frequency kf_r remains the final parameter of interest for increasing the luminosity.

The maximum number of bunches that can be filled in a storage ring corresponds to the harmonic number $h = f_{RF}/f_r$. In principle that can lead to $2h$ crossing points equally spaced along the circumference. However these crossing points which do not have the right low β values, to minimize the beam-beam effect, are called parasitic crossings since they will have a bad influence on the luminosity. Hence one is left with two possibilities:



(a)



(b)

Figure 2: Horizontal emittance from FODO cells

either the unwanted crossings are suppressed by local beam separation (of the order of $\pm 10\sigma$) in one plane or another, or the two beams circulate in two different rings with common straight sections for desirable crossings with low β^*.

Beam separation has been successfully used at CESR, Cornell [4] and LEP, CERN [5] but the total number of bunches per beam still remains quite limited (a few bunches/beam) for practical reasons related to the separation scheme, here called the "Pretzel scheme" (Fig. 3).

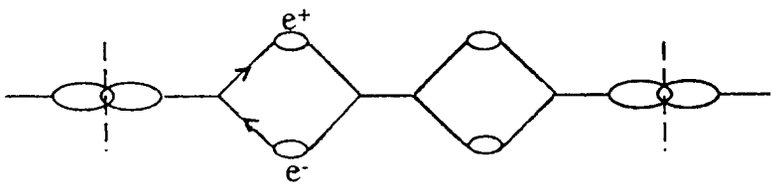


Figure 3: Pretzel scheme

Since factories aim at very high luminosities, the most reasonable approach finally consists of designing a double ring machine with one or perhaps two, common insertions for localizing the detectors. Only a few projects at low energy (PHI-factories) still use a single bunch per beam and are designed to be compact in order to increase the collision frequency either using superconducting magnets or some other novel idea.

3.4 Techniques for multibunch operation

The inverse of the collision frequency, $(kf_r)^{-1}$, represents the time distance between bunches. The shorter it is the higher the luminosity, at a given crossing, will be. However this distance is limited by the rings separation scheme under consideration. There are essentially three possibilities:

a) Electrostatic separation and head-on collisions

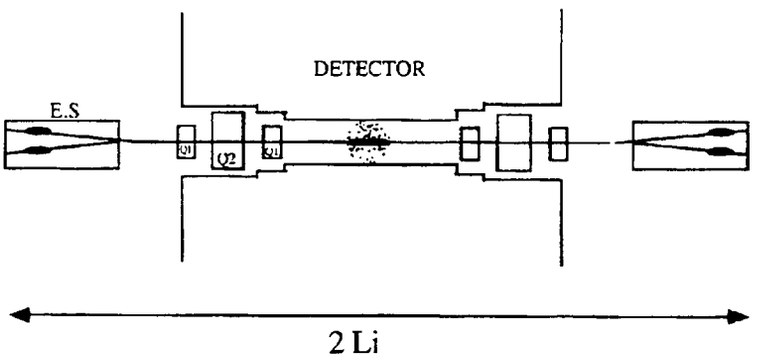


Figure 4: Head-on collisions with electrostatic beam separation

This scheme is sketched on Fig. 4. On both sides of the detector the electron and positron beams are separated naturally by means of electrostatic separators (ES) in two different rings. In practice, since ES can only provide small deflections, they are followed by independant vertical bends for the two rings, and after a certain distance

used for optical matching and total ring separation, reverse vertical bends bring the orbits back into horizontal planes again. Clearly in this scheme the distance $2L_i$ between the outputs of the two ES determines the minimum distance between bunches to avoid parasitic crossings:

$$kf_r \leq c/2L_i \tag{11}$$

where c is the particle velocity. It is assumed that beam separation at the output of the ES is large enough ($\geq 10\sigma_y$).

This scheme is mostly used in Tau-Charm Factory designs [6, 7], for which a total length $2L_i \approx 15$ m can accommodate the detector (including final focus doublet or triplet) and the electrostatic separators. Consequently collision rates up to 20 MHz can be aimed at.

b) Magnetic separation with large crossing angles

If the two beams with opposite charges and opposite velocities cross at an angle they can be further separated when entering off-center a defocusing quadrupole, symmetrically with respect to its axis. Such a double ring scheme with vertical crossing angle was used on the DORIS machine at DESY in Hamburg. This scheme in which the crossing angle was made in the plane of the beam flatness induced synchro-betatron coupling which led to enhancement of the non-linear beam-beam effect with a corresponding reduction of the luminosity.

Today results of numerous theoretical and simulation studies, predict that the situation would be much better if the large crossing angle is made in the plane where the beam size is the largest. A rough estimate of the maximum half crossing angle is given by the formula:

$$\phi_x \leq \sigma_x^*/\sigma_z \tag{12}$$

where the right hand side represents the ratio between the horizontal r.m.s beam size ($\sigma_x \gg \sigma_y$) and the r.m.s bunch length. Within this approximation crossing angles up to $\pm$ 20 mrd have been proposed and bunch distances as small as 0.6 m can be envisaged.

c) Magnetic separation with head-on collisions in asymmetric rings

Head-on collisions in the case of asymmetric double rings schemes ($E^+ \neq E^-$) are still possible with only magnetic separation. The trick consists of using a small dipole magnet right after the crossing point which provides deviation angles that depend on the beam energy (the sign of the bending radius being the same). After a certain distance the two beams are separated and hence can be further separated by proper location of the low β^* quadrupoles. This scheme can be applied either in the vertical or the horizontal plane. It is one of the essential features of B-factories. Detailed design studies have shown that distances between successive bunches as small as 1.3 m can be achieved in this way.

4 DESIGN ALTERNATIVES

4.1 Luminosity enhancement with beam monochromatization

When two beams collide, each having an r.m.s relative energy spread σ_ε, the absolute r.m.s energy dispersion in the C.M. is:

$$\sigma_W = \sqrt{2}\ \sigma_\varepsilon\ E_o \tag{13}$$

where E_o is the beam energy.

For the study of narrow resonance processes this is equivalent to a reduction of the effective luminosity. Beam monochromatization consists of reducing this quantity σ_W by creating a non-zero dispersion function D^* at the interaction point [8], with opposite polarities for the two beams, in one transverse plane or the other. The corresponding collision scheme is shown on Fig. 5 for the case of a vertical dispersion:

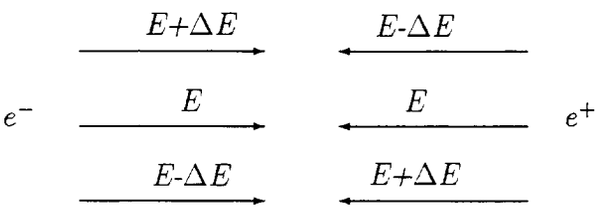


Figure 5: Monochromatization scheme

$$(D_y^*)^+ = -(D_y^*)^- \tag{14}$$

with the corresponding chromatic orbit displacement:

$$\Delta y = D_y^* \varepsilon \tag{15}$$

where $\varepsilon = \Delta E/E$ is the relative energy deviation of a particle. Hence an electron with an energy deviation $+\ \Delta E$ will meet a positron beam with an energy deviation $-\Delta E$. To first order, neglecting the vertical betatron beam size $\sigma_{y\beta}^*$, the absolute C.M. energy becomes:

$$W = 2E + O(\varepsilon^2) \tag{16}$$

Taking into account the finite vertical betatron beam size it can be shown that the r.m.s. absolute energy spread in the C.M. is expressed as [9]:

$$\Sigma_W = \sigma_W/\lambda \tag{17}$$

where:

$$\lambda = D_y^* \sigma_\varepsilon \Big/ \sigma_{y\beta}^* \tag{18}$$

is called the "gain factor".

Unfortunately, due to the fact that the total vertical beam size is now dominated by the dispersion effect:

$$\sigma_y^* = [\sigma_{y\beta}^{*2} + (D_y^* \sigma_\varepsilon)^2]^{1/2} \approx D_y^* \sigma_\varepsilon \tag{19}$$

this "gain factor" for the C.M. energy spread, becomes a "loss factor" for the effective luminosity:

$$L = L_o/\lambda \tag{20}$$

where L_o is the luminosity without monochromatization.

At this stage it is interesting to show that there is an alternative which allows one to recover the luminosity while keeping the benefit of a reduced C.M. energy spread.

Remembering that:

$$L \propto \left(\frac{\xi_x}{\beta_x^*} + \frac{\xi_y}{\beta_y^*} \right) \tag{21}$$

it is clear that the dramatic increase in vertical beam size due to the dispersion at the crossing point lowers ξ_y while only slightly affecting ξ_x. Consequently the first term in the bracket is no longer negligible and it is possible to increase its contribution by decreasing the value of β_x^*. Since it is hard to make both β^* very small, one can just produce a new optics configuration with $\beta_y^* \gg \beta_x^*$ by reversing the polarities of the insertion quadrupoles. The first term in the bracket has become the leading one and ξ_x can be adjusted to ξ_{max} for maximum luminosity, if σ_x^* is made small enough to follow the reduction of β_x^*. The later means that the new configuration needs a small horizontal emittance, as for instance in synchrotron radiation storage rings. The price to pay with small horizontal emittances, and still high bunch current, is a reduction of the beam lifetime due to intra-beam scattering.

4.2 Crab-crossing scheme

It is believed that crossing at an angle reduces the ξ_{max}. This was experienced first on the DORIS machine at DESY where the two beams (e^+ and e^-), coming from two different rings atop each other, were crossing at an angle. A systematic experiment was made, much later, on CESR, Cornell to measure the ξ_{max} value as a function of the crossing angle [10]. The results are shown on Fig. 6.

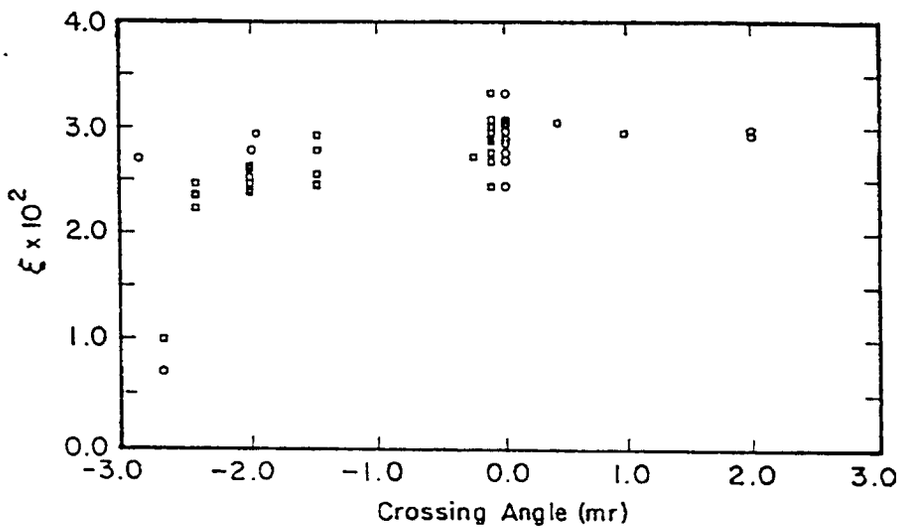


Figure 6: ξ_{max} versus crossing angle at CESR

From the analysis it seems that the effect of the angle ϕ becomes significant when:

$$\phi_{x,y} \geq \sigma_{x,y}^* / \sigma_z \tag{22}$$

Theoretical work as well as simulations indicate that crossing at an angle made the non-linear beam-beam effect worse due to coupling between synchrotron (longitudinal) motion and betatron (transverse) motion which increases the density of non-linear resonances in the tune diagram.

To overcome this synchro-betatron coupling it has been proposed to rotate the beams prior to collision such that they will perform head-on collisions in the C.M. frame [11, 12]. The beam can be tilted conveniently using RF deflecting cavities located symmetrically with respect to the crossing point (Fig. 7). A large crossing angle therefore makes it pos-

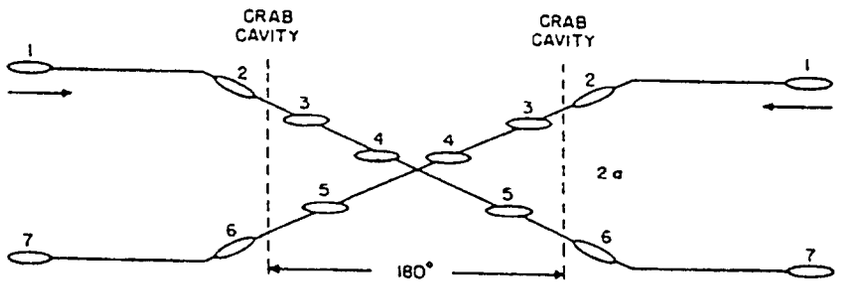


Figure 7: Crab-crossing scheme

sible to reduce the distance between circulating bunches while storing more bunches to produce higher luminosities. Notice that if crossing at an angle is made in the plane corresponding to the largest transverse beam size, and if the latter is large enough compared to the bunch length, crab-crossing is not necessary in principle.

5 PLANNED e^+e^- FACTORIES

5.1 Overview

Many B-factory projects have been worked out throughout the world and most of them have led to detailed technical proposals. The most advanced ones, considered to be ready for construction if funding is available, are [13, 14, 15]:

- TRISTAN II at KEK, Tsukuba, Japan
- PEP II at SLAC, Stanford, USA (collaboration SLAC/LBL/LLNL)
- CESR-B at Cornell, Ithaca, USA

All these projects have in common the use of existing sites and existing facilities and they are well advanced in terms of R&D. Then, comes the B-factory project in Novosibirsk as a piece of a future large accelerator complex called VEPP 5 that also includes a PHI factory and corresponding injectors [16].

Other proposals exist, such as the PSI project in Switzerland [17], the B-factory at CERN in the ISR tunnel [18] and the DESY B-factory [19] but they all seem to have been withdrawn from the world competition ; hence they will be given no more attention in the following.

As for Tau-Charm factories, the remaining candidates are:

- The DUBNA project in Russia [20], often presented as a part of a large accelerator complex that also includes synchrotron radiation, heavy ion and high resolution neutron source facilities.

- The European project, by a collaboration of different institutes, set up after the Spanish project [21] was withdrawn in late 1992.

PHI factories, at a more modest C.M. energy, are certainly the most advanced ones in terms of construction:

- DAFNE [22] at Frascati, Italy, is now half way to completion.
- Novosibirsk project [23] has started civil engineering as well as the construction of the injector linac.

Other proposals exist such as the UCLA [24] and KEK [25] projects.

5.2 Main characteristics of e^+e^- factories

The term factory is synonymous with very high luminosity colliders. It turns out that increasing the luminosity requires storing of high circulating current. This leads to a multibunch mode of operation since the bunch current is limited by the beam-beam interaction itself. However, many bunches in each beam, in a single ring, would in principle give additional crossing points around the circumference with a negative effect on the performance, unless electrostatic beam separation is arranged at these unwanted crossings. Such a scheme has been already successfully used on CESR, at Cornell with a few bunches in each beam, leading to a world record luminosity ($2.5\ 10^{32}\ cm^{-2}\ s^{-1}$). With a much larger number of bunches, the scheme becomes complicated and the next obvious solution consists of two separate rings with a common insertion region for desirable e^+e^- collisions.

Finally, all Beauty and Tau-Charm factories under consideration have been designed with two rings, either with one ring aside or atop the other one, in the same tunnel, with the same circumference and a single crossing point. The aim of these factories, compared to the present machines [26], is to gain about two orders of magnitude on the luminosity in the long term and at least one order of magnitude in the very short term. For B-factories the reference, as already mentioned, is the present CESR situation with $2.5\ 10^{32}\ cm^{-2}\ s^{-1}$, while for Tau-Charm factories the reference is the BEPC machine in Beijing with $10^{31}\ cm^{-2}\ s^{-1}$, both machines being single-ring machines and the latter operating in the single-bunch mode.

Double-ring machines, such as DCI in Orsay with a single bunch and four round beams colliding head-on for space charge compensation and DORIS in Hamburg with many flat bunches and a vertical crossing angle, have been built in the past without reaching their expected performances. However, taking into account the experience gained from them and the many single-ring machines operating throughout the world (beam-beam effect, single-bunch and multibunch effects, beam-gas interaction and intrabeam scattering) give us confidence that the new factory designs are safe extrapolations of well-known accelerator techniques and technologies. However, compared to existing machines, Beauty and Tau-Charm factories do present their own difficulties.

As for B-factories the electron and positron energies are different although the ring circumferences are the same ; hence B-factories enter into a new class of asymmetric colliders of which HERA (e-p collider) in Hamburg is the first one [27].

In the case of Tau-Charm factories, their design includes the option of beam monochromatization at the interaction point, which results in a much lower C.M. energy spread compared to the natural beam energy spread. However in most cases a back-up solution with a standard mode of operation is considered for practical reasons (better beam lifetime) as well as for safety reasons since a monochromatization scheme has never been tried experimentally [28, 29].

The design features of PHI factories differ quite a lot from one project to another. For instance, DAFNE and KEK projects use two rings with a large horizontal crossing angle and many bunches. The Novosibirsk project uses single bunch per beam and head-on collisions, in a single compact ring with a very unconventional geometry (two arcs separated with a single straight section) ; the high luminosity is obtained by having very small round beam spot sizes and equal mini β^* in both planes. The UCLA project is also a single ring with single bunch per beam and head-on collisions ; low β_y^* together with high collision frequency (compact ring with small circumference) is the key to high luminosity.

5.3 Beauty factory projects

They are all asymmetric ($E^- \neq E^+$) machines and the high energy ring is the electron one for the obvious reason of positron production cost.

Table 1
B-factories main parameters

	B-factory Cornell CESR B		B-factory SLAC/LBL/LLNL PEP II		B-factory KEK TRISTAN II		B-factory Novosibirsk	
	e^+	e^-	e^+	e^-	e^+	e^-	e^+	e^-
Energy/beam [GeV]	3.5	8	3.1	9	3.5	8	4	7
Luminosity [cm^{-2} s^{-1}]	$3.0\ 10^{33}$		$3.0\ 10^{33}$		$2.0\ 10^{33}$		$1.0\ 10^{34}$	
Circumference [m]	764		2200		3018		765	
N of bunch/beam	164		1658		1006		182	
Particules/bunch	$1.9\ 10^{11}$	$0.8\ 10^{11}$	$5.9\ 10^{10}$	$4.1\ 10^{10}$	$3.3\ 10^{10}$	$1.4\ 10^{10}$	$1.9\ 10^{11}$	$1.1\ 10^{11}$
Beam current [A]	2	0.9	2.14	1.48	0.52	0.22	2.1	1.2
Crossing angle	YES		NO		YES(small)		NO	
Crab crossing	YES		NO		NO		NO	
Monochromatization	NO		NO		NO		YES	
Beam lifetime [min]	-		1700		-		-	
ϵ_x/ϵ_y [nm rad]	135/3.75		97/3.9	48/1.9	19/0.19		4/0.92	5.8/0.92
ξ_x/ξ_y	0.04/0.03		0.03		0.05		0.017/0.05	
β_x^*/β_y^* [cm]	100/1.5		37.5/1.5	75/3.0	100/1		40/0.9	
D_x^*/D_y^* [cm]	-		-		-		-38/0	38/0
f_{RF} [MHz]	500		476		508		500	
V_{RF} [MV]	11.6	32.7	9.5	18.5	20	47	8.6	15.4
U_0 [MeV/turn]	0.7	4.8	1.2	3.6	0.9	4.1	0.6	3.2
σ_b [mm]	10		10		5		7.0	

All B-factories have been designed to operate with flat beams ($\sigma_y^* \ll \sigma_x^*, \beta_y^* \ll \beta_x^*$). A β_y^* of the order of 1 to 3 cm is considered to be a safe extrapolation of present achievements, allowing the first insertion quadrupole to stay at a reasonable minimum distance from the collision point. It is worth mentioning that an increase of the luminosity follows a reduction of β_y^* provided the r.m.s. bunch length σ_z is made such that $\sigma_z < \beta_y^*$; this will have some consequenses on the RF system.

Table 1 gives the design parameters of the B-factory projects and shows the typical kinds of flexibility one has in choosing the parameters to achieve the desirable luminosity.

In the case of CESR-B and TRISTAN II, the rings are separated horizontally in the tunnel and the beams after collision are also separated in the horizontal plane. In the case of Novosibirsk the beams are separated in the vertical plane after collision and one ring is atop the other. In the particular case of PEP II, the rings are separated vertically but the beams after collision are initially separated horizontally ; this scheme has a more complicated insertion but it does not cause fundamental problems.

The large differences in the machine circumferences are due to designing them to fit into existing tunnels, apart from the Novosibirsk case which has chosen the minimum size. Table 1 shows that between the two projects with large circumferences (PEP II and TRISTAN II) there are still some design differences:

- Due to high confidence in its powerfull injector, PEP II uses more current for a given luminosity and relaxes on β_y^* and ξ values compared to TRISTAN II. This fact also explains the difference in the emittances.

- PEP II uses head-on collisions, which is the most conventional scheme, and a bunch spacing of 1.3 m, while TRISTAN II uses a small crossing angle ($\pm$ 2.8 mrd) to enhance beam separation although it has a larger bunch spacing of 3 m and a smaller horizontal emittance. This partly looks like a different appreciation concerning the long range beam-beam effect at the first parasitic crossing following the interaction point. In fact, a crossing angle as small as $\pm$ 2.8 mrd is considered to have no particular effect on the beam-beam limit according to experimental evidence on CESR. Notice that the TRISTAN II design could come back to head-on collisions if necessary, but is proposed to be upgraded in the future (phase 2) with a large crossing angle ($\pm$ 10-20 mrd) and smaller bunch spacing (0.6 cm) leading to an increase of the luminosity by a factor 5.

- The two projects benefit from the beam energy asymmetry to initiate the beam separation after the collision point, using a common magnetic dipole. After passing this dipole the beams go off-center in the insertion quadrupoles and are deviated further away. In the case of TRISTAN II all these magnets, located inside the detector, are superconducting whereas they are permanent magnets in the case of PEP II. Since the PEP II detector is also warm, there is no need for cryogenic services in this area in their case.

- TRISTAN II has a smaller β_y^* value compared to PEP II. This partly explains its need for stronger focusing through superconducting insertion quadrupoles. It also explains the requirement of a smaller bunch length which indeed leads to a much higher RF voltage.

Considering now the two B-factory projects with much smaller circumferences, it is seen from Table 1 that, apart from their higher number of particles per bunch, they do show some intrinsic design features which can hardly be considered conventional:

- CESR-B uses a large horizontal crossing angle to separate the beams, which in principle is the most efficient separation scheme and which can be initiated from

magnets located outside the detector. In fact due to specific environment and detector constraints there is apparently no other choice. The drawback of this scheme is that the beam-beam interaction will induce strong synchro-betatron resonances and consequently reduce the maximum tolerable value of the space-charge parameter, ξ_{max}. To overcome this bad effect CESR-B proposes to use the "crab scheme". Theoretical studies are being carried out since no experiment can be made on existing machines. If successful the same scheme can be adopted to upgrade the luminosity of TRISTAN II (phase 2).

- The Novosibirsk project uses a monochromatization scheme which implies having a non-zero dispersion function at the crossing point and small emittances to maintain high luminosity.

5.4 Tau-Charm factory projects

They are also designed with two rings but are symmetric ($E^+ = E^-$). Hence the magnetic separation is no longer usable with head-on collisions and needs to be replaced by either electrostatic separation (ES) or a large crossing angle. Up to now the Tau-Charm factories have been essentially considered with electrostatic beam separation and head-on collisions since the initial requirement on the luminosity could be fulfilled this way. More recent pressure from the user community to increase the luminosity by another factor 3 to 4 would probably lead the designers to consider a crossing angle scheme as a future upgrade. But since no detailed work has been yet made in this direction it will be ignored in the following. On the other hand much emphasis has been placed on beam monochromatization to reduce the C.M. energy spread during collision. Some initial feasibility studies preferred this mode of operation while others remained more conventional. Finally, the Tau-Charm factory projects which today seem to be still in competition, the Dubna project for the Russian side and the so-called European collaboration (former Spanish project), have been considering versatile types of design which could switch from one mode of operation (monochromatization) to the other one (standard) by simple quadrupole strength settings. Although there is no demand from the potential users, a scheme with polarized beams has also been worked out [30].

Table 2 summarizes the main parameters of the Dubna and European projects and does not show any strong differences between them. In fact, this is not very surprising since strong collaboration exists between designers to converge towards the best approach that could fulfill the phase 1 performances (10^{33} cm^{-2} s^{-1}). The third column in Table 2 is a design example with a monochromatization mode of operation that can be switched back to a standard one. The difference in the circumferences here is not fundamental since it can be readjusted ; this in fact reflects that designs are still being worked out and they are not, as for B-factories, in the stage of detailed engineering proposals.

Much of the design philosophy for Tau-Charm factories follows that of Beauty factories. This is true for instance for the β^* values, the space-charge parameters and the short bunch length. However Tau-Charm factories have their own peculiarities which essentially come from the separation scheme and the monochromatization scheme.

Table 2
Tau-Charm factories main parameters

	Europe	Dubna	Monochrom.
Energy/beam at L_{max} [GeV]	2.0	2.2	2.0
Luminosity [cm^{-2} s^{-1}]	1.0 10^{33}	1.1 10^{33}	1.0 10^{33}
Circumference [m]	360	378	322
Bunches/beam	30	30	30
Particules/bunch	1.5 10^{11}	1.6 10^{11}	1.3 10^{11}
Beam current [A]	0.600	0.614	0.585
Crossing angle	NO	NO	NO
Crab crossing	NO	NO	NO
Monochromatization	NO	NO	YES
Beam lifetime [min]	290	210	120
ϵ_x/ϵ_y [nm rad]	400/20	482/-	19/1.2
ξ_x/ξ_y	0.04	0.035	0.04/0.035
β_x^*/β_y^* [cm]	20/1	-/1	1/15
f_{RF} [MHz]	500	476	500
V_{RF} [MV]	19	18	-
U_0 [MeV/turn]	0.14+wigglers	0.2	0.14
σ_b [mm]	10	7.5	-

5.5 PHI factory projects

PHI factories cover C.M. energies of about 1 GeV with luminosities in the range 10^{32} to 10^{33} cm^{-2} s^{-1}. The reference here is the old VEPP-2M storage ring operating since 1975 at the Budker Institute in Novosibirsk and which has reached a luminosity as high as 4.3 10^{30} cm^{-2} s^{-1}. So here again two orders of magnitude, at least, are aimed at with factories.

PHI factories, compared to higher energy factories, show quite a variety of design approaches as can be seen in Table 3.

The DAFNE project, half way to completion, and the KEK proposal enter in the class of "almost conventional technology" machines. Their design philosophy is based on multibunch beams circulating in two separate rings. However, the relatively small circumference makes it necessary to collide beams at a large angle in order to have many bunches without parasitic crossings. The choice of a horizontal crossing angle (in the plane where the beam size is large) is an alternative to escape from the DORIS problems (beam-beam induced synchro-betatron resonances with vertical crossing angle and flat beams). Small circumferences together with high collision frequencies lead to quite large, unusual circulating currents; from Table 3, in the case of DAFNE it can be seen that one can extract a current of 5.3 A per beam.

Table 3
PHI-factories main parameters

	Novosibirsk Phase I	UCLA Phase I	KEK	DAFNE
Number of rings	1	1	2	2
Number of crossings	1	1	1	2
Crossing type	Head-on	Head-on	Horizontal	Horizontal
Crossing angle [mrad]	0.0	0.0	20.0	$\pm$12.5
Energy/beam at L_{max} [MeV]	510.0	510.0	510.0	510.0
Luminosity [cm^{-2} s^{-1}]	>1.0 10^{33}	2.0 10^{32}	3.0 10^{33}	5.0 10^{32}
Circumference [m]	35.1	17.4	126.0	97.7
Bunches/beam	1	1	300	120
Particles/bunch	2.0 10^{11}	4.0 10^{11}	7.5 10^{10}	9.0 10^{10}
Crossing frequency [MHz]	17.1	17.2	714	368
ξ_x/ξ_y	>0.1	0.05	0.03	0.04
β_x^*/β_y^* [cm]	1.0/1.0	19.0/3.9	100.0/1.0	450.0/4.5
Coupling factor	1.0	0.2	0.01	0.01
σ_x^*/σ_y^* [mm]	0.07/0.07	0.71/0.14	1.1/0.011	2.1/0.020

Part of the difference in luminosity between DAFNE and KEK lies in the design value of β^*. Here DAFNE is not pushing too far and having to use the existing ADONE's hall may set a physical limit in the optimization of the insertion.

The other two projects, Novosibirsk and UCLA, differ drastically from the previous ones in the sense that they keep using a single bunch per beam in a single ring and increase the collision frequency by making the circumference very small. The UCLA compact ring is made of superconducting high field dipoles, while the Novosibirsk design lay-out looks like a "thigh bone" with a single straigth section and a compact arc at each extremity.

Boths projects, by suitable design values for the natural emittances, the coupling factor, and the β values at the interaction point, end with more particles per bunch. Let's emphasize here the originality of the Novosibirsk case which has choosen quite a challenging way of further increasing the luminosity by having equal mini-β values at the IP, equal transverse emittances hence round colliding beams, and finally hoping that such conditions will permit to reach higher values for the beam-beam parameter ξ according to simulation studies.

5.6 Specific R&D for factories

Although design approaches for factories are considered to be conventional, hence safe, there are still technological challenges to be met in connection with the necessary multibunch high current. Solutions for all these challenges exist, so let's review the main ones.

Interaction region

This is complicated by the fact that with two rings it is necessary to both focus and separate the beams. Solutions with or without superconducting magnets exist, and in the case of Tau-Charm factories the electrostatic separation uses similar technology to that on CERN machines. So finally the remaining challenge here is a proper integration of the insertion components with the detector in order to minimize the background, in particular that from the components which separate the beams.

The micro-beta scheme can be obtained in all cases without dramatic consequences on the dynamic aperture since chromaticities remain at a reasonable level and their correction by sextupoles uses quite conventional schemes. The constraint is caused by the fact that the first quadrupole needs to be located close enough to the collision point in order to limit the maximum value of the β function which is responsible for a large fraction of the ring chromaticity.

RF system

The RF system design is dominated by:

a) the high circulating current which gives high beam power losses due to synchrotron radiation,

b) the short bunches which require high cavity voltage,

c) many bunches with large peak and average currents which require very low impedances for higher-order modes (HOM) in cavities to avoid coupled-bunch instabilities.

The first two constraints look more like a cost problem. In principle both high power and high voltage favour superconducting cavities for which HOM dampers have already been developed by necessity. On the other hand superconducting cavities are still expensive and temperamental with high current electron beams.

In the case of CESR-B, where a long tradition of superconducting cavity development exists, single-cell superconducting cavities have been chosen and prototypes are being developed with emphasis on geometry and couplers [31] since accelerating voltage is not the real worry.

Considering that superconducting RF requires particular technological expertise and appropriate cleaning and cryogenic installations, the choice for PEP II has been normal conducting cavities. SLAC together with LBL and LLNL has launched an extensive R&D program on single-mode warm cavities [32]. A low power cavity with mode dampers has been successfully tested and now extensive electromagnetic and thermal studies are being made to produce a high power test cavity.

At KEK, for TRISTAN II the final choice is not yet obvious and R&D is going on, on both superconducting and normal conducting cavities [33, 34].

Although the constraints are less severe in the case of Tau-Charm factories, these projects at the moment seem to favour superconducting RF, following developments going on at CERN and DESY.

Most of the factory projects have chosen a frequency around 500 MHz. Although 500 kW klystrons exist in this frequency range, SLAC is launching a 1.2 MW prototype klystron.

In the particular case of CESR-B, a crab-cavity is also under development [35].

Vacuum system

A good vacuum pressure will guarantee long beam lifetime, less background in the detector and should reduce the ion trapping phenomena in the electron ring.

Factories with large circulating currents produce a large amount of synchrotron radiation power per unit length at specific locations around the circumference. Clearly the B-factories are much worse than Tau-Charm factories in this respect, and the use of high field wigglers to control the emittance makes them even more challenging. For instance the maximum power density from synchrotron radiation on TRISTAN II is about 33 kW/m in wiggler sections which is almost an order of magnitude above a standard situation. This power leads to large thermal stresses and requires particular attention to cooling. In addition the radiated photons will produce photo-electrons at a high rate which in turn extract molecules so increasing the pressure. Hence high pumping speed is necessary, but the lower the photo-desorption coefficient, the better the situation will be.

For all these reasons the choice of copper vacuum chambers has been validated in the case of B-factories. This choice could as well be extended to Tau-Charm factories, although in principle aluminium technology can still be used here. Copper vacuum chambers have already been used, on DCI in Orsay over a small fraction of the ring circumference, and on the HERA electron ring in Hamburg as the major part of the vacuum chamber. Though in principle copper vacuum systems can be regarded as a known technology, R&D work on prototypes is still necessary to establish the best procedures for optimizing chamber fabrication and efficient cooling within reasonable cost and according to specific B-factories performance level [36, 37, 38]. It is interesting to note also that the self-shielding property of copper would make lead shielding unnecessary.

Feed-back

With multibunch beams, coupled-bunch instabilities are normally expected to take place. Growth rates however are limited by mode damping in cavities and the use of a feed-back system is then possible. The worst coupled-bunch instabilities take place in the longitudinal space. Feed-back systems for multibunch operation have already been intensively developed at CERN, DESY, KEK and SLAC but factories need such a high current with so many bunches that it remains very important to continue developing more efficient feed-back systems. Here, existing synchrotron radiation storage rings with similar requirements are appropriate test facilities for feed-back developments. As an example, let's mention the experiment started on SPEAR/SSRL [39] of a bunch-by-bunch longitudinal feed-back system for PEP II which will soon be installed on the Advanced Light Source at LBL.

6 CONCLUSIONS

B-factories cannot be considered as accelerator physics ground breakers. However, they still need to be considered as engineering challenges in many respects.

In the case of Tau-Charm factories, the monochromatization scheme is an accelerator physics challenge; however, most designers are carefull enough to include a back-up

conventional mode of operation.

Finally, PHI factories are the most advanced ones in term of construction, especially DAFNE which in a couple of years should start answering many questions such as horizontal crossing angle and high multibunch current effects.

Although e^+e^- factories appear feasible, successful fulfilment of performances is an obligation which requires carefull choices.

REFERENCES

[1] P. Beloshitsky, Notes to a tunable lattice for realization of flat beam and monochromatization schemes of Tau-Charm factory, Preprint JINR E9-92-187, Dubna, 1992

[2] S. Krishnagopal and R. Siemann, Proc. 1989, IEEE Particle Accelerator Conference, 836 (1989)

[3] R.H. Helm and H. Wiedemann, Emittance in a FODO-cell lattice, PEP note 303, SLAC, Stanford, May 1979

[4] D. Rice, High luminosity operation of the Cornell Electron Storage Ring, Proc. 1989 IEEE Particle Accelerator Conference, March 20-23, 1989 Chicago, IL, p.444

[5] R. Bailey et al., Commissionning and operation of the LEP Pretzel Scheme, Proc. 1993 Particle Accelerator Conference, May 17-20, Washington, D.C. , p. 2013

[6] J.M. Jowett, Initial design of a Tau-Charm factory at CERN, CERN LEP-TH/87-56 (December 1987)

[7] J. Gonichon et al., Preliminary study of a high luminosity e^+e^- storage ring at a C.M. energy of 5 GeV, LAL/RT 90-02, Orsay (January 1990)

[8] M. Basseti, ADONE Internal Report E-18, Frascati (1974)

[9] A. Renieri, Possibility of achieving very high energy resolution in electron-positron storage rings, Preprint INF/756 (R), Frascati (1975)

[10] CESR Accelerator Group, Proceedings of B-factories: the state of the art in accelerators, detectors and physics, SLAC, April 6-10, 1992 - SLAC-400, p283

[11] R. Palmer, SLAC-PUB 4707 (1988)

[12] K. Oide and K. Yokoya, Phys. Rev. **A40**, (1989) p315

[13] S. Kurokawa, K. Satoh, E. Kikutani (Editors), Accelerator design of the KEK B-factory, KEK Report 90-24 (March 1991)

[14] An asymmetric B-factory based on PEP, Conceptual Design Report SLAC-372, LBL-PUB 5303 (1991)

[15] K. Berkelman et al. (CESR), Conceptual design for a B-factory based on CESR, CLNS 91-1050, Cornell University, Ithaca, NY 14853 (1991)

[16] N.S. Dikanski et al., Novosibirsk B-factory: status and perspectives, Proc. XVth International Conference on High Energy Accelerators, Hamburg, Germany July 20-24, 1992, p452, World Scientific

[17] R. Eichler et al., Proposal for an electron-positron collider for heavy flavour particle physics and synchrotron radiation, PSI PR-88-09 (1988)

[18] T. Nakada (editor), Feasibility study for a B-meson factory in the CERN ISR tunnel, CERN 90-02, PSI PR-90-08 (1990)

[19] HELENA, A beauty factory at Hamburg, Preprint DESY 92-41

[20] E.A. Perelstein et al., JINR Tau-Charm factory study, Proc. XVth International Conference on High Energy Accelerators, Hamburg, Germany, July 20-24, 1992, p448, World Scientific

[21] Y. Baconnier et al., A Tau-Charm factory laboratory in Spain combined with a synchrotron light source (a conceptual study), CERN/AC/90-07 (November 1990)

[22] The DAFNE project team - Overview of DAFNE, the Frascati Φ Factory, Proc. XVth International Conference on High Energy Accelerators, Hamburg, Germany, July 20-24, 1992, World Scientific Vol.1 p.115

[23] L.M. Barkov et al., Status of the Novosibirsk PHI factory project - 1991 IEEE Particle Accelerator Conference, May 6-9, 1991 San Francisco, Cal. Vol.1 p.183

[24] C. Pellegrini et al., A high luminosity superconducting mini collider for PHI meson production and particle beam physics, 1991 IEEE Particle Accelerator Conference, May 6-9, 1991 San Francisco, Cal. Vol.5 p.2856

[25] K. Hirata and K. Ohmi, Feasibility of a Φ factory in KEK, 1991 IEEE Particle Accelerator Conference, May 6-9, 1991 San Francisco, Cal. Vol.5 p.2847

[26] Review of particle properties, Phys. Rev. **D45**, Part 2 (1992)

[27] F. Willeke, Status of HERA, Proc. XVth International Conference on High Energy Accelerators, Hamburg, Germany July 20-24, 1992, World Scientific

[28] P. Beloshitsky, A magnet lattice for a Tau-Charm factory suitable for both standard scheme and monochromatization scheme, LAL/RT 92-09, Orsay (June 1992)

[29] A. Faus-Golfe and J. Le Duff, A versatile lattice for a Tau-Charm factory that includes a monochromatization scheme (low emittance) and a standard scheme (high emittance), LAL/RT 93-07, Orsay (June 1993). Also in the proceedings of the 1993 Particle Accelerator Conference, Washington, May 17-20, 1993 (to be published)

[30] A. Zholents, Polarized J/Φ mesons at a Tau-Charm factory with monochromatization scheme, CERN/SL/92-27 (AP) (June 1992)

[31] D. Moffat et al., Design, fabrication and testing of a ferrite-lined HOM load for CESR-B, Proc. 1993 Particle Accelerator Conference, Washington, May 17-20, 1993 (to be published). Also as internal report CLNS 93/1233, Newman Laboratory, Cornell University, Ithaca.

[32] R.A. Rimmer, RF cavity development for PEP II B-factory, PEP II/ME Note 2-93, SLAC, Stanford (January 1993)

[33] M. Suetake et al., Damped RF cavity in KEK, Proc. Workshop on B-factories: the state of the art in accelerators, detectors and physics, SLAC, April 6-10, 1992, SLAC-400 (November 1992) p189

[34] T. Shintake, Jpn. J. Appl. Phys. 31 (1992) p1567

[35] K. Akai et al., Crab cavity for the B-factories, Proc. XVth International Conference on High Energy Accelerators, Hamburg, Germany, July 20-24, 1992, World Scientific

[36] P.C. Marin et al., Extended study of photon stimulated gas desorption from OFHC copper by 3.75 keV critical energy photons, Proc. EPAC 92, Berlin, 24-28 March 1992, p132

[37] C. Foerster et al., Desorption measurements of copper and copper alloys for PEP II, Proc. XIIth International Vacuum Congress, VIIIth International Conference on Solid Surfaces, October 12-16, 1992, The Hague, Netherlands

[38] C. Perkins, Research and development for the PEP II vacuum system, SLAC-PUB 6039

[39] G. Oxoby et al., Bunch by bunch feed-back for PEP II, SLAC-PUB 6035 (January 1993)

BEAM-BEAM DYNAMICS

Eberhard Keil
CERN, CH-1211 Geneva 23

Abstract

This lecture discusses the dynamics of the beam-beam interaction in storage rings. The force which a bunch exerts on particles in another bunch, circulating in the opposite direction is derived from a beam-beam potential, and used to compute the incoherent beam-beam kick, the beam-beam tune shift and the beam-beam limit. The same potential is also used to compute the coherent beam-beam kick, tuneshift and limit, describing the action on the whole bunch circulating in the opposite direction. The variation of the beam-beam tune shift with the betatron amplitude, and the resulting tune spreads in the beams and footprints are also discussed. The lecture concludes with a discussion of the nonlinear resonances driven by the beam-beam effect.

1 INTRODUCTION

Beam-beam effects are one of the most important phenomena which limit the luminosity in a storage ring which for two equal beams is given by:

$$L = \frac{N^2 f k}{4\pi\sigma_x\sigma_y} \tag{1}$$

Here, N is the bunch population. f is the revolution frequency, and k is the number of bunches in one beam. The co-ordinate axes are labelled x for the radially outwards direction, y for the direction perpendicular to the median plane, and s along the equilibrium orbit; z may be either x or y. The horizontal and vertical rms beam radii at the collision points are called σ_x and σ_y, respectively. The bunch length will be called σ_s. For good luminosity L we want large values of N, f and k, and small values of σ_x and σ_y. Small values of σ_x and σ_y are achieved in low-β insertions. Their discussion is outside the scope of this lecture. It will become clear later on that beam-beam dynamics causes an upper limit on $N/\sigma_x\sigma_y$, and therefore needs to be studied.

1.1 Charge distributions

The most general charge distribution which will be considered is a Gaussian distribution in three dimensions (3D), given by:

$$\rho(x,y,s) = \frac{Ne}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_s}\exp\left(-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}-\frac{s^2}{2\sigma_s^2}\right) \tag{2}$$

We will often do calculations for cases in which the bunch length σ_s is much larger than the transverse radii σ_x and σ_y. The appropriate charge distribution in this case is a Gaussian distribution in two dimensions (2D) with line number density n:

$$\rho(x,y) = \frac{ne}{2\pi\sigma_x\sigma_y}G(x,y;\sigma_x,\sigma_y) = \frac{ne}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}\right) \tag{3}$$

1.2 Simplifying assumptions

In most of this paper, I will make the following simplifying assumptions:

1. The beam-beam collisions are head-on, i.e. the two colliding bunches follow the same trajectory in opposite direction, without offset and without crossing angle.

2. The bunches are long with $\sigma_s \gg \sigma_z$. Hence, I can do the electrodynamics in the two transverse dimensions only.

3. The orbit functions β_x and β_y are constant through beam-beam collisions, or at least $\beta_x, \beta_y \gg \sigma_s$. Hence, I can neglect their variation along the beam-beam collisions, which was called the "hourglass effect" [1], although the consequences, a reduction of the luminosity [2], and an increase in the beam-beam tune shift [3], were published much earlier.

4. When the mathematics gets too heavy and tables of integrals would be needed, I do not treat bunches with arbitrary σ_x and σ_y, but treat round beams with $\sigma_x = \sigma_y = \sigma$ instead.

5. I often assume collision between bunches moving at relativistic speeds with $\gamma \gg 1$, and assume $\beta \approx 1$. Here $\beta = v/c$ is the bunch velocity in units of the speed of light c.

6. I consider collisions between particles of opposite charge, and hence attractive beam-beam forces. For collisions of particles of the same charge and repulsive beam-beam forces, the signs of many equations must be changed.

2 ELECTRO-MAGNETIC FIELD OF A MOVING BUNCH

In the case of a Gaussian charge distribution in two dimensions, given by Eq. (3), the beam-beam fields can be derived from a beam-beam potential $U(x, y; \sigma_x, \sigma_y)$ which has the following integral representation for arbitrary σ_x and σ_y:

$$U(x, y; \sigma_x, \sigma_y) = \frac{ne}{4\pi\epsilon_0} \int_0^\infty \frac{\exp\left(-\frac{x^2}{2\sigma_x^2+t} - \frac{y^2}{2\sigma_y^2+t}\right)}{\sqrt{(2\sigma_x^2+t)(2\sigma_y^2+t)}} dt \tag{4}$$

This expression was derived by Kellogg [4], Houssais [5] at the University of Rennes, Kheifets [6], and Takayama [7]. The expression for the potential U in three dimensions is very similar. As usual, the electric field $\vec{E}$ is obtained by taking the gradient of U, i.e. $\vec{E} = -\nabla U$. The components of $\vec{E}$ were derived in closed form by Augustin [8] and Bassetti and Erskine [9]. Talman [10] gives a derivation of the potential U and of the components of $\vec{E}$ which are for $\sigma_x > \sigma_y$:

$$E_x - iE_y = -\frac{ine}{2\epsilon_0\sqrt{2\pi(\sigma_x^2-\sigma_y^2)}} \left[w\left(\frac{x+iy}{\sqrt{2(\sigma_x^2-\sigma_y^2)}}\right) - \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) w\left(\frac{\frac{x\sigma_y}{\sigma_x} + \frac{iy\sigma_x}{\sigma_y}}{\sqrt{2(\sigma_x^2-\sigma_y^2)}}\right) \right] \tag{5}$$

Here $w(z)$ is the complex error function [11]. It is sometimes convenient to leave the electric field components in their integral form, as will be done for calculating the coherent

beam-beam kicks. The components of the magnetic field $\vec{B}$ in the laboratory frame are most easily obtained by observing that the electro-magnetic field of a bunch is an electrostatic field in its rest frame, and that $\vec{B}$ vanishes there. Hence, the magnetic field $\vec{B} = \vec{\beta} \times \vec{E}/c$ of a bunch moving with speed $\vec{\beta}c$ in the negative s-direction of the laboratory frame follows from the electric field $\vec{E}$. In Cartesian coordinates, the relations between the components of the electric field $\vec{E}$ and magnetic field $\vec{B}$ are:

$$B_x = \beta E_y/c \qquad\qquad B_y = -\beta E_x/c \tag{6}$$

In round beams with $\sigma_x = \sigma_y = \sigma$, we use cylindrical coordinates and find that the electric field $\vec{E}$ and magnetic field $\vec{B}$ have only a radial and an azimuthal component, respectively:

$$E_r = -\frac{ne}{4\pi\epsilon_0}\frac{\partial}{\partial r}\int_0^\infty \frac{\exp\left(-\frac{r^2}{2\sigma^2+t}\right)}{2\sigma^2+t}\mathrm{d}t \qquad\qquad B_\phi = \frac{ne\beta c\mu_0}{4\pi}\frac{\partial}{\partial r}\int_0^\infty \frac{\exp\left(-\frac{r^2}{2\sigma^2+t}\right)}{2\sigma^2+t}\mathrm{d}t \tag{7}$$

It can be verified without tables of integrals and derivatives, differentiating first and then integrating, using $1/(2\sigma^2+t)$ as integration variable, that these expressions, combined with the Lorentz force in Eq. (8), yield Eq. (10).

3 INCOHERENT BEAM-BEAM EFFECTS

In this section, we first use the electro-magnetic field of a moving bunch to calculate the incoherent beam-beam force which a bunch exerts on the particles in another bunch, moving in the opposite direction, during the collision. We then integrate this force over the collision, and obtain the incoherent beam-beam kick. Finally, we compute the incoherent linear beam-beam tune shift and the incoherent beam-beam limit.

3.1 Incoherent beam-beam force

The incoherent beam-beam force $\vec{F}$ is that force which a bunch exerts on individual particles of the bunch travelling in the opposite direction. It follows from the Lorentz force acting on a particle with charge $-e$:

$$\vec{F} = -e\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{8}$$

In the case of round beams with $\sigma_x = \sigma_y = \sigma$, and of long bunches with $r \ll \sigma_s$, the electro-magnetic field has only two non-vanishing components, E_r and B_ϕ. The most direct and transparent way of finding E_r and B_ϕ is using Gauss's and Ampere's laws, respectively, in cylindrical coordinates, instead of working with Eqs. (6) and (7). The result is:

$$2\pi r E_r = (1/\epsilon_0)\int_0^r 2\pi r'\rho(r')\mathrm{d}r' \qquad\qquad 2\pi r B_\phi = \mu_0\int_0^r 2\pi r'\beta c\rho(r')\mathrm{d}r' \tag{9}$$

The dependences of E_r and B_ϕ on the radius r are the same. Hence, the Lorentz force has only an r component F_r. For a round beam with a Gaussian distribution in two dimensions, given by Eq. (3), the integral can be expressed in closed form:

$$F_r(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0 r}\left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{10}$$

The sign of β^2 corresponds to the case of the particles and the bunch moving in opposite directions. When particle and bunch travel in the same direction, the sign of β^2 is negative. The sign of the incoherent beam-beam force $F_r(r)$ depends on the signs of the charges in the two bunches. If the two beams contain particles of opposite charge, the force is attractive, and the sign is negative, as shown in Eq. (10). In Cartesian co-ordinates with $r = \sqrt{x^2 + y^2}$, the forces become:

$$F_x(x, y) = F_r(r)x/r \qquad F_y(x, y) = F_r(r)y/r \tag{11}$$

The incoherent beam-beam force $F_r(r)$ is shown in Figure 1. For $r \ll \sigma$, the force $F_r(r)$ increases linearly, and is given by:

$$F_r(r) = -\frac{ne^2(1+\beta^2)r}{4\pi\epsilon_0\sigma^2} \qquad \text{for } r \ll \sigma \tag{12}$$

The linear beam-beam force is similar to that in a quadrupole, but in contrast to a quadrupole it has the same sign in both x and y directions. For $r \gg \sigma$, the force drops like $1/r$:

$$F_r(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0 r} \qquad \text{for } r \gg \sigma \tag{13}$$

For elliptical Gaussian beams with $\sigma_x > \sigma_y$, it follows from Eqs. (5), (6) and (8) that the beam-beam force can be written as:

$$F_x - iF_y = -e(1+\beta^2)(E_x - iE_y) \tag{14}$$

As Eq. (6) this result applies to arbitrary charge distributions in a bunch.

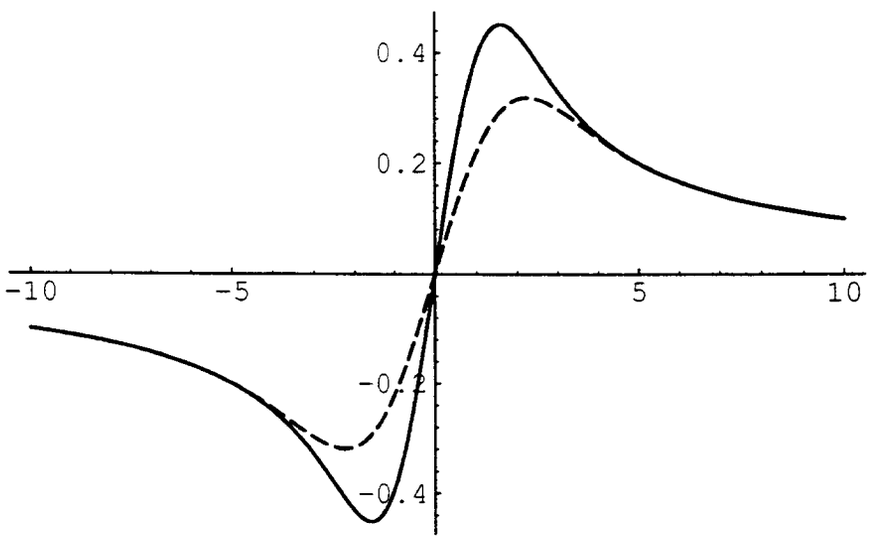


Figure 1: Comparison of the incoherent (full line) and coherent (dashed line) beam-beam kicks for round Gaussian beams. The abscissa is drawn in units of σ; the ordinate is arbitrary.

3.2 Incoherent beam-beam kick

The beam-beam force due to a bunch moving in the negative s-direction at speed v can be obtained from the two-dimensional force in Eq. (10) by multiplying it with the longitudinal density distribution, centered at $s = -vt$:

$$F_r(r,s,t) = -\frac{Ne^2(1+\beta^2)}{(2\pi)^{3/2}\epsilon_0 r\sigma_s}\left[1-\exp\left(-\frac{r^2}{2\sigma^2}\right)\right]\exp\left[-\frac{(s+vt)^2}{2\sigma_s^2}\right] \tag{15}$$

This approximation is justified when the bunch length σ_s is much larger than the transverse beam radii σ_x and σ_y. The beam-beam kick $\Delta r'$ is obtained from the beam-beam force by integration over the collision, remembering that the test particle is at $s = vt$:

$$mc\beta\gamma\Delta r' = \int_{-\infty}^{+\infty} F_r(r,s,t)\mathrm{d}t \tag{16}$$

Introducing the classical particle radius $r_0 = e^2/4\pi\epsilon_0 mc^2$, assuming $\beta \approx 1$, and performing the integration over t, we find:

$$\Delta r' = -\frac{2Nr_0}{\gamma r}\left[1-\exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{17}$$

For $r \ll \sigma$, the beam-beam kick approaches the following limit, where δ is the inverse focal length of the quadrupole representing the beam-beam kick:

$$\Delta r' = -\frac{Nr_0 r}{\gamma\sigma^2} = -r\delta \qquad \text{for } r \ll \sigma \tag{18}$$

3.3 Incoherent linear beam-beam tune shift

The following calculation applies to both the horizontal and the vertical plane. Therefore, I use the general subscript z. Using Eq. (18) for $z \ll \sigma$, the linear map M in the z-plane through an arc of the storage ring with phase advance $2\pi Q_z$ and amplitude function β_z at the collision point and the beam-beam collision becomes:

$$M = \begin{pmatrix} \cos 2\pi Q_z & \beta_z \sin 2\pi Q_z \\ -\beta_z^{-1}\sin 2\pi Q_z & \cos 2\pi Q_z \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\delta & 1 \end{pmatrix} \tag{19}$$

The trace of the map M yields the perturbed tune $(Q_z + \xi_z)$:

$$\mathrm{Tr}(M) = 2\cos 2\pi(Q_z+\xi_z) = 2\cos 2\pi Q_z - \beta_z\delta\sin 2\pi Q_z \tag{20}$$

Expanding the trigonometric functions and assuming $\beta \approx 1$ as in Eq. (18) yields for the linear beam-beam tune shift $\xi_z \ll 1$:

$$\xi_z = \beta_z\delta/4\pi = \frac{Nr_0\beta_z}{4\pi\gamma\sigma^2} \tag{21}$$

It is easy enough to compute the linear beam-beam tune shifts for Gaussian beams with an elliptical cross section and rms radii σ_x and σ_y, by expanding the functions appearing in Eq. (5) into Taylor series, keeping only the lowest order terms in the real and imaginary

parts, respectively, and repeating the calculations leading to Eq. (21). For $\beta \approx 1$ the result for the beam-beam tune shift in the z-plane, ξ_z, is:

$$\xi_z = \frac{N r_0 \beta_z}{2\pi\gamma\sigma_z(\sigma_x + \sigma_y)} \tag{22}$$

Inserting the vertical beam-beam tune shift ξ_y into the luminosity formula Eq. (1), and assuming $\sigma_y \ll \sigma_x$, yields the following expression which is often used in the discussion of storage ring performance:

$$L \approx \frac{N f k \xi_y \gamma}{2 r_0 \beta_y} \tag{23}$$

The product Nfk in the numerator is the circulating current in one beam, often limited by effects other than beam-beam collisions, i.e. collective effects and/or RF power. Beam-beam effects enter into Eq. (23) by the factor ξ_y in the numerator. Therefore, all storage rings operate close to the beam-beam limit. The energy in terms of γ also appears in the numerator. It is also advantageous to have a small value of β_y at the collision point, since it appears in the denominator. Therefore, the beam-beam collisions take place in low-β insertions in all modern storage rings.

3.4 Incoherent beam-beam limit

The limit of stability of the linear betatron oscillations is reached when $\mathrm{Tr}(M) = \pm 2$ in Eq. (20). Solving it for $\hat{\xi}_z$ leads to the following expression for $\hat{\xi}_z > 0$:

$$\hat{\xi}_z = \begin{cases} (1/2\pi)\cot \pi Q_z & \text{for } 0 < Q_z \bmod 1 < 0.5 \\ -(1/2\pi)\tan \pi Q_z & \text{for } 0.5 < Q_z \bmod 1 < 1 \end{cases} \tag{24}$$

The linear beam-beam limit $\hat{\xi}_z$, shown in Figure 2, is periodic in Q_z with period one half. For most fractional tunes this limit is much higher than the observed limit $\xi \approx 0.03$. This suggests that the observed beam-beam limit ξ is not caused by the linear beam-beam limit $\hat{\xi}_z$.

4 COHERENT BEAM-BEAM EFFECTS

Coherent beam-beam effects arise from the forces which an exciting bunch exerts on a whole test bunch during a beam-beam collision. Based on the material already presented, we will first calculate the coherent beam-beam kicks, then linearize them to obtain the linear coherent tune shift, and finally compute the coherent beam-beam limit.

4.1 Coherent beam-beam kicks

Coherent beam-beam kicks are the kicks which a whole test bunch receives by colliding with a whole bunch travelling in the opposite direction. They are obtained from the incoherent beam-beam kicks by integration over the charge distribution in the test bunch. The symmetry of the beam-beam force implies that the coherent beam-beam kick vanishes for head-on collisions. Let us treat the two beams on an equal footing, with the first beam acting on the second beam and vice-versa, and introduce offsets $\bar{x}_\pm$ and $\bar{y}_\pm$ of the two beams. In order to evaluate the coherent horizontal beam-beam kick $\Delta\bar{x}'_+$ for Gaussian beams with elliptical cross section, we follow [18] and start from the integral

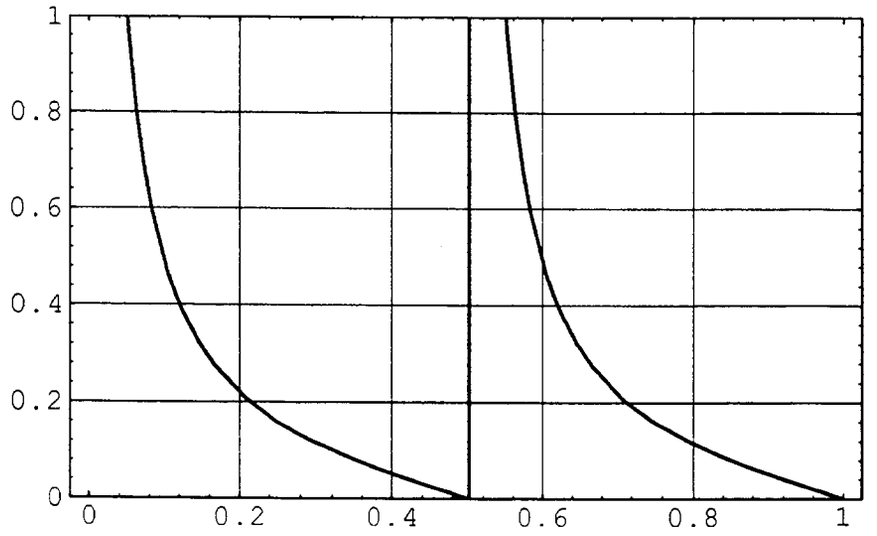


Figure 2: Incoherent beam-beam limit $\hat{\xi}_z$ as a function of the tune Q_z. The abscissa is Q_z, the ordinate is $\hat{\xi}_z$.

representation of the beam-beam potential in Eq. (4), integrated over the beam-beam collision, using Eqs. (3), (6), (14), and (16):

$$-\frac{2\pi\sigma_x\sigma_y\gamma\Delta\bar{x}'_+}{Nr_0} = \int_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y G(x-\bar{x}_+, y-\bar{y}_+; \sigma_x, \sigma_y) \frac{\partial}{\partial x} \int_0^\infty \frac{\exp\left(-\frac{(x-\bar{x}_-)^2}{2\sigma_x^2+t} - \frac{(y-\bar{y}_-)^2}{2\sigma_y^2+t}\right)}{\sqrt{(2\sigma_x^2+t)(2\sigma_y^2+t)}}\mathrm{d}t \tag{25}$$

The equation for the coherent vertical beam-beam kick $\Delta\bar{y}'_+$ is very similar. An integration by parts of the right hand side moves the differential operator from the beam-beam potential to the Gaussian density distribution:

$$= -\int_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y \frac{\partial}{\partial x} G(x-\bar{x}_+, y-\bar{y}_+, \sigma_x, \sigma_y) \int_0^\infty \frac{\exp\left(-\frac{(x-\bar{x}_-)^2}{2\sigma_x^2+t} - \frac{(y-\bar{y}_-)^2}{2\sigma_y^2+t}\right)}{\sqrt{(2\sigma_x^2+t)(2\sigma_y^2+t)}}\mathrm{d}t \tag{26}$$

Because of the symmetry of G in x and $\bar{x}_+$ the differentiation with respect to x may be replaced by a differentiation with respect to $\bar{x}_+$:

$$= \int_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y \frac{\partial}{\partial \bar{x}_+} G(x-\bar{x}_+, y-\bar{y}_+, \sigma_x, \sigma_y) \int_0^\infty \frac{\exp\left(-\frac{(x-\bar{x}_-)^2}{2\sigma_x^2+t} - \frac{(y-\bar{y}_-)^2}{2\sigma_y^2+t}\right)}{\sqrt{(2\sigma_x^2+t)(2\sigma_y^2+t)}}\mathrm{d}t \tag{27}$$

Interchanging integrals and differentiation moves the differentiation outside the integrals:

$$= \frac{\partial}{\partial \bar{x}_+} \int_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y G(x-\bar{x}_+, y-\bar{y}_+, \sigma_x, \sigma_y) \int_0^\infty \frac{\exp\left(-\frac{(x-\bar{x}_-)^2}{2\sigma_x^2+t} - \frac{(y-\bar{y}_-)^2}{2\sigma_y^2+t}\right)}{\sqrt{(2\sigma_x^2+t)(2\sigma_y^2+t)}}\mathrm{d}t \tag{28}$$

The integrals in x and y have the same form, and can be done analytically. The remaining integral in t is very similar to the beam-beam potential in Eq. (4):

$$U_c = \frac{ne}{4\pi\epsilon_0}\int_0^\infty \frac{\exp\left(-\frac{(\bar{x}_+-\bar{x}_-)^2}{4\sigma_x^2+t} - \frac{(\bar{y}_+-\bar{y}_-)^2}{4\sigma_y^2+t}\right)}{\sqrt{(4\sigma_x^2+t)(4\sigma_y^2+t)}}\mathrm{d}t \tag{29}$$

The coherent beam-beam potential U_c contains the differences of the offsets $\bar{x}_\pm$ and $\bar{y}_\pm$ instead of x and y, and a factor 4 instead of 2 in front of the σ's. In the case of round Gaussian beams this yields the following expression for the coherent beam-beam kick $\Delta\bar{r}'$ with $\bar{r} = \sqrt{(\bar{x}_+ - \bar{x}_-)^2 + \bar{y}_+ - \bar{y}_-)^2}$:

$$\Delta\bar{r}' = -\frac{2Nr_0}{\gamma\bar{r}}\left[1 - \exp\left(-\frac{\bar{r}^2}{4\sigma^2}\right)\right] \tag{30}$$

Figure 1 shows a comparison of the incoherent and the coherent beam-beam kicks. The coherent kick has half the slope of incoherent kick for $\bar{r} \ll \sigma$. The two kicks are equal for $\bar{r} \gg \sigma$.

4.2 Linear coherent beam-beam tune shift

Using the same arguments as in the case of the incoherent beam-beam kicks, we find for the coherent beam-beam kick for $\bar{r} \ll \sigma$ for round Gaussian beams with $\beta \approx 1$:

$$\Delta\bar{r}' = -\frac{Nr_0\bar{r}}{2\gamma\sigma^2} = -\bar{r}\Delta \qquad \text{for } \bar{r} \ll \sigma \tag{31}$$

Here, $\Delta = \delta/2$ is the inverse focal length of the quadrupole representing the linear coherent beam-beam kick. The linear coherent beam-beam tune shift $\Xi_z \ll 1$ can be calculated for $\bar{x}_+ - \bar{x}_- \ll \sigma$ and $\bar{y}_+ - \bar{y}_- \ll \sigma$ and becomes just one half of the linear incoherent tune shift ξ_z shown in Eq. (21):

$$\Xi_z = \frac{Nr_0\beta_z}{8\pi\gamma\sigma^2} \qquad \text{for } \bar{x}_+ - \bar{x}_- \ll \sigma \text{ and } \bar{y}_+ - \bar{y}_- \ll \sigma \tag{32}$$

This result can be generalized for elliptical Gaussian beams:

$$\Xi_z = \frac{Nr_0\beta_z}{4\pi\gamma\sigma_z(\sigma_x+\sigma_y)} \qquad \text{for } \bar{x}_+ - \bar{x}_- \ll \sigma_x \text{ and } \bar{y}_+ - \bar{y}_- \ll \sigma_y \tag{33}$$

Not surprisingly, this is one half of the incoherent beam-beam tune shift in Eq. (22). If the two beams have different rms radii at the interaction points, all σ's appearing in Eq. (33) must be replaced by $\sqrt{(\sigma_+^2 + \sigma_-^2)/2}$ [18].

4.3 Coherent beam-beam limit

I work in the approximation $z \ll \sigma$ and describe the bunch motion in two beams and one plane by matrices. The rotation matrix R operates on the transpose of the vector $\vec{z} = (z_+, z'_+, z_-, z'_-)$, and transports bunches between collision points. In the case of just

one bunch in each of the two beams, the order of R is four:

$$R = \begin{pmatrix} C_+ & \beta_+ S_+ & 0 & 0 \\ -S_+/\beta_+ & C_+ & 0 & 0 \\ 0 & 0 & C_- & \beta_- S_- \\ 0 & 0 & -S_-/\beta_- & C_- \end{pmatrix} \tag{34}$$

Here $C_\pm = \cos\mu_\pm$ and $S_\pm = \sin\mu_\pm$, $\mu_\pm$ is the phase advance between the collision points, and the index $\pm$ marks the beam. The $\beta_\pm$ are taken at the interaction point(s), the $\alpha_\pm$ there are assumed to vanish. The motion of each bunch is described by a 2×2 block matrix. The block matrices for all bunches are arranged along the main diagonal. The kick matrix K describes the beam-beam kicks and in the case of equal bunches is given by:

$$K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\Delta & 1 & +\Delta & 0 \\ 0 & 0 & 1 & 0 \\ +\Delta & 0 & -\Delta & 1 \end{pmatrix} \tag{35}$$

The kick matrix K has ones along the main diagonal. For each pair of bunches there are four kick terms Δ which link the changes of the slopes to the positions of the two bunches which are colliding at a collision point. The generalization to $k > 1$ bunches is obvious. Then the kick matrix K contains four kick terms for all pairs of bunches which collide at the same time. The stability of the one-turn map consisting of $2k$ R-matrices alternating with $2k$ K-matrices of order $4k$ can be analysed by computing its eigenvalues. Closed analytical solutions for the coherent beam-beam limit $\hat{\Xi}$ are known in symmetrical cases [19]. When all $2k$ interaction points and all k bunches in the two beams have the same parameters, and if all $2k$ phase advances between the interaction points are the same, the coherent beam-beam limit $\hat{\Xi}$ is:

$$\hat{\Xi} = \frac{\cos(\pi Q/k) - \cos(\{Q+1\}\pi/k)}{2\pi \sin(\pi Q/k)} \tag{36}$$

Here, $\{\ldots\}$ denotes the integral part. Figure 3 shows the coherent beam-beam limit $\hat{\Xi}$ for $k = 4$ bunches in each beam, colliding in eight collision points. $\hat{\Xi}$ is periodic every four units of tune Q, and approaches zero when the tune Q approaches an integer from below. With increasing k, $\hat{\Xi}$ decreases. For $k = 4$, it is already much smaller than the incoherent limit $\hat{\xi}$, as becomes obvious by comparing Figures 2 and 3.

5 NONLINEAR BEAM-BEAM EFFECTS

The nonlinear variation of the beam-beam force with r in a round Gaussian beam, shown in Eq. (10), causes a tune shift $\Delta Q(a)$ with amplitude a and a tune spread in the beams, and drives nonlinear resonances. The usual method for treating these effects is Hamiltonian perturbation theory. Figure 4 shows $\Delta Q(a)/\xi$ as a function of a/σ for round Gaussian beams.

5.1 Hamiltonian formalism

Following [20] I start from the differential equation for betatron oscillations in the z coordinate with a periodic linear driving term $K(s)$ and a periodic beam-beam kick,

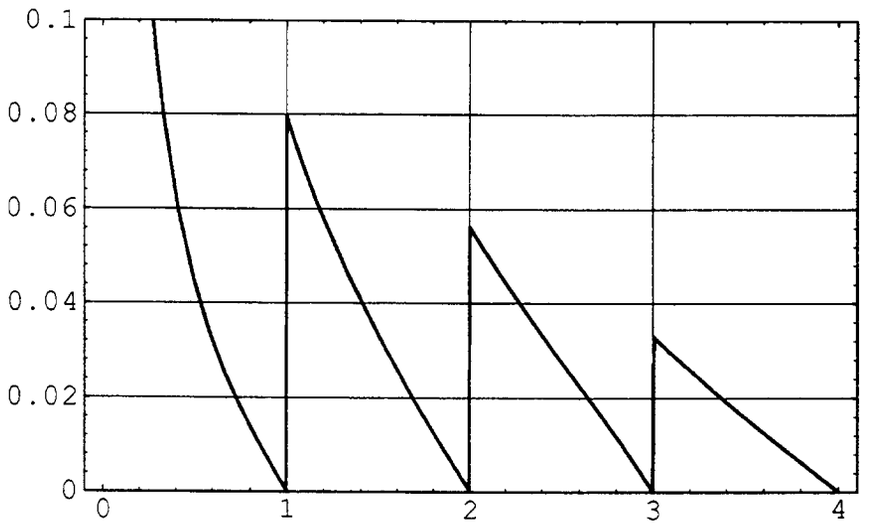


Figure 3: Coherent beam-beam limit $\hat{\Xi}$ for $k = 4$ bunches colliding in eight collision points as a function of the tune Q. The abscissa is Q. The ordinate $\hat{\Xi}$ is periodic every four units of Q.

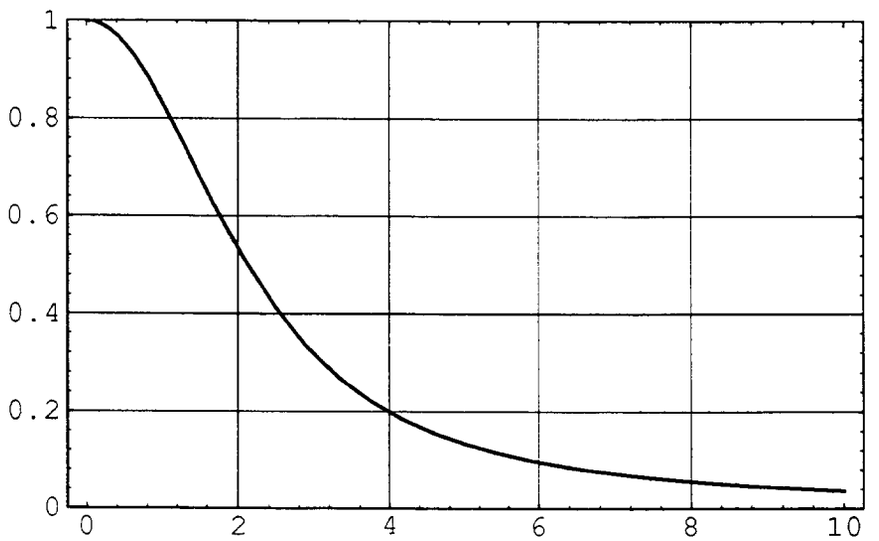


Figure 4: Tune shift versus amplitude for round Gaussian beams. The abscissa is the amplitude a in units of σ. The ordinate is the ratio $\Delta Q(a)/\Delta Q(0)$.

obtained from Eqs. (17) and (21), and restrict the derivation to one dimension and round Gaussian beams:

$$\frac{\mathrm{d}^2 z}{\mathrm{d}s^2} + K(s)z = -\frac{8\pi\sigma^2\xi}{\beta_z}\left[\frac{1-\exp(-z^2/2\sigma^2)}{z}\right]\delta_p(s) \tag{37}$$

Here $\delta_p(s)$ is the periodic δ-function with period $2\pi R$, where R is the average radius, assuming just one beam-beam collision per turn. I now apply a Courant-Snyder transformation [21] to normalized betatron phase space (η, η') with $\eta = z/\sqrt{\beta_z}$ and $\theta = \int \mathrm{d}s/(Q\beta_z)$:

$$\frac{\mathrm{d}^2\eta}{\mathrm{d}\theta^2} + Q^2\eta = -\frac{8\pi Q\xi\sigma^2}{\beta_z\eta}\left[1-\exp\left(-\frac{\beta_z\eta^2}{2\sigma^2}\right)\right]\delta_p(\theta) \tag{38}$$

The new independent variable θ and $\delta_p(\theta)$ have period 2π, β_z is taken at the collision point, and Q is the tune. The solutions of the homogeneous equation are circles, and the constants ϵ and ϕ are given by the initial conditions:

$$\eta(\theta) = \sqrt{\epsilon}\cos(Q\theta+\varphi) = \sqrt{\epsilon}\cos\phi \tag{39}$$

To solve the inhomogeneous Eq. (38), we assume that ϵ and ϕ are slowly varying functions of θ, use the method of variation of constants, and obtain the following set of equations for η and η', where in the second equation we already neglect terms in ϵ' and ϕ', where the prime ($'$) denotes differentiation with respect to θ:

$$\eta = \sqrt{\epsilon}\cos\phi \qquad\qquad \eta' = -Q\sqrt{\epsilon}\sin\phi \tag{40}$$

Changing to action-angle variables with

$$\epsilon = \eta^2 + (\eta'/Q)^2 \qquad\qquad \phi = -\arctan(\eta'/Q\eta) \tag{41}$$

and differentiating, we obtain first-order differential equations for ϵ and ϕ:

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}\theta} = 2\eta\eta' + \frac{2\eta'\eta''}{Q^2} = \frac{16\pi\xi\sigma^2\eta'}{Q\eta\beta_z}\left[1-\exp\left(-\frac{\beta_z\eta^2}{2\sigma^2}\right)\right]\delta_p(\theta) \tag{42}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}\theta} = -\frac{1}{1+[\eta'/(Q\eta)^2]}\left(\frac{\eta''}{Q\eta} - \frac{\eta'^2}{Q\eta^2}\right) = Q + \frac{8\pi Q\xi\sigma^2}{\epsilon\beta_z}\left[1-\exp\left(-\frac{\beta_z\eta^2}{2\sigma^2}\right)\right]\delta_p(\theta) \tag{43}$$

The final step in first order perturbation theory is substituting the unperturbed solution into the right-hand sides, and replacing the periodic δ-function by its Fourier expansion $\delta_p(\theta) = (1/2\pi)\sum_{m=-\infty}^{\infty}\exp(-im\theta)$ which yields:

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}\theta} = 8\xi\sigma^2\sin\phi\sqrt{\epsilon/\beta_z}\left[\frac{1-\exp(-\beta_z\eta^2/2\sigma^2)}{\eta\sqrt{\beta_z}}\right]\sum_{m=-\infty}^{\infty}\exp(-im\theta) \tag{44}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}\theta} = Q + \frac{4\xi\sigma^2\cos\phi}{\sqrt{\epsilon\beta_z}}\left[\frac{1-\exp(-\beta_z\eta^2/2\sigma^2)}{\eta\sqrt{\beta_z}}\right]\sum_{m=-\infty}^{\infty}\exp(-im\theta) \tag{45}$$

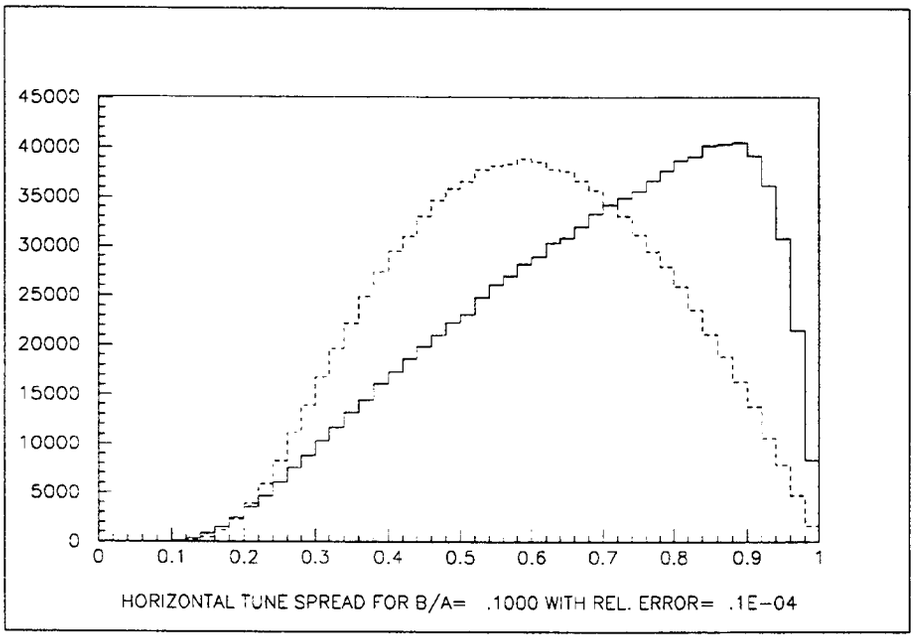


Figure 5: Distribution of the incoherent horizontal (full line) and vertical (dashed line) beam-beam tune shifts for a Gaussian beam with $\sigma_y/\sigma_x = 0.1$. The abscissa is in units of the linear tune shifts $\xi_x = \xi_y$.

5.2 Tune shift with amplitude

The tune shift with amplitude follows from Eq. (45) by only taking the constant term in the Fourier expansion of the δ-function, and averaging the remainder over all phases ϕ [20]. Observing that $\cos^2\phi = (1+\cos 2\phi)/2$, and using the integral representation of the modified Bessel function I_0 of order 0 [12], we find:

$$\overline{\frac{\mathrm{d}\phi}{\mathrm{d}\theta}} = Q + \frac{\xi}{\alpha\pi}\int_0^{2\pi}\left[1-\exp(-\alpha\cos^2\varphi)\right]\mathrm{d}\phi = Q + (2\xi/\alpha)\left[1-\exp(-\alpha/2)I_0(\alpha/2)\right] \quad (46)$$

Here $\alpha = \epsilon\beta_z/(2\sigma^2) = (a/\sigma)^2/2$ where a is the betatron amplitude. The extra term on the right hand side is the nonlinear tune shift $\Delta Q(a)$ with amplitude, which is shown in Figure 4:

$$\Delta Q(a)/\xi = (2/\alpha)\left[1-\exp\left(-\alpha/2\right)I_0\left(\alpha/2\right)\right] \quad (47)$$

The tune shift $\Delta Q(a)$ is approximately equal to ξ for $a \ll \sigma$. However, for $a \gg \sigma$, the tune shift $\Delta Q(a)$ is much smaller than ξ. This behaviour of $\Delta Q(a)$ is completely different from that due to multipole components of the magnetic guiding field in a storage ring. For components with eight or more poles, the absolute value of the tune shift is a monotonically increasing function of the amplitude in first order. The tune shifts with amplitude of elliptical Gaussian beams are functions of the horizontal and vertical amplitudes, a and b, and of the ratio σ_y/σ_x [22, 23]. The tune shifts with amplitude cause a tune spread within the beam. For the Gaussian density distribution in Eq. (3), the distributions in the amplitudes a and b are Rayleigh distributions given by:

$$R(a) = (a/\sigma_x^2)\exp(-a^2/2\sigma_x^2) \qquad R(b) = (b/\sigma_y^2)\exp(-b^2/2\sigma_x^2) \quad (48)$$

The results of a Monte Carlo calculation [22] of the distribution functions of the horizontal and vertical beam-beam tune shifts for a beam with $\sigma_y/\sigma_x = 0.1$, based on 10^6 random pairs of amplitudes (a, b), are shown in Figure 5. The tune spreads are almost as large as the linear beam-beam tune shifts ξ_x and ξ_y, respectively. The horizontal distribution has a peak close to ξ_x and a long tail towards smaller tune shifts. The vertical distribution is more symmetrical.

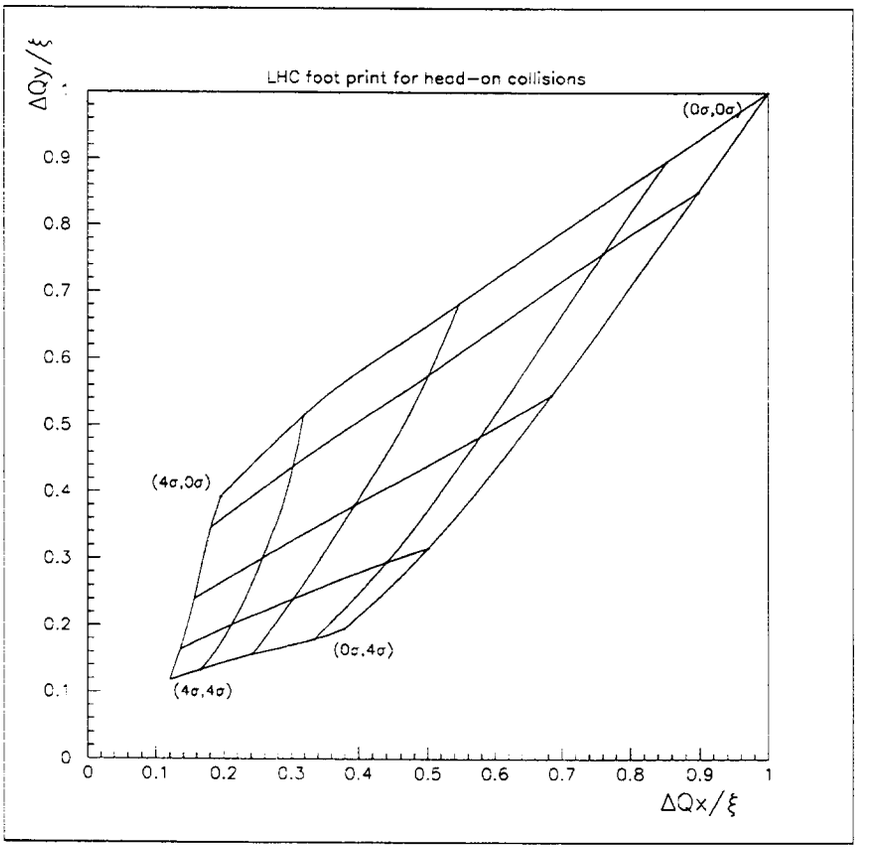


Figure 6: Footprint of head-on beam-beam collisions of round Gaussian beams in the LHC. Abscissa and ordinate are the amplitude-dependent beam-beam tune shifts ΔQ_x and ΔQ_y, respectively, in units of the linear beam-beam tune shift ξ. The lines mark constant amplitudes of horizontal and vertical betatron oscillations.

Another way of representing the effects of the beam-beam tune shift is known as footprints, and shown in Figure 6. Here one plots the tunes Q_x and Q_y in the (Q_x, Q_y)-plane with the amplitudes of the betatron oscillations as a parameter. In the case of equal emittances, equal β-functions at the interaction points in the horizontal and vertical plane, and head-on collisions, the footprint is symmetrical with respect to the diagonal line $Q_x = Q_y$. Particles with vanishing betatron amplitudes have the highest tune shifts $\Delta Q_x \approx \Delta Q_y \to \xi$, particles with large amplitudes the smallest ones. The tune shifts are concentrated in the neighbourhood of the diagonal line $\Delta Q_x \approx \Delta Q_y$.

5.3 Beam-beam resonances

In order to compute the excitation of nonlinear resonances driven by the beam-beam collisions, the slowly varying terms in the equations of motion, Eqs. (44) and (45), must be identified and extracted. We obtain closed expressions for the resonance excitation by following [20], and replacing the function in the square bracket by its Fourier transform $\tilde{g}(\omega)$:

$$\frac{1-\exp(-\beta_z\eta^2/2\sigma^2)}{\eta\sqrt{\beta_z}} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{g}(\omega)\exp(i\omega\sqrt{\epsilon\beta_z}\cos\phi)\mathrm{d}\omega \tag{49}$$

Using the identities for Bessel functions J_n of order n [13]

$$J_{n-1}(x)+J_{n+1}(x)=(2n/x)J_n(x) \qquad J_{n-1}(x)-J_{n+1}(x)=2J_n'(x) \tag{50}$$

and the following relation which follows from [24] after a little algebra

$$\exp(i\omega\sqrt{\epsilon\beta_z}\cos\phi)=\sum_{k=-\infty}^{\infty}J_k(\omega\sqrt{\epsilon\beta_z})\exp\left[ik(\phi+\pi/2)\right] \tag{51}$$

and integrating once by parts, the equations for ϵ and ϕ take the form:

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}\theta} = -\frac{4\xi\sigma^2}{\pi\beta_z}\sum_{m,n=-\infty}^{\infty}ni^n\exp\left[i(n\phi-m\theta)\right]\int_{-\infty}^{\infty}\tilde{g}(\omega)J_n(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega/\omega \tag{52}$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}\theta} = Q+\frac{2i\xi\sigma^2}{\pi\beta_z\epsilon}\sum_{m,n=-\infty}^{\infty}i^n\exp\left[i(n\phi-m\theta)\right]\int_{-\infty}^{\infty}\tilde{g}'(\omega)J_n(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega \tag{53}$$

These equations are still an exact representation of the original differential equation and the beam-beam kicks. Let us define a resonance by choosing two relatively prime integers p and q such that the phase $\chi=p\phi-q\theta$ can be considered as varying more slowly than other phases, and keep in the double sum only terms of the form $\exp\left[i\ell(p\phi-q\theta)\right]$, where ℓ is any integer – positive, negative, or zero. Then the equations for ϵ and χ become approximately:

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}\theta} = -\frac{4p\xi\sigma^2}{\pi\beta_z}\sum_{\ell=-\infty}^{\infty}\ell\exp\left[i\ell(\chi+p\pi/2)\right]\int_{-\infty}^{\infty}\tilde{g}(\omega)J_{\ell p}(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega/\omega \tag{54}$$

$$\frac{\mathrm{d}\chi}{\mathrm{d}\theta} = (pQ-q)+\frac{2ip\xi\sigma^2}{\pi\beta_z\epsilon}\sum_{\ell=-\infty}^{\infty}\exp\left[i\ell(\chi+p\pi/2)\right]\int_{-\infty}^{\infty}\tilde{g}'(\omega)J_{\ell p}(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega \tag{55}$$

The sums in the equations for ϵ and χ can be simplified, using $J_{-2n}(x)=J_{2n}(x)$:

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}\theta} = -\frac{8ip\xi\sigma^2}{\pi\beta_z}\sum_{\ell=1}^{\infty}\ell\sin\left[\ell(\chi+p\pi/2)\right]\int_{-\infty}^{\infty}\tilde{g}(\omega)J_{\ell p}(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega/\omega \tag{56}$$

$$\begin{aligned}\frac{\mathrm{d}\chi}{\mathrm{d}\theta} = (pQ-q)+\frac{4ip\xi\sigma^2}{\pi\beta_z\epsilon}\Bigg[&\frac{1}{2}\int_{-\infty}^{\infty}\tilde{g}'(\omega)J_0(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega\\ &+\sum_{\ell=1}^{\infty}\cos\left[\ell(\chi+p\pi/2)\right]\int_{-\infty}^{\infty}\tilde{g}'(\omega)J_{\ell p}(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega\Bigg]\end{aligned} \tag{57}$$

We now limit the sums over ℓ to the term with $\ell = 1$. This is a good approximation [20] for resonances of high order p. According to Eq. (56), the fixed points occur at $\chi + p\pi/2 = 0$ and π. Their amplitudes are obtained by setting Eq. (57) equal to zero. The result is:

$$-\frac{pQ-q}{\xi} = \frac{2ip\sigma^2}{\pi\beta_z\epsilon}\left[\int_{-\infty}^{\infty}\tilde{g}'(\omega)J_0(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega \pm 2\int_{-\infty}^{\infty}\tilde{g}'(\omega)J_p(\omega\sqrt{\epsilon\beta_z})\mathrm{d}\omega\right] \quad (58)$$

$$= (2p/\alpha)\left[1 - \exp(-\alpha/2)I_0(\alpha/2) \pm 2\exp(-\alpha/2)I_{p/2}(\alpha/2)\right] \quad (59)$$

Here, the abbreviation $\alpha = \beta_z\epsilon/2\sigma^2$ in Eq. (47) has been used again, as well as the explicit equation [20] $\tilde{g}'(\omega) = -2i\delta(\omega) + i\sigma\sqrt{2\pi}\exp(-\omega^2\sigma^2/2)$ and the identity

$$(1/\sqrt{2\pi})\int_{-\infty}^{\infty}\mathrm{d}u\exp(-u^2/2)J_p(u\sqrt{2\alpha}) = \exp(-\alpha/2)I_{p/2}(\alpha/2) \quad (60)$$

Since $\tilde{g}(\omega)$ is an even function of ω, the integral vanishes when p is odd, i.e. resonances of

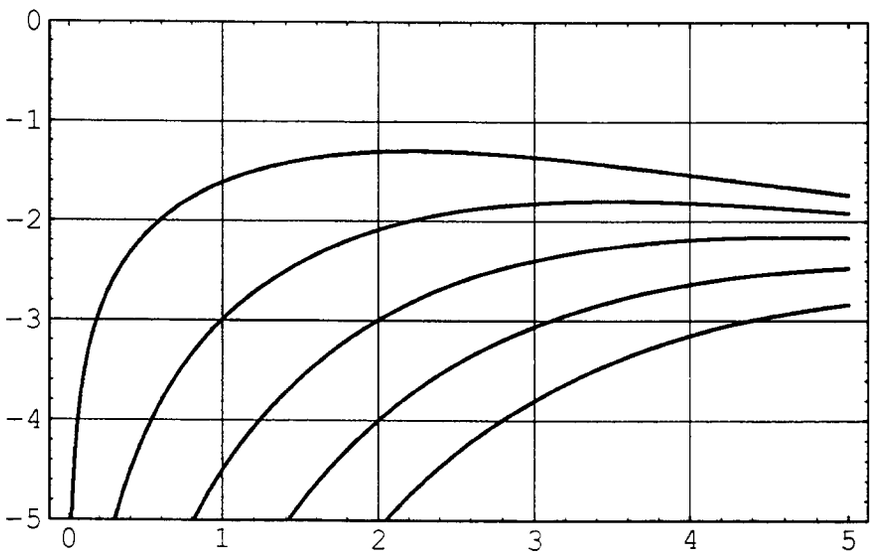


Figure 7: Resonance width for round Gaussian beams versus amplitude. The abscissa is the amplitude in units of σ. The ordinate is $\log_{10}(W_p/\xi)$, i.e. the resonance width in units of ξ. The orders of the resonances are $p = 4, 6, 8, 10, 12$, starting at the top.

odd order are not driven by head-on beam-beam collisions. For even p it can be expressed [14] in terms of the confluent hyper-geometric function $M(\frac{p+1}{2}, p+1, -\alpha)$. Applying a Kummer transformation [15] changes the sign of α. The resulting expression in M can be written in terms of modified Bessel functions [16]. The accumulated expression of Γ functions, powers of 2, etc., is reduced to unity by applying the duplication formula [17] to $\Gamma(2z)$. The first two terms in the square bracket of Eq. (59) are the amplitude dependent beam-beam tune shift which we already encountered in Eq. (47). The term following the $\pm$ sign yields the range of the tune Q for which the amplitudes of the fixed points are

smaller than that of the particle with scaled amplitude α, i.e. for which the particle is in the resonance. Solving Eq. (59) for this range of Q, we obtain the width of the resonances $W_p(\alpha)$ of order p driven by the beam-beam collisions:

$$W_p(\alpha) = \xi(4/\alpha)\exp(-\alpha/2)I_{p/2}(\alpha/2)) \tag{61}$$

Figure 7 shows $\log_{10}(W_p/\xi)$ of the resonance width as a function of the betatron amplitude a in units of the rms beam radius σ, i.e. of $a/\sigma = \sqrt{2\alpha}$. The resonance width decreases with the order p of the resonances. The maximum width moves to higher a/σ with increasing p. Resonances of odd order are not excited because of the symmetry of the beam-beam force for head-on collisions.

References

[1] S. Milton, CBN 89-1 (Cornell 1989).

[2] SPEAR Group, IEEE Trans Nucl.Sci **NS-20** (1973) 838.

[3] G.E. Fischer, SPEAR-154 (SLAC 1972).

[4] O.D. Kellogg, *Foundations of Potential Theory* (Dover, New York, 1953).

[5] B. Houssais, Thesis. University of Rennes (1967).

[6] S. Kheifets, PETRA-Kurzmitteilung 119 (DESY 1976).

[7] K. Takayama, Lett. Nuovo Cim. **34** (1982) 190.

[8] J.E. Augustin, *Turn by turn study of e^+e^- space charge effects* (SLAC 1973).

[9] M. Bassetti and G.A. Erskine, CERN-ISR-TH/80-06 (1980).

[10] R. Talman, in AIP Conference Proc. **153**, ed. by M. Month and M. Dienes (1987) 789.

[11] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965) 297.

[12] ibid., 374.

[13] ibid., 361.

[14] ibid., 486.

[15] ibid., 505.

[16] ibid., 377.

[17] ibid., 256.

[18] K. Hirata, Nucl. Instr. Meth. **A269** (1988) 7.

[19] K. Hirata and E. Keil, to be published.

[20] A.G. Ruggiero and L. Smith, PEP Summer Study (SLAC 1973).

[21] E.D. Courant and H.S. Snyder, Ann. Phys. **3** (1958) 1.

[22] E. Keil, CERN/ISR-TH/72-7 (1972).

[23] M. Meddahi and R. Schmidt, CERN-SL/90-15 (AP) (1990).

[24] G.N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge 1944) 22.

BEAMSTRAHLUNG AND DISRUPTION

L.Z. Rivkin
Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland

Abstract
Similar to the case of the storage ring colliders, beam-beam effects in future linear colliders will limit the maximum achievable luminosity. If the strong mutual focusing of the beams (high *disruption*) may lead to enhancement of the luminosity, the accompanying radiation by the particles in the collective fields of the opposite bunch (*beamstrahlung*) may limit the extent to which this enhancement can be utilised, by creating a large spread in the center-of-mass collision energies and producing intolerable levels of backgrounds. An attempt is made to present a simple qualitative introduction to these phenomena.

1. INTRODUCTION

The next generation electron-positron colliders are designed to produce a much higher rate of collisions, or much higher luminosity, than what has been achieved to date.

In case of circular colliders, the demand for higher statistics, higher precision measurements resulted in the designs of the so-called "factories": ϕ, τ-charm and B meson colliders. The required luminosity is one to two orders of magnitude higher than the luminosity attained at those collision energies in the past (e.g. B - meson factories presently under construction at SLAC [1] and KEK [2] are designed for the peak luminosity of $L \sim 3 \cdot 10^{33} \mathrm{cm}^{-2}\mathrm{s}^{-1}$ with the plans to reach the luminosity of $10^{34} \mathrm{cm}^{-2}\mathrm{s}^{-1}$ at the later stage of the projects).

Synchrotron radiation losses limit the collision energies of the storage ring based colliders, as the radiation power rises very fast with particle energy:

$$P_{SR} \propto E^4 \tag{1}$$

On the other hand, linear colliders are considered today the way to higher center-of-mass collision energies [3]. Because the basic point cross section falls as the square of the energy, the required luminosity for the 500 to 1000 GeV c.m. energy electron-positron colliders has to be also of the order of $10^{34}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$, in order to insure reasonable event rates.

1.1 Beam-beam effects in linear colliders

The luminosity of a collider, in terms of the number of particles per bunch N, bunch collision frequency f, and the effective interaction area A at the collision point, can be written as

$$L = \frac{N^2 f}{A} \tag{2}$$

The collision frequency in the linear collider designs is typically much lower than in the circular colliders; so is the bunch population N. Thus, to achieve high luminosity, the transverse beam sizes and the effective interaction area A is planned to be much smaller than in circular colliders (typically > 100 $\mu\mathrm{m}^2$), leading to very high transverse bunch densities.

The beam-beam interaction limits the maximum attainable transverse bunch density. In the context of circular colliders, one speaks of maximum beam-beam tune shift. In the linear collider context the beam-beam interaction is very strong; one speaks of *disruption.*

Particles in one bunch passing through the opposite bunch are subject to very strong transverse focusing forces. The bunches focus each other, resulting in enhancement of the luminosity (the so-called *pinch effect* [4]). When this effect is strong, each bunch emerges from the collision point with a large spread in the angles of the particle trajectories. This *disrupted* bunch has to be guided carefully away from the detector to minimise the backgrounds.

When the beams are offset transversely with respect to each other, each of them is deflected as a whole by the fields of the opposite bunch (one can think of a strong attractive force between two opposite currents). This deflection of the bunch can be measured by the downstream beam position monitors and can be used to center the beams at the collision point. This method has been studied and used successfully at the Stanford Linear Collider (SLC) [5].

The benefit of possible luminosity enhancement is offset by the fact that particles radiate in the strong electromagnetic fields of the opposite bunch. The stronger the fields are and the higher the energy of the particle, the more energy it can lose during collisions. The probability that a given particle will lose a substantial portion of its energy to radiation before colliding with another particle in the opposite beam becomes high. This leads to a large *spread in the center-of-mass collision energies*. Moreover, hard photons emitted during collisions can create pairs of low energy electrons and positrons or muons that can result in unacceptable levels of the *backgrounds* in the detector.

2. FIELDS OF A SINGLE BUNCH

Let us first look at the electromagnetic fields created by a single bunch of charged particles. One of the easiest ways to write down the fields is to consider the bunch in a co-moving inertial frame, where the charges are at rest. The Coulomb and Gauss laws then immediately give us the electric field, and recalling the Lorentz transformations for the fields, it is easy to write the expressions for the electric and magnetic fields in the laboratory frame.

2.1 Lorentz transformations

The component of the electric field in the lab frame parallel to the bunch velocity remains unchanged,

$$\mathbf{E}_{\parallel} = \mathbf{E}'_{\parallel}$$

while the transverse fields transform as follows:

$$\mathbf{E}_{\perp} = \gamma \mathbf{E}'_{\perp} \quad \text{and} \quad \mathbf{B}_{\perp} = \frac{\mathrm{v}}{c^2}\mathbf{E}_{\perp}$$

where the primed quantities are in the rest frame of the bunch.

2.2 Linear charge density

Consider an infinitely long bunch with a linear charge density λ in the lab frame, moving with velocity v (in the rest frame the charge density is γ times smaller). The transverse electric field a distance r away from the bunch can be easily calculated using the Gauss law (here we are using cylindrical system of coordinates (r, θ, s) with s taken along the bunch axis):

$$E_r = \frac{\lambda}{2\pi\varepsilon_0 r}$$

and since the same bunch represents a current $I = v\lambda$, using Ampere's law it is also easy to write down the transverse magnetic field at a distance r away from the bunch:

$$B_\theta = \frac{\mu_0 \lambda}{2\pi r} v = E_r \frac{v}{c^2}$$

2.3 Round beams: uniform distribution

The above expressions are true well outside the bunch, and we are interested in the collective fields of a bunch inside or near to it. Let us first write the fields for a long bunch of length l and total number of particles N. The bunch is a cylinder of radius a with uniform transverse particle distribution. Again, from the Gauss law, we have for the electric fields:

$$E_r = \frac{eN}{2\pi\varepsilon_0 l} \cdot \frac{1}{r} \quad (r > a)$$

$$E_r = \frac{eN}{2\pi\varepsilon_0 l} \cdot \frac{r}{a^2} \quad (r < a)$$

and the magnetic field can be computed from these as in the previous section. The plot of the fields (normalised to their maximum value at the edge of the bunch) as a function of distance away from the bunch axis is shown in the figure below.

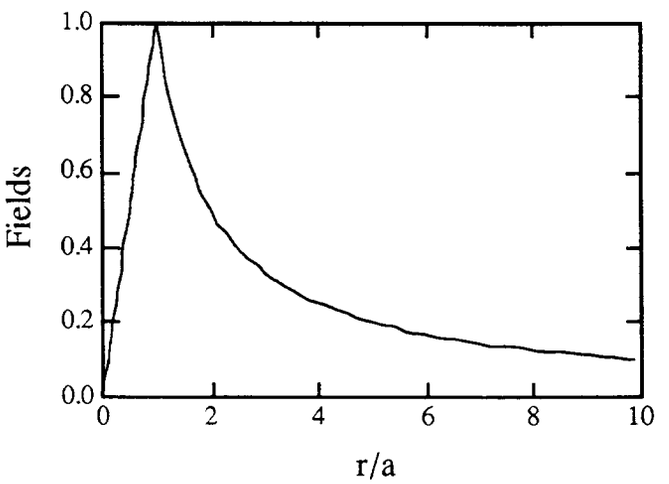


Fig. 1 Electric and magnetic fields of a bunch with uniform charge distribution. The distance from the center of the bunch, r, is in units of the transverse bunch radius.

2.4 Round beams: Gaussian distribution

In the case of a round transverse Gaussian distribution, similar calculation yields the following result:

$$E_r = \frac{eN}{2\pi\varepsilon_0 l\sigma}\left(\frac{1-e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}}{r/\sigma}\right)$$

where σ is the transverse beam size (standard deviation). The bracket in the equation above is plotted in Fig. 2 as a function of distance from the bunch axis in units of beam size.

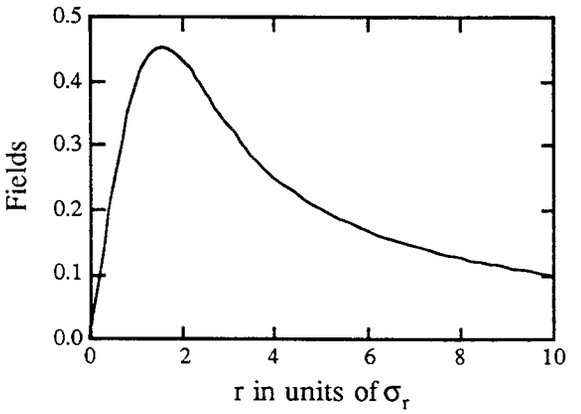


Fig. 2 Electric and magnetic fields of a bunch with Gaussian charge distribution. The distance from the center of the bunch, r, is in units of the transverse beam size.

2.5 Force seen by a test particle

What is the force a particle feels due to the fields that we have just written down? There are two cases to consider: a particle that belongs to the same bunch and a particle belonging to the opposite bunch. In both cases we can simply write down the Lorentz force:

$$\text{Force} = e(\mathbf{E} + [\mathbf{v} \times \mathbf{B}])$$

We can now substitute into the expression above the components of the electric and magnetic fields we have calculated in the previous section. In the case of the co-moving particle, a "fellow traveller", the result is:

$$\text{Force} = eE_r - evB_\theta = eE_r\left(1-\frac{v^2}{c^2}\right) = \frac{1}{\gamma^2}eE_r$$

and one sees the fine cancellation between the force due to the electric and the force due to the magnetic fields that occurs for ultra-relativistic particles; the resulting "space charge" force is suppressed by a very large factor. Thus for high energy particles the space charge force – the force it feels from all the other particles in its own bunch, or the force due to the collective electromagnetic fields of its own bunch – can be completely neglected in most cases (can you think of the circumstances, under which this delicate balance might be destroyed?).

The situation is drastically different for the case of a particle that belongs to the opposite bunch. For the particle moving in the opposite direction the contributions to the force from the electric and magnetic fields add:

$$\text{Force} = eE_r + evB_\theta = eE_r\left(1+\frac{v^2}{c^2}\right) \approx 2eE_r$$

and the resulting force becomes independent of the test particle energy. Using the expression for the electric field generated by a round Gaussian bunch, we have:

$$F = \frac{e^2 N}{\pi\varepsilon_0 l}\frac{1}{r}\left(1-e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}\right)$$

As would be expected, the force vanishes inversely proportional to the distance away from the bunch (for $r >> \sigma$). Moreover, for the test particles passing through the bunch close to its axis, the force is *linear* with r!

$$F = \frac{e^2 N}{2\pi\varepsilon_0 l\sigma^2}\cdot r$$

and it is a *focusing* force.

Of course, the reason we have written down the force for the case of a *round* bunch is because the expressions for the general case are much more complicated [6]. Another comparatively simple case is that of a very flat beam: e.g. when the vertical size much smaller than the horizontal one.

2.6 Beam-beam kick

We are finally in a position to write down the change in a test particle trajectory due to the beam-beam force. The particle transverse momentum is changed by:

$$\Delta p_\perp = \int_{-\infty}^{\infty} F dt$$

and the change in the slope of the particle trajectory is:

$$\Delta r' \equiv \frac{\Delta p_\perp}{P} = -\frac{2e}{\gamma mc}\int_{-\infty}^{\infty} E_r dt = -\frac{2e}{\gamma mc}E_r\cdot\Delta t$$

The time during which the particle feels the beam-beam force is just

$$\Delta t = \frac{l}{2c}$$

since both the test particle and the bunch (of length l) are moving in opposite directions with velocities very close to the speed of light.

For our simple case of the round Gaussian bunch, the beam-beam *kick* experienced by the bunch is then:

$$\Delta r' = -\frac{2Nr_e}{\gamma}\frac{1}{r}\left(1 - e^{-r^2/2\sigma^2}\right) \quad \text{where} \quad r_e \equiv \frac{e^2}{4\pi\varepsilon_0 mc^2}$$

and near the axis the kick is linear with the distance away from the bunch axis:

$$\Delta r' = -\frac{Nr_e}{\gamma\sigma^2}\cdot r$$

In the more general case of a Gaussian transverse beam distribution with unequal transverse beam sizes, near the axis the kick is:

$$\Delta x' = -\frac{2Nr_e}{\gamma\sigma_x\left(\sigma_x + \sigma_y\right)}\cdot x \quad \text{and} \quad \Delta y' = -\frac{2Nr_e}{\gamma\sigma_y\left(\sigma_x + \sigma_y\right)}\cdot y$$

These expressions show that near the axis the bunch looks to a test particle like a lens that focuses in both transverse directions.

3. DISRUPTION

The bunch, then, looks like a focusing lens to a test particle that traverses it in the opposite direction near its axis. It is equivalent to a lens with the focal lengths:

$$\frac{1}{f_{x,y}} = \frac{2Nr_e}{\gamma\sigma_{x,y}\left(\sigma_x + \sigma_y\right)}$$

And in the linear colliders case, a very strong lens indeed. For the Stanford Linear Collider [5] parameters (γ about 100'000, σ about a micron, $5\cdot10^{10}$ particles per bunch), the focal length is of the order of a millimeter!

When dealing with such strong lenses, we have to take into account the fact, that not only the slope of the trajectory of the test particle changes as it crosses the lens, but there is also an appreciable change in the transverse position, or impact parameter. In other words, we can no longer treat this lens in thin approximation (a simple kick to the trajectory and no change in position), but rather take into account the finite bunch length. As is illustrated in Fig. 3, a particle that traverses a bunch of length σ_S and whose trajectory slope changes by $\Delta r'$, also suffers a change in the impact parameter of

$$\Delta r = \Delta r' \cdot \sigma_s$$

and since

$$\Delta r' = -\frac{1}{f}\cdot r \qquad \Rightarrow \qquad \Delta r = -\frac{\sigma_s}{f}\cdot r$$

The relative change in the impact parameter is the quantity that in the context of the linear colliders is called *disruption* [7], that is

$$D = \frac{\Delta r}{r} = -\frac{\sigma_s}{f}$$

It is equal to one when the focal length of the beam-beam lens becomes as short as the bunch length. The disruption is *defined* as follows [8]:

$$D_{x,y} \equiv \frac{2Nr_e}{\gamma} \frac{\sigma_s}{\sigma_{x,y}\left(\sigma_x + \sigma_y\right)}$$

and it coincides with the expression above for the case of the Gaussian bunches:

$$D_{x,y} = -\frac{\sigma_s}{f_{x,y}} \qquad \text{(Gaussian distribution)}$$

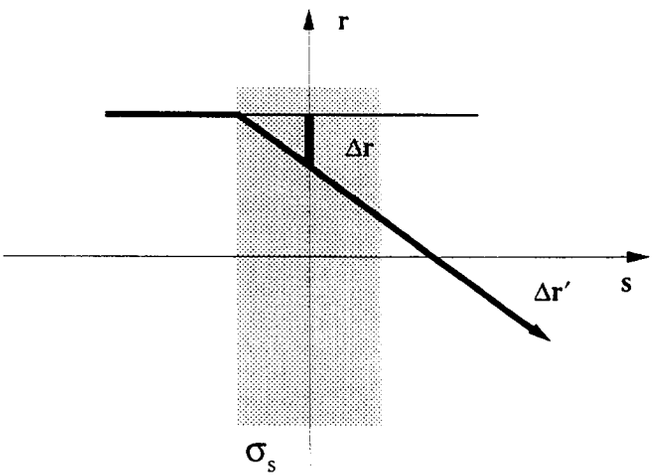


Fig. 3 Change in the test particle trajectory (both slope and impact parameter change) as it traverses the opposite bunch.

3.1 Connection with the storage ring jargon

In the case of a storage ring collider (see the lectures on the beam-beam dynamics by E. Keil in these proceedings [6]), the linear part of the beam-beam force (for the particles with small transverse offsets) can be viewed as an additional focusing lens at the interaction point. Such a gradient error results in a tune shift:

$$\Delta Q = -\frac{1}{4\pi} \cdot \frac{1}{f} \cdot \beta^*$$

where f is the focal length of the lens and β^* is the value of the beta function at the interaction point. The beam-beam limit is characterised in the ring colliders by the maximum achievable *beam -beam tune shift* (also called *beam -beam strength parameter*)

$$\xi_{x,y} \equiv -\frac{1}{4\pi} \cdot \frac{1}{f_{x,y}} \cdot \beta^*_{x,y}$$

where the focal lengths $f_{x,y}$ for a Gaussian bunch, from the previous section, are:

$$\frac{1}{f_{x,y}} = \frac{2Nr_e}{\gamma\sigma_{x,y}\left(\sigma_x + \sigma_y\right)}$$

We can express the focal lengths $f_{x,y}$ in terms of the beam-beam strength parameter, using the above definitions:

$$\frac{1}{f_{x,y}} = -4\pi\xi_{x,y} \cdot \frac{1}{\beta^*_{x,y}}$$

The maximum value of the beam-beam strength parameter achieved in the ring colliders has been limited to below 0.06.

An interesting observation [9] can be made here: a "typical" particle with a transverse offset of one standard deviation σ_r (here r stands for either x or y) will get an angular kick

$$\Delta r' = -\frac{\sigma_r}{f}$$

Comparing it to the spread in angles in the bunch, σ'_r, we can write:

$$\frac{\Delta r'}{\sigma'_r} = -\frac{\sigma_r}{f} \cdot \frac{1}{\sigma'_r} = -\frac{\beta^*}{f} = -4\pi\xi$$

that is, the beam tune shift of $\xi \sim 0.1$ corresponds to the angular kicks or disruption angles starting to exceed σ'_r.

Both quantities, the beam-beam strength parameter and the disruption, characterise the strength of the beam-beam interaction and, for Gaussian bunches, we can write down formally:

$$D_{x,y} = -\frac{\sigma_s}{f_{x,y}} = 4\pi\xi_{x,y} \cdot \frac{\sigma_s}{\beta^*_{x,y}}$$

Maximum achievable disruption may be a relevant design parameter in the ring colliders that operate with values of the beta function at the interaction point that are comparable to the bunch length [10].

3.2 Disruption and luminosity enhancement

During collisions the bunches focus each other, leading to the increase in luminosity. The luminosity enhancement factor, H_D is defined as the ratio of the luminosity under disruption to the geometric luminosity (i.e. when the beam-beam effects are ignored):

$$H_D \equiv \frac{L}{L_0} = \frac{\sigma_{x0}\sigma_{y0}}{\sigma_x \sigma_y}$$

There are very few analytical results on this parameter. Most of the insight is gained with extensive simulations [8].

3.3 Kink instability

Another important effect that has been studied in the simulations is the behaviour of the luminosity as a function of transverse beams offset at the collision point. When the beam-beam effect is weak, the luminosity is degraded simply due to the reduced overlap between the beams. In the strong beam-beam force regime an offset can lead to the development of the so-called kink instability. As a result, the luminosity enhancement as a function of disruption saturates [8]. Simulations have been used to estimate the tolerance with which the beams have to be kept centered with respect to each other. For example, study of the flat beams collisions indicated that for offsets

$$\frac{\Delta y}{\sigma_y} \leq D_y$$

simulations show no significant reduction in luminosity. In fact, for small transverse offsets, the beam-beam force has a beneficial effect on the luminosity, pulling the beams towards each other.

The luminosity is degraded when the beams are offset in transverse direction, e.g. if the beams are offset vertically and the beam-beam effects are weak [11]:

$$L \propto \exp\left[-\left(\frac{\delta y}{2\sigma_y}\right)^2\right]$$

An example of simulation taking into account the beam-beam effect is shown in the Fig. 4 below, and demonstrates the expected gain in luminosity from the pinch effect in one of the designs of a future linear collider final focus system [12].

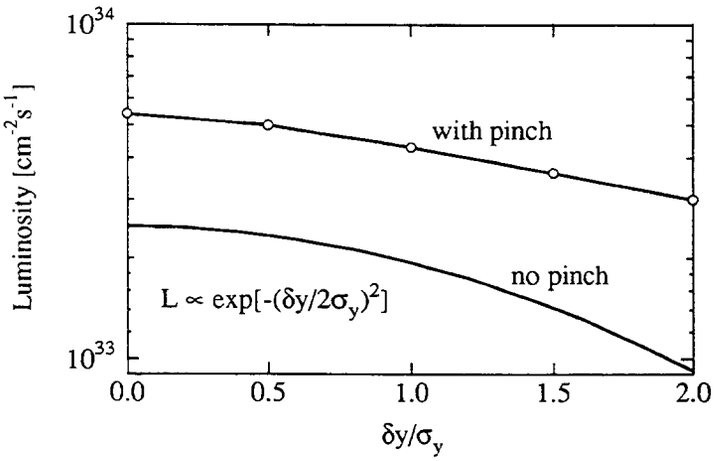


Fig. 4 Luminosity vs. vertical offset. The lower curve is the analytical prediction, the upper data is the result of simulations, taking into account the beam-beam interaction.

3.4 Beam-beam deflections

One of the most important successes of the Stanford Linear Collider is the development of the use of the beam-beam deflections as a tool for steering the micron-size beams into collision, for maintaining collisions and for monitoring and tuning such beam characteristics as the beam transverse sizes at the interaction point and beams overlap.

The geometry of the beam-beam deflections is shown in Fig. 5 and in Fig. 6 the beam-beam deflection angle is plotted as a function of the beams offset. In the formula shown, the offset in the exponent is normalised by the effective beam spot size at the collisions point (the two beam sizes added in quadrature), that can be recovered from the measurements.

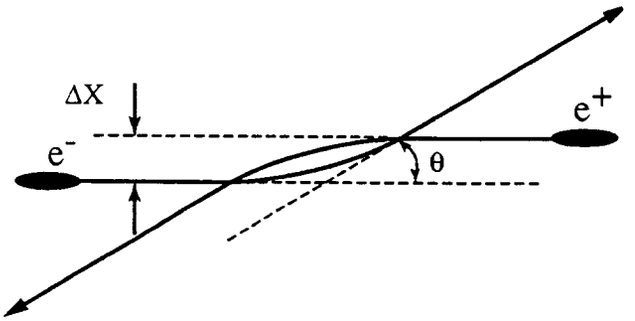


Fig. 5 Geometry of the beam-beam deflections

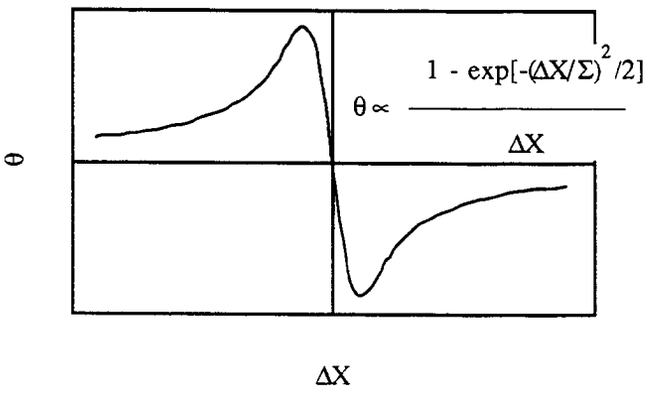


Fig. 6 Beam-beam deflection angle as a function of the transverse beams offset

4. BEAMSTRAHLUNG

During collisions the particles experience transverse acceleration by the collective fields of the opposite bunch (as we have seen, these fields, at the edge of the bunch, can reach very high values indeed). As a result, particles radiate, and this radiation has been called "beamstrahlung", by analogy with the term bremsstrahlung.

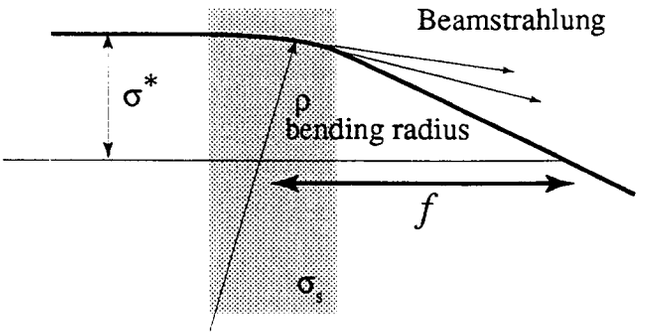


Fig. 7 Trajectory of a typical particle incident on the opposite bunch is bent during the passage. The particle radiates beamstrahlung photons in the process.

In Fig. 7 is shown a trajectory of a typical incident particle as it traverses the opposite bunch off-axis. The particle's trajectory is bent towards the axis by the focusing action of the opposite bunch and, taking as an example the case of the Stanford Linear Collider, the magnetic field that the particle sees can reach $B \sim 200$ Tesla! Beamstrahlung was first observed at SLC and was used as a tool to optimise the luminosity, as illustrated in Fig. 8. As the vertical offset between the beams was changed, the flux of beamstrahlung photons was measured downstream of the collision point. The maximum flux was observed when the beams were separated on the order of one standard deviation, or when the majority of the particles in one beam saw the maximum field of the opposite bunch. The dip in the photon flux corresponds to the beams being centered, as confirmed by the beam deflection signal in the top part of the graph.

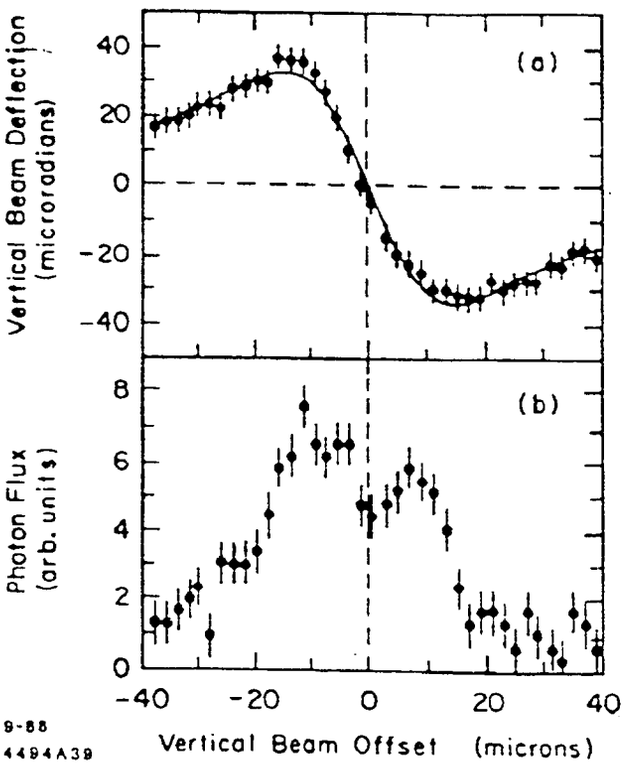


Fig. 8 Beamstrahlung and beam-beam deflection signals. Experimental data from SLC. The dip in the middle of the beamstrahlung signal corresponds to minimum beams offset.

It is useful to think of this radiation as synchrotron radiation in the collective electromagnetic fields of the bunch. Under certain conditions, the properties of beamstrahlung can indeed be estimated using the classical synchrotron radiation formulae. But in the context of the designs of the future linear colliders, with particle energies and transverse particle densities reaching very high values, this approximation breaks down and the quantum nature of this radiation has to be taken into account fully.

4.1 Qualitative picture of synchrotron radiation

Let us briefly recall the main features of synchrotron radiation. Consider a charged particle that is bent by a uniform magnetic field (e.g. in a dipole magnet) and is following a trajectory with a bending radius ρ. Because of the Lorentz boost, the radiation is thrown forward and is contained in a very narrow cone around the tangent to the trajectory. The typical opening angle of the cone is

$$\theta \approx \frac{1}{\gamma}$$

where γ is the Lorentz factor. An observer, looking along the tangent to the trajectory, sees radiation that is emitted from a very short length of the trajectory:

$$l_R \approx \frac{\rho}{\gamma}$$

The time interval between the first photons emitted by the particle as it enters this short piece of the orbit, and the last photons that it emits as it is disappearing from the observer's view, is equal to the difference in time it takes the particle to traverse this length and the time it takes the photons to cross this part of the trajectory. The typically ultra relativistic particles, travelling with the speed very close to the speed of light, stay very close behind the photons they emit, thus the time duration of the resulting radiation pulse is very short:

$$\Delta t = \frac{\rho}{\gamma}\cdot\frac{1}{\beta c} - \frac{\rho}{\gamma}\cdot\frac{1}{c} = \frac{\rho}{\gamma\beta c}(1-\beta)$$

and recalling the definition of the Lorentz factor,

$$1-\beta = \frac{1-\beta^2}{1+\beta} \equiv \frac{1}{2\gamma^2}$$

the typical frequency of the emitted radiation is

$$\omega \sim \frac{1}{\Delta t} \sim 2\frac{c\gamma^3}{\rho}$$

The observer sees the radiation coming in very short flashes separated by the revolution time. In fact, the spectrum of the synchrotron radiation contains frequencies that are harmonics of the revolution frequency ω_0. It extends all the way to the high frequencies of the order of γ^3 times the revolution frequency and decays exponentially above these frequencies. Typically one speaks of the critical frequency or critical photon energy that is chosen such that half of the power is emitted below this photon energy:

$$\varepsilon_c = \hbar\omega_c = \frac{3}{2}\frac{\hbar c}{\rho}\gamma^3$$

or, in "practical" units,

$$\varepsilon_c[\mathrm{GeV}] = 0.664 E^2[\mathrm{TeV}] \cdot B[\mathrm{Tesla}]$$

4.2 Beamstrahlung parameter Υ

The above expression for the typical photon energy is only an approximation that fails when we consider the case of very high particle energies and very high fields. For the typical parameters of the future linear collider designs it can result in photon energies that exceed the initial particle energy. The description of beamstrahlung [8] that takes into account the quantum nature of the process of emission (in particular, the recoil of the particle when it emits a photon) characterises the spectrum and intensity of the radiation using a new parameter Υ:

$$\Upsilon \equiv \frac{2}{3}\frac{\varepsilon_c}{E} = \gamma\frac{\langle B\rangle}{B_c}$$

where $\langle B\rangle$ is the average electromagnetic field seen by the particle while traversing the opposite bunch. It has contributions from both the electric and the magnetic fields of the bunch. The constant B_c, called the Schwinger critical field strength, is:

$$B_c = \frac{m_e^2 c^2}{e\hbar} = 4.42 \cdot 10^9 \text{ Tesla}$$

The average field, calculated for a Gaussian bunch, results in the following expression for the beamstrahlung parameter:

$$\Upsilon_{\mathrm{mean}} \approx \frac{5}{6}\frac{N r_e^2 \gamma}{\alpha\sigma_s\left(\sigma_x + \sigma_y\right)}$$

The beamstrahlung parameter is not constant during collision, in particular, with strong disruption, due to pinch effect, the beam sizes change and it can become much higher than the value calculated for the parameters of the bunch just before collision.

4.3 Beamstrahlung spectrum

The expressions for the energy of the emitted photons and for average intensity of the radiation now acquire additional factors dependent on the beamstrahlung parameter that regain "sanity" and no longer violate energy conservation. For example, the typical photon energy approaches the particle energy without exceeding it. Thus high values of Υ indicate that the spectrum consists mostly of very hard photons. The spectrum changes from the classical synchrotron radiation spectrum, as illustrated in Fig. 9, towards almost triangular shape, dropping sharply at the particle energy.

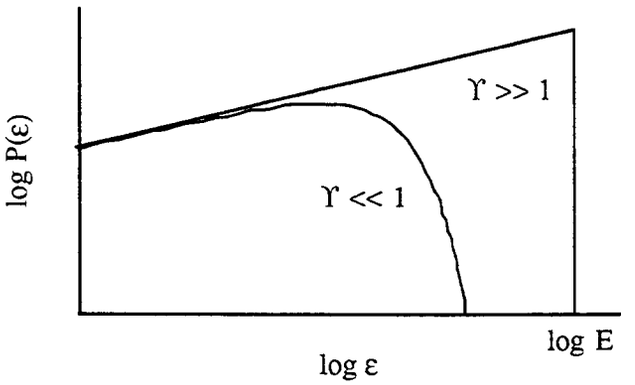


Fig. 9 Beamstrahlung power vs. photon energy (log-log scale)

Average photon energy in the quantum regime limit is about a quarter of the particle energy:

$$\frac{\langle\varepsilon\rangle}{E} \approx \frac{1}{2}\Upsilon \quad \text{when } \Upsilon << 1$$

$$\frac{\langle\varepsilon\rangle}{E} \approx \frac{1}{4} \quad \text{when } \Upsilon >> 1$$

An important characteristic of the emitted radiation is the average number of photons emitted, it serves as a measure of backgrounds in the detector. The average number of photons per crossing [13], emitted by a single particle, can be written approximately as:

$$n_\gamma \approx \frac{5}{2}\frac{\alpha^2\sigma_s}{r_e\gamma}\Upsilon\cdot\frac{1}{\left(1+\Upsilon^{2/3}\right)^{1/2}}$$

The expression is valid for $\Upsilon < 5$.

4.5 Collision energy spread

With a large amount of beamstrahlung the energy spread of the beams increases. Particles that have radiated are no longer at the design energy, resulting in the increased spread in the center-of-mass collision energies.

In quantum regime, the probability to emit very hard photons is high, and the energy distribution in the bunch becomes rather complicated, bearing little resemblance to a Gaussian distribution. Thus it is not very useful to characterise the distribution with its second moment (standard deviation) and energy distribution is obtained from the simulations.

High disruption results in the luminosity enhancement, but due to accompanying high beamstrahlung, the enhancement occurs mainly for the collisions with lower than the design center-of-mass energies.

4.6 Backgrounds

Beamstrahlung photons are a source of backgrounds in the detector. One of the most important processes is *pair creation.* A photon can convert into low energy pairs (e^+e^-, $\mu^+\mu^-$) in close encounters with individual particles of the opposite bunch, this is called *incoherent pair*

creation. There is also a possibility of *coherent pair creation* by a photon that converts in the collective electromagnetic field of the oncoming bunch. One of the particles in a pair is *defocused* by the opposite bunch and scatters into the detector.

One of the possible ways to try to reduce the backgrounds due to beamstrahlung is to use *flat beams*, i.e. when $\sigma_y << \sigma_x$. This keeps the beamstrahlung parameter low, since

$$\Upsilon \propto \frac{1}{\sigma_x + \sigma_y}$$

also helping with the collisions center-of-mass energy spread.

5. CONCLUSIONS

The above considerations of the phenomena of disruption and beamstrahlung may help the reader to understand qualitatively the following implications for future linear collider designs.

In order to avoid excessive backgrounds and collision energy spread due to large beamstrahlung and to delay the onset of the kink instability (instability to small transverse offsets of the colliding bunches) disruption should be limited to moderate values. The results of the simulations then tend to predict limited luminosity enhancement: less than a factor of two in the case of flat bunches and maybe up to a factor of six for the round beams.

Present linear collider designs tend to schemes where trains of small bunches are collided, limiting the luminosity per collision and compensating for it by having many bunches per train.

ACKNOWLEDGEMENTS

Many thanks to Pisin Chen, Albert Hofmann, Werner Joho, Olivier Napoli and Bruno Zotter for materials and stimulating discussions that were invaluable in the preparation of these notes.

REFERENCES

[1] PEP-II An Asymmetric B Factory, Conceptual Design Report, SLAC-418 (1993).

[2] B-Physics Task Force, Accelerator Design of the KEK B-Factory, KEK Report 90-24 (1991).

[3] R.B. Palmer, Prospects for High-energy e^+e^- Linear Colliders, Ann. Rev. Nucl. Part. Sci. 40:529-592, 1990.

[4] R. Holebeek, NIM, **184** (1984) 333.

[5] SLC Conceptual Design Report, SLAC-229 (1982).

[6] E. Keil, Beam-beam Dynamics, These proceedings.

[7] J.E. Augustin, et al., Proceedings of the Workshop on Possibilities and Limitations of Accelerators and Detectors, Batavia (1978).

[8] K. Yokoya and P. Chen, in Frontiers of Particle Beams: Intensity Limitations, Lecture Notes in Physics **400**, Springer-Verlag, 1992.

[9] K.L. Brown, Luminosity and Tune Shift in e^+e^- Storage Rings, SLAC-PUB-4366 (1987).

[10] S. Milton, Evidence of a Disruption Limit in Storage Rings, Preprint PSI PR-90-05 (1990)
S. Milton and L. Rivkin, Proc. EPAC90, Nice (1990), p.255.

[11] O. Napoly, The Luminosity for Beam Distributions with Errors and Wakefield Effects in Linear Colliders, CERN SL/92-34 (AP), CLIC Note 173 (1992).

[12] O. Napoly, A Large Aperture Final Focus System for TESLA, Proc. IEEE Part. Acc. Conf., Washington DC (1993).

[13] P. Chen, Beamstrahlung and the QED, QCD Backgrounds in Linear Colliders, SLAC-PUB-5914 (1992).

SINGLE AND MULTIPLE TOUSCHEK EFFECTS

J. Le Duff
Laboratoire de l'Accélérateur Linéaire, Orsay, France

1. INTRODUCTION

Particles in a circulating bunch execute transverse betatron oscillations around the equilibrium orbit. Since the transverse velocities are statistically distributed, these particles can be scattered by collisions so transferring transverse momenta into longitudinal momenta. Such a change in the direction of the particle's momentum can lead to a strong variation of its energy, due to a relativistic effect; if the relative energy variation exceeds the energy acceptance of the machine the particle is lost.

The effect was first recognized by B. Touschek (1963) on the small e^+e^- storage ring ADA, by observation of the stored bunch lifetime. The theoretical model was first worked out be C. Bernadini, B. Touschek and J. Haissinski [1-4] and later on by some others [5,6].

For low and medium energy electron storage rings, operating with flat, dense beams, only the horizontal betatron motion produces sufficiently high relative transverse velocities for the energy variations to lead to particle loss. The theory, however, did not predict accurately experimental lifetime observations in the lowest part of the ADA energy range where measured values were much better than those calculated. H. Bruck first pointed out that this could be due to an anomalous increase in the transverse beam dimensions. A model was worked out [7,8] showing that smaller energy transfers, not contributing directly to particle losses, were responsible for the increase. These small transfers, having a high probability, act like an additional noise on the quantum radiation and, as a consequence, the balance from the damping of betatron oscillations leads to a new steady state.

Both the single and multiple Touschek effects appear in H. Bruck's book [9]. The following presentation follows essentially the original work with, however, some extension to the case of strong focusing machines. In 1974 the multiple effect theory was generalized by A. Piwinski [10] to take proton storage rings into account and was renamed 'Intrabeam Scattering' [11,12].

For a long time the Touschek effects were almost forgotten since the operating conditions of e^+e^- circular colliders, with large emittances to minimize beam-beam effects, made them negligible. More recently, however, they came back into evidence with the construction and operation of very small emittance electron (or positron) storage rings dedicated to synchrotron radiation physics.

2. SINGLE TOUSCHEK EFFECT

2.1 A rough estimate of the effect

Particles in a bunch execute transverse betatron oscillations around the equilibrium orbit (Fig. 1a). In the moving frame of the bunch the motion becomes purely transverse, neglecting the slow synchrotron motion. Coulomb scattering will occur for particles having different transverse velocities and will result in an energy transfer from the transverse plane to the longitudinal direction (Fig. 1b).

A particle with horizontal betatron amplitude $\hat{x}$ will have a maximum horizontal velocity

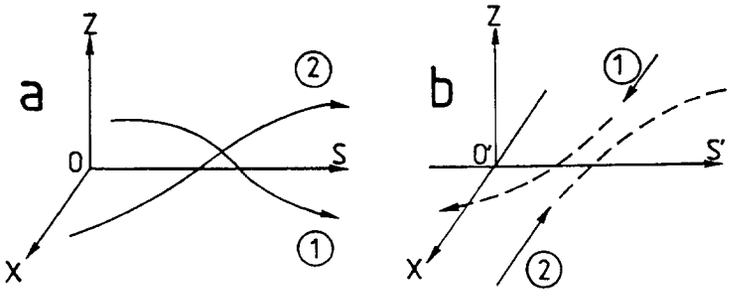


Fig. 1 Coulomb scattering of two particles inside the bunch. (a) Laboratory system, (b) Centre-of-mass system.

$$x' = \frac{\hat{x}}{\beta_x} = \frac{\hat{x}}{\lambda\!\!\!^{-}_x} = \frac{p_x}{p}$$

where $\lambda\!\!\!^{-}_x$ is the betatron wavelength and βx the envelope function. It corresponds to a transverse momentum $p_x = p\hat{x} / \beta_x$. Consider a typical machine with $\hat{x} \approx 1.10^{-4}$ m and $\beta_x \approx$ 10 m with an operating energy of 800 MeV. The transverse momentum can be as much as 8 keV. When transferred into the longitudinal direction it becomes $\Delta E = \gamma p_x = 12.5$ MeV or $\Delta E / E = 1.56\%$ which obviously is of the order of magnitude of a typical energy acceptance (e.g. RF acceptance). Since the particles have a Gaussian transverse energy distribution, events with higher energy transfers can occur.

Generally, storage rings operate with flat beams such that the vertical size is at least an order of magnitude smaller than the horizontal one. Hence the vertical betatron motion will contribute much less to the losses.

2.2 Total cross section for particle losses

It is usual to consider the Coulomb scattering of two particles in their centre-of-mass (c.m.) system in which the particles will have equal but opposite momenta. After the collision each particle is scattered by an angle θ as shown in Fig. 2 for one particle. The number of particles dN which are scattered into the differential solid angle $d\Omega$ is generally represented by the formula

$$dN = d\sigma \; N \; n \; dx$$

where N is the number of test particles with velocity $v = dx/dt$ hitting the target of density n, and where $d\sigma$ is the differential cross section for the physical process: $d\sigma = \sigma \, d\Omega$. Integration over all values of interest for the solid angle would give the corresponding total cross section σ_T. The rate of events for such a physical process is then

$$\frac{dN}{dt} = \sigma_T \; v \; N \; n$$

where $N = n \, dV$ since the test beam and the target are common.

In the present case, the Coulomb scattering differential cross section for electrons (or positrons) is given by the Möller formula (non-relativistic case)

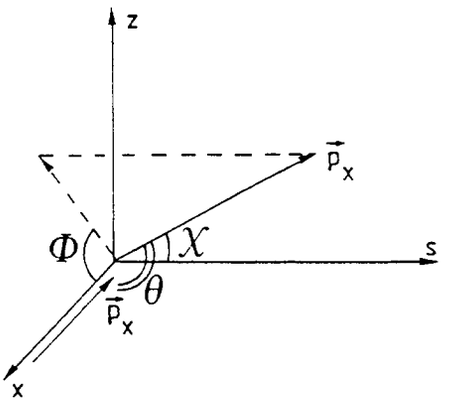


Fig. 2 Coulomb scattering in the c.m. system

$$\frac{d\sigma}{d\Omega} = \frac{4r_0^2}{(v/c)^4}\left[\frac{4}{\sin^4\theta} - \frac{3}{\sin^2\theta}\right]$$

where v is the relative velocity in the c.m. system.

The momentum transfer into the longitudinal direction is

$$\Delta p = p_x|\cos\ \chi| \ ,$$

$2p_x$ being the relative transverse momentum. The particle is lost if $\gamma\Delta p \geq \varepsilon_{RF}$ where ε_{RF} is the momentum acceptance of the RF (assuming this is the limiting acceptance). Hence the total cross section for particle losses is obtained by integrating the differential cross section over angles such that:

$$|\cos\ \chi| \geq \varepsilon_{RF} / \gamma p_x \ .$$

Putting $\mu = \varepsilon_{RF} / \gamma p_x$ one obtains

$$\sigma_T = \int_{|\cos\ \chi| \geq \mu} d\sigma$$

and since $\cos\ \theta = \sin\ \chi \cos\ \phi$

$$\sigma_T = \frac{4r_0^2}{(v/c)^4} \int_0^{\arccos\mu} \sin\ \chi\ d\chi \cdot 2 \int_0^{\pi} d\phi \left[\frac{4}{\left(1-\sin^2\chi\ \cos^2\phi\right)^2} - \frac{3}{\left(1-\sin^2\chi\ \cos^2\phi\right)}\right] .$$

The first integral is bounded to zero and arc cos μ knowing that the final state is the same for the two particles. Performing the integration leads to:

$$\sigma_T = \frac{8\pi r_0^2}{(v/c)^4}\left[\frac{1}{\mu^2} - 1 + \ln\ \mu\right]$$

remembering that this formula is valid in the c.m. system.

2.3 Touschek lifetime

The total rate of collisions leading to losses can be written as

$$\frac{dN}{dt} = \int_V \frac{\sigma_T}{\gamma^2} \; v \; n^2 \; dV$$

where V is the bunch volume. The factor γ^2 takes account of the Lorentz transformation of the product $\sigma_T v$ from the c.m. to the laboratory system.

Since we are considering only effects which take place in the horizontal plane the integration is automatically performed in the vertical and longitudinal planes, hence:

$$\frac{dN}{dt} = \frac{N^2}{\gamma^2} \frac{1}{4\pi\sigma_z\sigma_s} \int_V \sigma_T v \rho\left(x_1, x_1'\right)\rho\left(x_2, x_2'\right) dV$$

where

$$\rho(x,x') = \frac{\beta_x}{2\pi\sigma_x^2} e^{-\left[x^2+\left(\beta_x x' - 1/2\beta_x' x\right)^2\right]/2\sigma_x^2}$$

is the horizontal phase space distribution function in a strong focusing machine, and where $x_1 = x_2$ in order for the collision to take place. Note that the contribution to the horizontal transverse velocity from the dispersion function has been neglected.

In the laboratory system the relative velocity is simply:

$$\left(\frac{v}{c}\right)_{\text{lab}} = x_2' - x_1' \; .$$

The integral now becomes

$$\frac{dN}{dt} = \frac{N^2}{\gamma^2} \frac{\beta_x^2}{16\pi^3 \; \sigma_x^4\sigma_z\sigma_s} \int_{-\infty}^{+\infty} \sigma_T(v) \; v \; e^{-\left(Ax_1^2 + Bx_1 + C\right)/2\sigma_x^2} dx_1 dx_1' dx_2'$$

where

$$A = 2 + \frac{1}{2}\beta_x'^2 \; , \quad B = -\beta_x\beta_x'\left(x_1' + x_2'\right) \text{ and } C = \beta_x^2\left(x_1'^2 + x_2'^2\right) .$$

Since σ_T and v do not depend on the position x_1 the integration with respect to that variable is straightforward

$$\frac{dN}{dt} = \frac{N^2}{\gamma^2} \frac{\beta_x^2\sqrt{2\pi}}{16\pi^3 \; \sigma_x^3\sigma_z\sigma_s\sqrt{A}} \int_{-\infty}^{+\infty} \sigma_T(v) \; v \; e^{\frac{\beta_x^2\beta_x'^2}{8A\sigma_x^2}\left(x_1' + x_2'\right)^2 - \frac{\beta_x^2}{2\sigma_x^2}\left(x_1'^2 + x_2'^2\right)} dx_1' dx_2' \; .$$

If we now make the following change to the variables:

$$u_1 = \dot{x_2}$$
$$u_2 = \dot{x_2} - \dot{x_1}$$

such that $\sigma_T(v)v$ will only depend on u_2 and the integration on u_1 is simplified, then

$$\frac{dN}{dt} = \frac{N^2}{\gamma^2}\frac{\beta_x}{16\pi^2\sigma_x^2\sigma_z\sigma_s}\int \sigma_T(v)\ v\ e^{-\frac{1}{2\sigma_x^2}\frac{\beta_x^2}{2}u_2^2}du_2\ .$$

Note that $u_2 = (v/c)_{\text{lab}}$; hence one has

$$u_2 = \gamma^{-1}v/c$$

where v is the relative velocity in the c.m. system. One can then write

$$\left(\frac{dN}{dt}\right)_{\text{loss}} = 2\left(\frac{dN}{dt}\right)_{\text{collision}} = 2\frac{N^2}{\gamma^2}\frac{\beta_x}{16\pi^2\sigma_x^2\sigma_z\sigma_s}\int \sigma_T(v)v\ e^{-\frac{\beta_x^2}{4\sigma_x^2\gamma^2}\left(\frac{v}{c}\right)^2}\frac{1}{\gamma}d\left(\frac{v}{c}\right)$$

But we can also introduce the individual particle momentum p_x in the c.m.

$$p_x = \frac{|\vec{p}_1 - \vec{p}_2|}{2} = \frac{m_0 v}{2}$$

where m_0 is the rest mass. The loss rate then becomes

$$\left(\frac{dN}{dt}\right)_{\text{loss}} = \frac{N^2}{\gamma^2}\frac{\beta_x}{8\pi^2\sigma_x^2\sigma_z\sigma_s}\int_{\varepsilon_{RF}/\gamma}^{\infty}\sigma_T(p_x)\frac{2p_x}{m_0}\ e^{-\left\{\frac{\beta_x^2 p_x^2}{\sigma_x^2\gamma^2 m_0^2c^2}\right\}}\frac{2}{\gamma m_0 c}dp_x$$

where the first bound means that particles can be lost only if their initial momentum exceeds the momentum acceptance in the c.m.

Using the value $\sigma_T(v)v$ as calculated above and introducing the quantity:

$$\sigma_p = \frac{\gamma m_0 c\sigma_x}{\beta_x}$$

which is known to be the r.m.s. of individual transverse particle momenta, one obtains

$$\left(\frac{dN}{dt}\right)_{\text{loss}} = \frac{N^2}{\gamma^3}\frac{\beta_x r_0^2 c^3 m_0^2}{4\pi\sigma_x^2\sigma_z\sigma_s}\int_{\varepsilon_{RF}/\gamma}^{\infty}\frac{1}{p_x^3}\left[\frac{\gamma^2 p_x^2}{\varepsilon_{RF}^2} - 1 + \ln\frac{\varepsilon_{RF}}{\gamma p_x}\right]e^{-p_x^2/\sigma_p^2}dp_x$$

which can be written in the following form, well known in the literature

$$\frac{1}{N}\frac{dN}{dt} = \frac{1}{\tau} = \frac{Nr_0^2 c}{8\pi\sigma_x\sigma_z\sigma_s}\frac{\lambda^3}{\gamma^2}D(\xi)$$

where τ is the lifetime and with the parameter definition:

$$\lambda^{-1} = (\Delta E / E)_{RF} = \varepsilon_{RF} / \gamma \ m_0 c$$

$$\xi = \left(\varepsilon_{RF} / \gamma \ \sigma_p\right)^2$$

and $D(\xi)$ being the universal function

$$D(\xi) = \sqrt{\xi}\left\{-\frac{3}{2}e^{-\xi} + \frac{\xi}{2}\int_{\xi}^{\infty}\frac{\ln u \ e^{-u}}{u}du + \frac{1}{2}(3\xi - \xi \ \ln \xi + 2)\int_{\xi}^{\infty}\frac{e^{-u}}{u}du\right\}$$

for which the plot is shown in Fig. 3. Note that this function presents a flat maximum equal to 0.3 in the ξ range generally of interest.

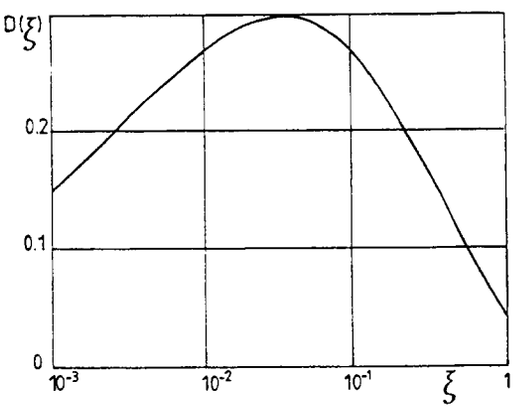


Fig. 3 Plot of $D(\xi)$

The formula shows a dependence of the parameters with the azimuth s. This is true for the bunch dimensions but it can also be true for λ if the vacuum chamber constitutes the real energy acceptance (this can happen for instance when energy transfer occurs in a region of high dispersion). The average effect over the whole circumference is

$$\frac{1}{\tau} = \frac{1}{2\pi R}\int\frac{1}{\tau(s)}ds = \left\langle\frac{1}{\tau(s)}\right\rangle .$$

Up to now we have neglected the influence of the dispersion function on the parameters of the Coulomb scattering. A more precise study would include the assumption that the overall phase space coordinates of the two particles are:

$$\text{particle 1} \quad \left[x_1, x_1', D_x\varepsilon_1, D_x'\varepsilon_1\right]$$

$$\text{particle 2} \quad \left[x_2, x_2', D_x\varepsilon_2, D_x'\varepsilon_2\right]$$

where $\varepsilon = \Delta E / E$ is the relative energy deviation, and D_x the horizontal dispersion function.

Two particles having the same transverse position are now defined as

$$x_2 = x_1 + D_x(\varepsilon_1 - \varepsilon_2)$$

and the relative velocity in the laboratory system becomes

$$\left(\frac{v}{c}\right)_{\text{lab}} = \left(x_2' + D_x'\varepsilon_2\right) - \left(x_1' + D_x'\varepsilon_1\right) .$$

All the integrations performed under these more general assumptions, lead to the formula:

$$\frac{1}{\tau} = \frac{N r_0^2 c}{8\pi \ \sigma_z \sigma_s \sigma_{x\beta} \sqrt{1 + \sigma_{xs}^2 / \sigma_{x\beta}^2}} \frac{\lambda^3}{\gamma^2} D_x(\xi)$$

where $\sigma_{x\beta}$ and σ_{xs} are respectively the r.m.s. horizontal contributions from betatron and synchrotron motions

$$\sigma_x^2 = \sigma_{x\beta}^2 + \sigma_{xs}^2 .$$

The other parameters have the same meaning as before, but with the more general formula for the r.m.s. relative momentum distribution

$$\sigma_p = \frac{\gamma m_0 c}{2\sqrt{C_1}}$$

with:

$$C_1 = \frac{\beta_x^2}{4\sigma_{x\beta}^2} - \frac{B_1^2}{4A_1}$$

$$A_1 = \frac{1}{4\sigma_\varepsilon^2} + \frac{1}{4\sigma_{x\beta}^2}\left[D_x^2 + \left(\beta_x D_x' - \beta_x' D_x / 2\right)^2\right]$$

$$B_1 = \frac{\beta_x}{2\sigma_{x\beta}^2}\left[\beta_x D_x' - \beta_x' D_x / 2\right]$$

$$\sigma_{xs}^2 = D_x^2 \sigma_\varepsilon^2$$

where σ_ε is the r.m.s. of the relative energy spread.

The derivatives of the optical functions only appear in the argument ξ. Since the function $D(\xi)$ has a flat maximum they will not contribute very much to the lifetime. Finally the correction from the dispersion can only increase the lifetime, at least if the vacuum chamber is not the momentum aperture.

3. MULTIPLE TOUSCHEK EFFECT

3.1 Basic idea

Let us now examine the smaller scattering angles which are insufficient to produce particle losses but which nevertheless disturb the particle statistical distribution (noise source).

A first obvious consequence of the Coulomb collisions is a statistical change of the particle energies in the longitudinal phase space, leading to an increase of the energy spread.

A spontaneous energy deviation induces an energy oscillation around the nominal energy (synchrotron motion) damped by the synchrotron radiation. It is well known that a set of harmonic oscillators excited by random fluctuations, with r.m.s. $\langle \delta E^2 \rangle$ at the rate $\dot{N}$ will have a Gaussian distribution with standard deviation

$$\sigma_E^2 = \frac{1}{4} \dot{N} \langle \delta E^2 \rangle \tau_E$$

where τ_E is the damping time for the synchrotron motion.

A rapid change in the particle energy will give, in addition, a jump δx of the radial particle position ($\delta x = D_x \delta E/E$), according to the finite dispersion function D_x. This leads to an induced betatron motion which is then also damped by the synchrotron radiation. The statistical result is a new steady state for the particle distribution in the horizontal plane of motion

$$\sigma_x^2 = \frac{1}{4} \dot{N} \langle \delta x^2 \rangle \tau_x \ .$$

These r.m.s. values from multiple Coulomb scattering must be added quadratically to r.m.s. values related to other random effects (such as the quantum radiation) as soon as the noise sources appear to be independent

$$\sigma_{\text{total}}^2 = \sum \sigma_{\text{partial}}^2 \ .$$

3.2 Small-angle Coulomb scattering

At small angles the Möller formula for the differential cross section reduces to

$$\frac{d\sigma}{d\Omega} = \frac{16 r_0^2}{(v/c)^4} \frac{1}{\sin^4 \theta}$$

where v is the relative velocity in the c.m. system that we will assume to be essentially horizontal. The angular change of the momentum gives a momentum component perpendicular to the horizontal axis (Fig. 2)

$$p_\perp = p_x \sin\theta \qquad dp_\perp = p_x \cos\theta d\theta \approx -p_x d\theta$$

with

$$p_x = \frac{m_0 v}{2}$$

where m_0 is the rest mass of the electron since we still assume non-relativistic transverse motions.

Hence

$$\frac{d\sigma}{d\Omega} = \frac{r_0^2 m_0^4 c^4}{p_\perp^4}$$

and since

$$d\Omega = -2\pi \sin\theta d\theta$$

$$d\sigma = 2\pi \left[\frac{r_0 m_0^2 c^2}{p_x} \right]^2 \frac{1}{p_\perp^3} dp_\perp \ .$$

This shows that the probability is higher for small perpendicular momentum transfers, and that the noise will mostly come from these small angle collisions.

Note that

$$p_\perp = \frac{2 r_0 p_0^2}{b \ q}$$

where b is the impact parameter, $p_0 = m_0 c$, and $q = 2p_x$ the relative momentum.

3.3 Energy spread due to multiple scattering

The quantity of interest is, in the laboratory system:

$$\dot{N}\left\langle \delta E^2 \right\rangle = c^2 \dot{N}\left\langle \delta p_s^2 \right\rangle$$

which, from the Lorentz transformation is equivalent to:

$$\dot{N}\left\langle \delta E^2 \right\rangle = \frac{1}{2} c^2 \gamma \left\{ \dot{N}\left\langle \delta p_\perp^2 \right\rangle \right\}_{c.m.}$$

where the factor 1/2 takes account of the fact that the probability is the same for transfers occurring in the vertical and longitudinal directions.

Considering one test particle, the rate of events from multiple scattering with the rest of the bunch is

$$\dot{N} = \sigma \ v \ n$$

and since

$$n_{c.m.} = \gamma^{-1} n_{\text{lab}}$$

one can write

$$\dot{N}\left\langle \delta E^2 \right\rangle = \frac{1}{2} c^2 n \left\{ \sigma \ v \left\langle \delta p_\perp^2 \right\rangle \right\}_{c.m.} \ .$$

The mean quadratic transferred momentum can be obtained from the differential cross section:

$$\left\langle \delta p_\perp^2 \right\rangle = \frac{\int p_\perp^2 d\sigma}{\int d\sigma} = \frac{1}{\sigma} \int p_\perp^2 d\sigma$$

where the integral is performed over all transfers corresponding to a given relative velocity in the c.m.

$$\sigma\left\langle \delta p_{\perp}^{2} \right\rangle = 2\pi\left(r_0 m_0^2 c^2\right)^2 \frac{1}{p_x^2} \ln\left[\frac{p_{\perp \max}}{p_{\perp \min}}\right]$$

and where

$$p_{\perp \max} = p_x \ .$$

The minimum momentum transfer is related to the maximum distance which can occur between two particles. In the original model this was taken as the beam half-height. However, at long distances, the interaction between two particles is disturbed by the presence of the other particles (collective effects). Hence a more realistic maximum distance, or maximum impact parameter for free binary collisions would be:

$$b_{\max} \cong n^{-1/3}$$

and the corresponding lower limit for the momentum transfer becomes

$$p_{\perp \min} = \frac{2 r_0 p_0^2}{b_{\max} q} \ .$$

Hence one can write

$$\sigma\left\langle \delta p_{\perp}^{2} \right\rangle = 2\pi\left(r_0 m_0^2 c^2\right)^2 \frac{1}{p_x^2} \ln\left[\frac{p_x}{p_m}\right]^2$$

with

$$\left[\frac{p_x}{p_m}\right]^2 = \frac{p_x^2 b_{\max}}{r_0 p_0^2}$$

where

$$p_m = p_0 \sqrt{\frac{r_0}{b_{\max}}}$$

can be considered as a lower limit on the effective initial momentum (this effect comes from the use of the classical treatment).

Now, in order to take into account all possible relative velocities (or momenta) and all possible particle locations within the bunch, one needs to average the expression: $n\left\{\sigma v \left\langle \delta p_{\perp}^{2} \right\rangle\right\}_{c.m.}$.

However, considering Gaussian distributions and the fact that velocity distribution is independent of particle location one has

$$\overline{n\left\{\sigma v \left\langle \delta p_{\perp}^{2} \right\rangle\right\}_{c.m.}} = \bar{n}\overline{\left\{\sigma v \left\langle \delta p_{\perp}^{2} \right\rangle\right\}_{c.m.}}$$

with

$$\bar{n} = N \frac{\iiint \rho \cdot \rho dV}{\iiint \rho dV} = \frac{N}{2^3 \pi^{3/2} \sigma_x \sigma_z \sigma_s}$$

and

$$\overline{\left\{\sigma v\left\langle \delta p_{\perp}^{2}\right\rangle\right\}}=\int\frac{2p_x}{m_0}\sigma\left\langle \delta p_{\perp}^{2}\right\rangle P(p_x)dp_x$$

where $P(p_x)dp_x$ is the differential probability acting on the variable p_x.

Since individual particle momenta have a Gaussian distribution, and since

$$p_x=\frac{|\vec{p}_1-\vec{p}_2|}{2}$$

then

$$\sigma_{p_x}^2=\frac{1}{2}\sigma_p^2\ .$$

The transverse momentum being invariant in a Lorentz transformation

$$\sigma_p=\gamma\ \ p_0\sigma_{x'}$$

where $\sigma_{x'}$ is given from betatron dynamics considerations:

$$\sigma_{x'}=\frac{\sigma_x}{\beta}\ .$$

Finally the bracket becomes

$$\overline{\left\{\sigma v\left\langle \delta p_{\perp}^{2}\right\rangle\right\}}_{c.m.}=\frac{4\sqrt{\pi}r_0^2m_0^3c^4}{\sigma_p}\int_{p_m}^{\infty}\frac{1}{p_x}\ln\left[\frac{p_x}{p_m}\right]^2e^{-p_x^2/\sigma_p^2}dp_x\ .$$

With a suitable change of variable

$$\chi=\left[\frac{p_x}{\sigma_p}\right]^2$$

it takes the more convenient form:

$$\overline{\left\{\sigma v\left\langle \delta p_{\perp}^{2}\right\rangle\right\}}_{c.m.}=\frac{2\sqrt{\pi}r_0^2p_0^3c}{\sigma_p}f(\chi_m)$$

with

$$f(\chi_m)=\int_{\chi_m}^{\infty}\frac{1}{\chi}\ln\left(\frac{\chi}{\chi_m}\right)e^{-\chi}d\chi$$

$$\chi_m=\left[\frac{p_m}{\sigma_p}\right]^2=\frac{r_0p_0^2}{b_{\max}\sigma_p^2}$$

Now putting all terms together one gets for the r.m.s. energy spread from multiple scattering:

$$\sigma_E^2 = \frac{N r_0^2 p_0^2 c^3 \beta_x \ \tau_E}{2^5 \pi \ \gamma \sigma_x^2 \sigma_z \sigma_s} f(\chi_m) \ .$$

The function $f(\chi_m)$ is tabulated on Fig. 4 for a practical range of χ_m values.

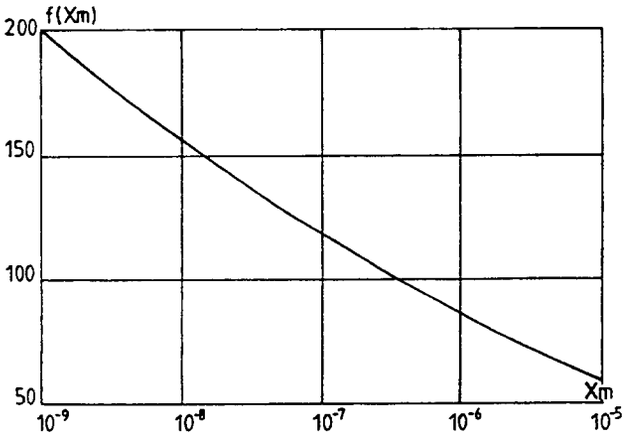


Fig. 4 Plot of the function $f(\chi_m)$

In practice the total energy spread will be the result of the contributions from the present effect and the quantum radiation. These two effects are not correlated so they add quadratically

$$\left(\sigma_E^2\right)_{\text{total}} = \left(\sigma_E^2\right)_{\text{rad.}} + \left(\sigma_E^2\right)_{\text{Tousch.}} \ .$$

This formula is self consistent since the bunch volume itself depends on the total energy spread.

3.4 Bunch lengthening and emittance growth

In the longitudinal phase space, energy oscillations are correlated with phase oscillations. The equilibrium distribution of particles around the synchronous phase is described by the r.m.s. bunch length directly proportional to the r.m.s. energy spread:

$$\sigma_s = R\alpha \frac{\omega_s}{\Omega_s} \frac{\sigma_E}{E_s}$$

where R is the physical ring radius, ω_s the angular revolution frequency, Ω_s the angular synchrotron frequency, E_s the nominal energy, and α the momentum compaction.

Since the multiple Touschek effect corresponds to random energy changes, the horizontal betatron motion and its equilibrium state will be affected due to finite dispersion

$$\sigma_x = \left[\beta_x U_x + D_x^2\right]^{1/2} \frac{\sigma_E}{E}$$

where the invariant U_x is given by the lattice optical functions:

$$U_x = \frac{J_E}{J_x}\left\langle \frac{1}{\beta_x}\left[D_x^2 + \left(\beta_x D_x' - \frac{1}{2}\beta_x' D_x\right)^2 \right] \right\rangle .$$

Note that in this expression the averaging is made in the bending magnets only.

Generally, with no bending in the vertical plane, the r.m.s. vertical beam dimension comes from residual coupling and is expressed through a coupling coefficient K:

$$U_z = K\ U_x$$

$$\sigma_z = \sqrt{K\ \beta_z U_x}\,\frac{\sigma_E}{E_s} .$$

Hence the effective bunch volume will grow as $(\sigma_E / E_s)^3_{\text{total}}$ and σ_p as $(\sigma_E / E_s)_{\text{total}}$.

Considering the $f(\chi_m)$ is not a fast varying function of χ_m the multiple Touschek effect will decrease approximately as

$$(\sigma_E)_{\text{Tousch.}} \propto \frac{1}{(\sigma_E^2)_{\text{total}}} ,$$

this behaviour remaining valid for all beam dimensions.

Since the natural bunch volume due to quantum radiation decreases when the energy decreases, and since the damping time increases rapidly at the same time, it is clear that the effect increases rapidly with decreasing ring energy. Hence a good starting approach consists of evaluating the effect in the absence of synchrotron radiation and then trying to locate the energy threshold for the multiple Touschek effect to occur (Fig. 5).

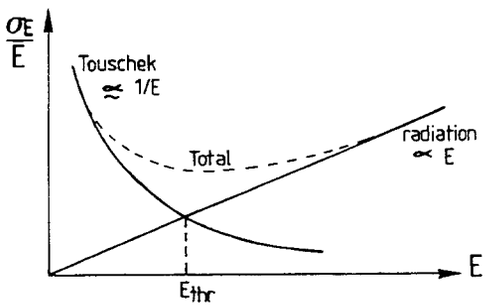


Fig. 5 Energy threshold for multiple Touschek effect

In that case the practical formulae for the effect become:

$$\left(\frac{\sigma_E}{E_s}\right)^6 = \frac{N\ r_0^2 \beta_x \tau_E \Omega_s f(\chi_m)}{2^5 \pi \gamma^3 \left[\beta_x U_x + D_x^2\right] \sqrt{K \beta_z U_x}\,\alpha}$$

$$\chi_m = \frac{N^{1/3}\ r_0 \beta_x^2 Q_s^{1/3}}{2\sqrt{\pi}\gamma^2 \left[\beta_x U_x + D_x^2\right]^{7/6} (K\beta_z U_x)^{1/6} R^{1/3} \alpha^{1/3}} \left(\frac{\sigma_E}{E_s}\right)^{-3}$$

where $Q_s = \Omega_s / \omega_s$.

The system is solved by putting

$$\left(\frac{\sigma_E}{E_s}\right)^6 = A \; f(\chi_m)$$

$$\left(\frac{\sigma_E}{E_s}\right)^3 = \frac{B}{\chi_m}$$

where x_m is the solution of the following equation

$$f(\chi_m) = \frac{1}{A}\left(\frac{B}{\chi_m}\right)^2 .$$

* * *

REFERENCES

[1] C. Bernardini et al., Phys. Rev. Lett. Vol. 10 (1963), p. 407.

[2] J. Haissinski, Calcul des pertes de particules dues aux collisions à l'intérieur d'un même paquet. Cas d'ADA. (Rétablissement de la formule de MM. Bernardini et Touschek.) Rapport technique interne No. 41-63. Anneaux de Stockage. LAL Orsay (22 avril 1963).

[3] J. Haissinski, Corrections relativistes du calcul de l'effet ADA. Rapport technique 52 bis - 63. Anneaux de stockage. LAL, Orsay (17 juin 1963).

[4] J. Haissinski, Thèse (Orsay 1965).

[5] U. Volkel, DESY 67/5 (1967).

[6] J. Le Duff, RT 1-72. Anneaux de Collisions. LAL Orsay (1972).

[7] H. Bruck and J. Le Duff, Beam enlargement in storage rings by multiple Coulomb scattering. Proceedings of the 5th Int. Conf. on High Energy Accelerators, Frascati, 1965 (INFN, Frascati, 1965).

[8] J. Le Duff, Thèse (Orsay 1965).

[9] H. Bruck, Accélérateurs circulaires de particules. Presses Universitaires de France (1966).

[10] A. Piwinski, Proc. 9th Int. Conf. on High Energy Accelerators, Stanford, 1974 (Conf. 740522, USAEC, Washington, 1974), p. 405.

[11] A. Piwinski, Proc. of the CERN Accelerator School, Paris, 1984, CERN 85-19 (1985).

[12] A.H. Sørensen, Proc. of the CERN Accelerator School, Aarhus, 1986, CERN 87-10 (1987).

STOCHASTIC COOLING

D. Möhl
CERN, Geneva, Switzerland

Abstract

This paper describes the main analytical approaches to stochastic cooling. The first is the time domain picture in which the beam is rapidly sampled and a statistical analysis is used to describe the cooling behaviour. The second is the frequency domain picture, which is particularly useful since the observations made on the beam are mainly in this domain. This second picture is developed in detail to assess ingredients of modern cooling theory like mixing and signal shielding and to illustrate some of the diagnostic methods. Finally the use of a distribution function and the Fokker-Plank equation are discussed, which give the most complete description of the beam during the cooling.

1 INTRODUCTION

Beam cooling aims at reducing the size and the energy spread of a particle beam circulating in a storage ring. This reduction of size should not be accompanied by beam loss; thus the goal is to increase the particle density.

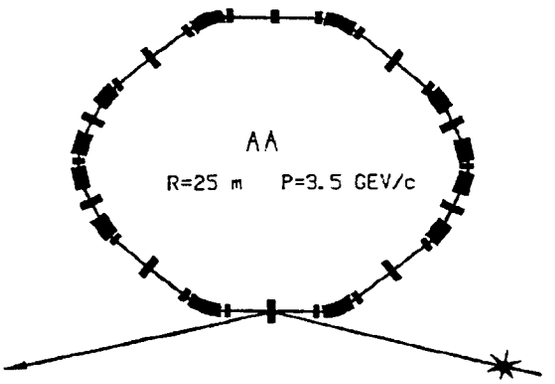
Since the beam size varies with the focusing properties of the storage ring, it is useful to introduce normalized measures of size and density. Such quantities are the (horizontal, vertical and longitudinal) emittances and the phase-space density. For our present purpose they may be regarded as the (squares of the) horizontal and vertical beam diameters, the energy spread, and the density, normalized by the focusing strength and the size of the ring to make them independent of the storage ring properties.

Phase-space density is then a general figure of merit of a particle beam, and cooling improves this figure of merit.

The terms beam temperature and beam cooling have been taken over from the kinetic theory of gases. Imagine a beam of particles going around in a storage ring. Particles will oscillate around the beam centre in much the same way that particles of a hot gas bounce back and forth between the walls of a container. The larger the mean square of the velocity of these oscillations in a beam the larger the beam size. The mean square velocity spread is used to define the beam temperature in analogy to the temperature of the gas which is determined by the kinetic energy $0.5\ mv_{rms}^2$ of the molecules.

Why do we want beam cooling? The resultant increase of beam quality is very desirable for at least three reasons:

i) Accumulation of rare particles
Cooling to make space available so that more beam can be stacked into the same storage ring. The Antiproton Accumulator (AA) at CERN is an example of this (see Fig. 1).



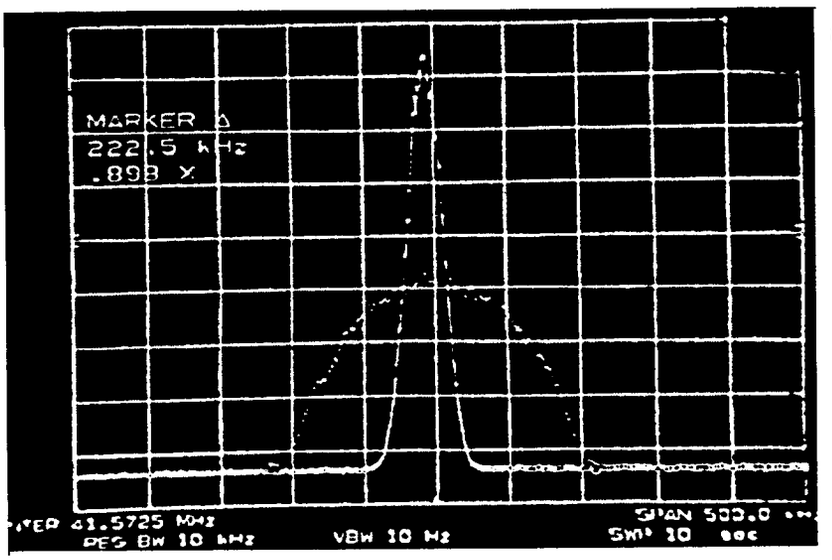
QUANTITY	BEAM IN	STACK		GAIN
N	10^7	6×10^{11}		6×10^4
E_h	100π	3.5π	[mm•mrad]	28
E_v	100π	2.0π	[mm•mrad]	50
$\Delta p/p$	7.5	2.0	$[10^{-3}]$	4
$\frac{N}{E_h \cdot E_v \Delta p/p}$	130	4×10^{10}		3×10^8

Figure 1: The CERN Antiproton Accumulator (AA). Sketch and table of performance with number of particles, horizontal emittance, vertical emittance and momentum spread of incoming beam and of stack after 24 h of accumulation (design values).

ii) Improvement of interaction rate and resolution
Cooling to provide sharply collimated and highly mono-energetic beams for precision experiments with colliding beams or beams interacting with fixed targets. The Low Energy Antiproton Ring (LEAR) at CERN is an example of this (see Fig. 2).

iii) Preservation of beam quality
Cooling to compensate for various mechanisms leading to growth of beam size and/or loss of stored beam. Again LEAR is an example of this application.

Several cooling techniques are operative or have been discussed [1]: Electron beams have a tendency to cool 'by themselves' owing to the emission of radiation as the orbit is curved. The energy radiated decreases very strongly with increasing rest mass of the particles. For (anti-)protons and heavier particles, radiation damping is negligible at energies currently accessible in accelerators. 'Artificial' damping had therefore to be devised, and two such methods have been successfully put to work during the last decade: i) cooling of heavier particles by the use of an electron beam — this is the subject of lectures on electron cooling; and ii) stochastic cooling by the use of a feedback system, which will be discussed in this chapter.



a) Momentum cooling at injection in LEAR; $\sqrt{dN/dp}$ *displayed against momentum;* 3×10^9 *antiprotons, before and after 3 minutes of cooling.* $\Delta p/p$ *is reduced by a factor 4.*
b) Comparison of the cooled beam extracted from LEAR to the low energy antiproton beams previously obtained in secondary beam lines.

Figure 2: An example of momentum spread cooling and properties of the cooled beam from the CERN Low Energy Antiproton Ring (LEAR) compared to a secondary beam used before 1980.

2 SIMPLIFIED THEORY, TIME-DOMAIN PICTURE

2.1 The basic set-up

The arrangement for cooling of the horizontal beam size is sketched in Fig. 3. Assume, for the moment, that there is only one particle circulating. Unavoidably, it will have been injected with some small error in position and angle with respect to the ideal orbit (centre of the vacuum chamber). As the focusing system continuously tries to restore the resultant deviation, the particle oscillates around the ideal orbit. Details of these 'betatron oscillations' [2] are given by the focusing structure of the storage ring, namely by the distribution of quadrupoles and gradient magnets (and higher-order 'magnetic lenses') which provide a focusing force proportional to the particle deviation (and to higher-order powers of the deviation).

For the present purpose, we can approximate the betatron oscillation by a purely sinusoidal motion. The cooling system is designed to damp this oscillation. A pick-up

electrode senses the horizontal position of the particle on each traversal. The error signal — ideally a short pulse with a height proportional to the particle's deviation at the pick-up — is amplified in a broad-band amplifier and applied on a kicker which deflects the particle by an angle proportional to its error.

In the simplest case, the pick-up [3] consists of a plate to the left of the beam and a plate to the right of it. If the particle passes to the left, the current induced on the left plate exceeds the current on the right one and vice versa. The difference between the two signals is a measure of the position error. The 'kicker' is, in principle, a similar arrangement of plates on which a transverse electromagnetic field is created which deflects the particle [3].

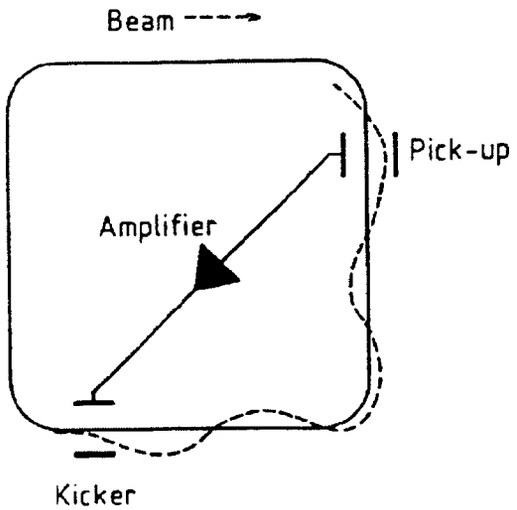


Figure 3: The principle of (horizontal) stochastic cooling. The pick-up measures horizontal deviation and the kicker corrects angular error. They are spaced by a quarter of the betatron wavelength λ_β (plus multiples of $\lambda_\beta/2$). A position error at the pick-up transforms into an error of angle at the kicker, which is corrected.

Since the pick-up detects the position and the kicker corrects the angle, their separation is chosen to correspond to a quarter of the betatron oscillation (plus an integer number of half wavelengths if more distance is necessary). A particle passing the pick-up at the crest of its oscillation will then cross the kicker with zero position error but with an angle which is proportional to its displacement at the pick-up. If the kicker corrects just this angle the particle will from thereon move on the nominal orbit. This is the most favourable situation (sketched as Case 1 in Figs. 4 and 5). A particle not crossing the pick-up at the crest of its oscillations will receive only a partial correction (Cases 2 and 3 in Figs. 4 and 5). As we shall see later, it will then take several passages to eliminate the oscillation.

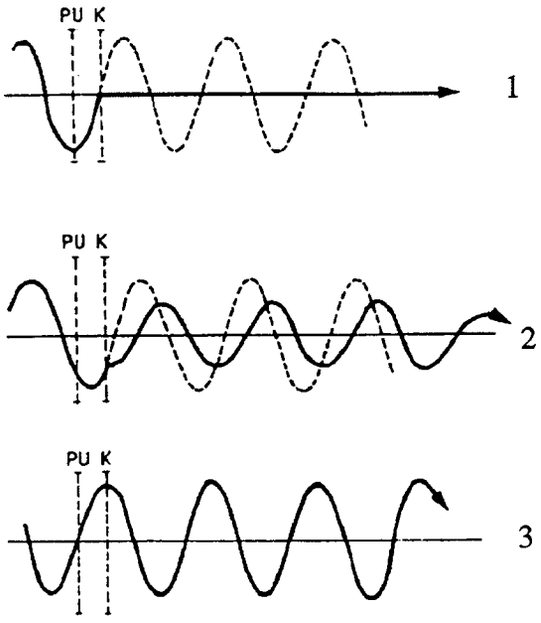


Figure 4: The importance of betatron phase: Particle 1 crosses the pick-up with maximum displacement. Its oscillation is (ideally) completely cancelled at the kicker. Particle 2 arrives at an intermediate phase; its oscillation is only partly eliminated. Particle 3 arrives with the most unfavourable phase and is not affected by the system.

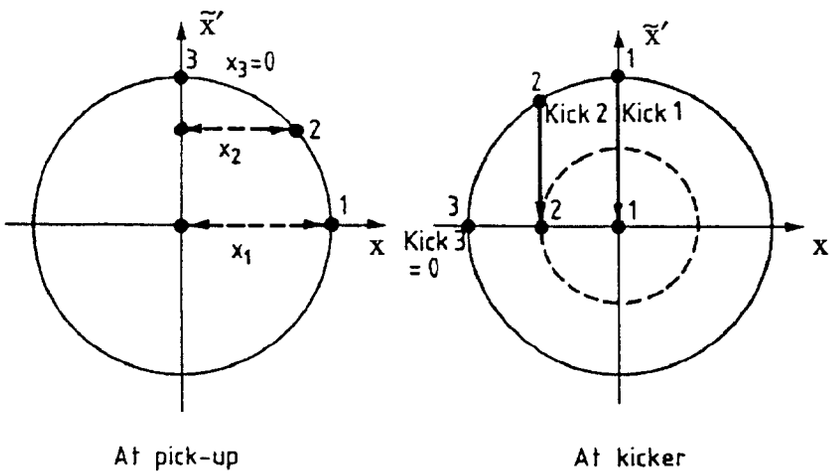


Figure 5: Phase space representation of betatron cooling. The same as for Fig. 4 except that a 'polar diagram' $\tilde{x}' = f(x)$ is used to represent the betatron motion $x = \hat{x} \sin [Q(s/R) + \psi_0]$, $x' = (R/Q)\, x' = \hat{x} \cos [Q(s/R) + \psi + 0]$. The undisturbed motion of a particle is given by a circle with the radius equal to the betatron amplitude $\hat{x}$. Kicks correspond to a jump of x'. The cooling system tries to put particles onto smaller circles. Particles 1, 2 and 3 are sketched with the most favourable, the intermediate, and the least favourable initial phase, respectively. As the number of oscillations per turn is different from an integer or half-integer, particles come back with different phases on subsequent turns and all particles will be cooled progressively.

Another particularity of stochastic cooling is easily understood from the single particle model (Fig. 3): the correction signal has to arrive at the kicker at the same time as the test particle. Since the signal is delayed in the cables and the amplifier, whereas a high-energy particle moves at a speed close to the velocity of light, the cooling path has usually to take a short cut across the ring. Only at low and medium energy ($v/c < 0.5$) is a parallel path feasible.

We have thus familiarized ourselves with two constraints on the distance pick-up to kicker: taken along the beam, this distance is fixed, or rather quantized, owing to the required phase relationship of the betatron oscillation; taken along the cooling path this length is fixed by the required synchronism between particle and signal. A change of energy (particle velocity) and/or a change of the betatron wavelength will therefore require special measures. Incidentally, the first of these two conditions, is due to the oscillatory nature of the betatron motion. For momentum spread cooling in a coasting beam, where the momentum deviation of a particle is constant rather than oscillatory, this constraint does not come into play and a greater freedom in the choice of pick-up-to-kicker distance exists.

It is now time to leave the one-particle consideration and turn our attention to a beam of particles which oscillate incoherently i.e. with different amplitudes and with random, initial phase.

By beam cooling we shall now mean a reduction with time of the amplitude of each individual particle. To understand stochastic cooling, we will next have a closer look at the response of the cooling system. This permits us to discern groups of particles — so called samples — which will receive the same correcting kick during a passage through the system.

2.2 Notion of beam samples

To be able to analyse the response of the cooling system, let us start with an excursion into elementary pulse and filtering theory [4]. What we would like to take over is a bandwidth/pulse-length relation known as the Küpfmüller or Nyquist theorem*:

If a signal has a Fourier decomposition of band-width $\Delta f = W$, *then its 'typical' time duration will be*

$$T_s = 1/(2W) \ .$$

This is illustrated in Fig. 6, where we sketch the Fourier spectrum of a pulse and the resulting time-domain signal. Clearly the two representations are linked by a Fourier transformation, and this permits us to check the theorem.

---

* The bandwidth/pulse length relation was introduced by Nyquist and independently by Küpfmüller in 1928. This theorem is closely-related to the more general sampling theorem of communication theory: If a function S(t) contains no frequencies higher than W cycles per second, it is completely described by its value $S(mT_s)$ at sampling points spaced by $\Delta t = T_s = 1/2W$ (i.e. taken at the 'Nyquist rate' 2W; see, for example, J.A. Betts, Signal Processing and Noise, English Universities Press, London, 1970). The Küpfmüller theorem is a consequence of Parseval's relation for the Fourier transform. For the distributions S(t) of interest here it holds if we define $T_s = (1/S^2_{max}) \int_{-\infty}^{\infty} S^2_{(t)}\, dt$ and $2\,\pi W = (\int_0^{\infty} s(\omega) d\omega)^2 / \int_0^{\infty} s^2(\omega)\, d\omega$.

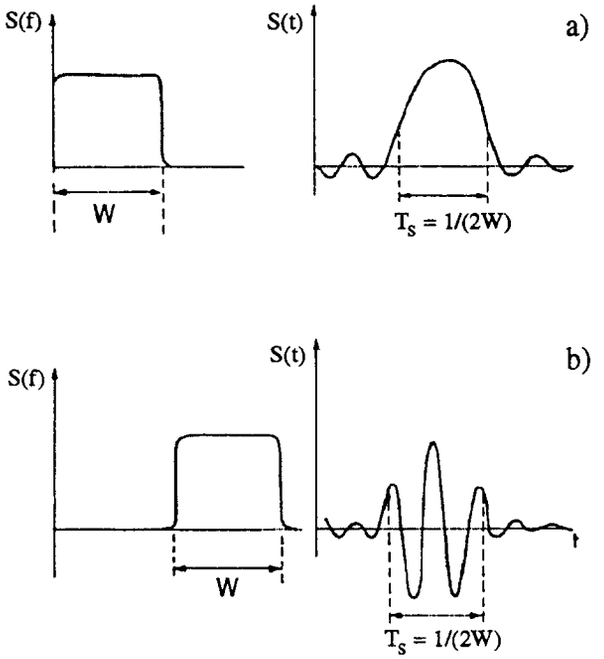


Figure 6: Illustration of the Küpfmüller-Nyquist relation: a signal whose Fourier decomposition S(f) has a bandwidth W, has a typical time duration $T_s = 1/(2W)$. Illustration for a 'low-pass' [case (a)] and a 'band-pass' signal [case (b)].

For curiosity, note the difference between a pulse with a low-frequency and a high frequency spectrum (both cases are sketched in Fig. 6). In spite of the different shape of the time-domain signal, the 'typical duration' is in both cases 1/(2W).

A corollary to the theorem is well known to people who design systems for transmitting short pulses:

When a short pulse is filtered by a low-pass or band-pass filter of bandwidth W, the resulting pulse has a 'typical' time width (see Fig. 7)

$$T_s = 1/(2W) \; . \tag{2.1}$$

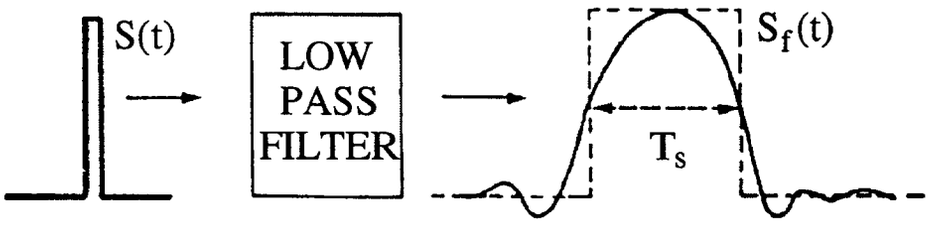


Figure 7: Input and output signal S(t) of a low-pass system and 'rectangular' approximation to the output pulse $S_f(t)$.

In this form, the theorem is directly applicable to our cooling problem, to which we now return. Passing through the pick-up, an off-axis particle induces a short pulse with a length given by the transit time. Owing to the finite bandwidth (W) of the cooling

system, the corresponding kicker signal is broadened into a pulse of length T_s. To simplify considerations, we approximate the kicker pulse by a rectangular pulse of total length T_s (Fig. 8).

A test particle passing the system at t_0 will then be affected by the kicks due to all particles passing during the time interval $t_0 \pm T_s/2$. These particles are said to belong to the sample of the test particle. In a uniform beam of length T (revolution time), there are $\ell_s = T/T_s = 2\,WT$ equally spaced samples of length T_s with

$$N_s = N/(2WT) \text{ particles per sample} . \tag{2.2}$$

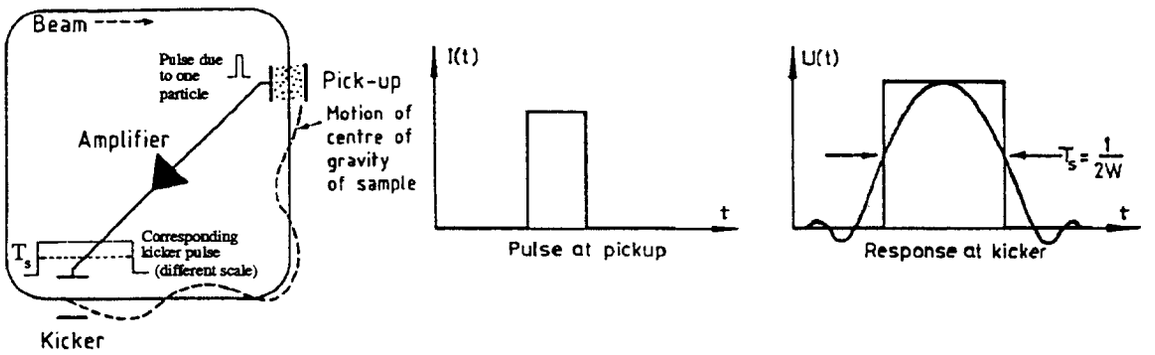


Figure 8: Pick-up signal of a particle and corresponding kicker pulse (idealized). The test particle experiences the kicks of all other particles passing within time $-T_s/2 \leqslant \Delta t \leqslant T_s/2$ of its arrival at the kicker. These particles are said to belong to the sample of the test particle. Cooling may be discussed in terms of the centre-of-gravity motion of samples.

Table 1: An example of samples corresponding to cooling at injection in LEAR [5]

No. of particles in the beam	N	10^9
Revolution time	T	0.5 μs
Transit time in one pick-up unit	T_t	0.1 ns
Cooling system bandwidth	W	250 MHz
Sample length	T_s	2 ns
No. of samples per turn	$\ell_s = T/T_s$	250
No. of particles per sample	N_s	4×10^6

2.3 Coherent and incoherent effects

The model of samples has allowed us to subdivide the beam into a large number of slices which are treated independently of each other by the cooling system. If the bandwidth can be made large enough so that there are no other particles in the sample of the test particle, then the single-particle analysis is still valid. However, to account for the reality of some million particles per sample, we have to go a step further and do

some simple algebra. This will permit us to discern two slightly different pictures of the cooling process. In the 'test particle picture' we shall view cooling as the competition between: i) the 'coherent effect' of the test particle upon itself via the cooling loop; and ii) the 'incoherent effect', i.e. the disturbance to the test particle by the other sample members (see Fig. 9). In the 'sampling picture' we shall understand stochastic cooling as a process where samples are taken from the beam at a rate ℓ_s per turn. By measuring and reducing the average sample error, the error of each individual particle will (on the average) slowly decrease.

A few simple equations will illustrate these pictures. Let us denote by x the error of the test particle and assume that the corresponding correction at the kicker is proportional to x, say λx. With no other particles present, the error would be changed from x to a corrected

$$x_c = x - \lambda x \tag{2.3}$$

i.e. the test particle receives a correcting kick,

$$\Delta x = -\lambda x \ . \tag{2.4}$$

In reality the kicks $-\lambda x_i$ of the other sample members have to be added, and the corrected error after one turn and the corresponding kick are

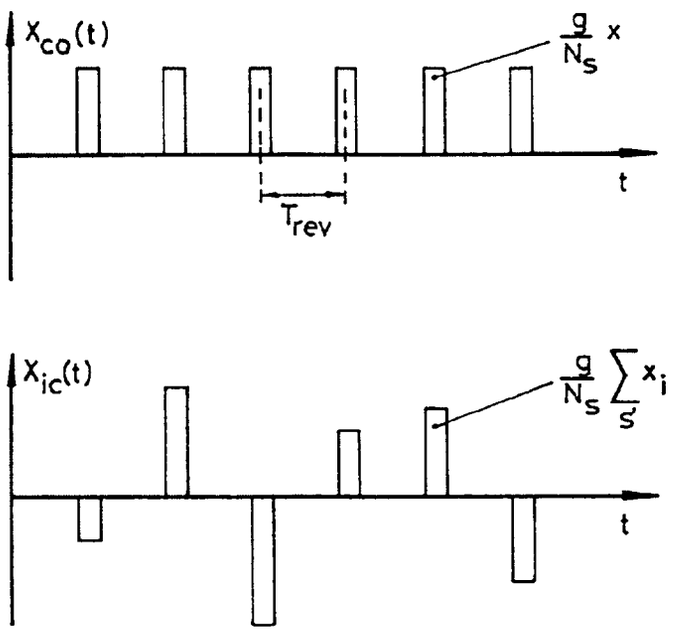


Figure 9: Cooling system signals for the test particle picture. Signals at the instant of passage of the test particle are sketched. The upper trace gives the coherent correction signal due to the test-particle itself. The lower trace sketches the incoherent signal due to the other particles in the sample. The kick experienced is the sum of coherent and incoherent effects. If the amplification is not too strong and the sample population is small, the coherent effect which is systematic will predominate over the random heating by the incoherent signals.

$$x_c = \underbrace{x - \lambda x}_{\text{coherent effect}} - \underbrace{\sum_{s'} \lambda_i x_i}_{\text{incoherent effect}}$$

$$\Delta x = -\lambda x - \sum_{s'} \lambda_i x_i \ . \tag{2.5}$$

In our rectangular response model, $\lambda_i = \lambda$ is the same for all sample members. Hence, we can also write

$$\Delta x = -\lambda x - \lambda \sum_{s'} x_i \ . \tag{2.6}$$

Equations (2.5) and (2.6) clearly exhibit the 'coherent' and the 'incoherent' effects mentioned above. The sum labelled s' includes all particles in the sample except the test particle. You may want to rewrite this sum including the test particle (this sum will be labelled s) and interpret it in terms of the average sample error (the sample centre of gravity if you like), which is by definition

$$\langle x \rangle_s = \frac{1}{N_s} \sum_{s} x_i \ . \tag{2.7}$$

Equations (2.5) and (2.6) then become

$$x_c = x - (\lambda N_s)\langle x \rangle_s \equiv x - g\langle x \rangle_s \tag{2.8a}$$

$$\Delta x = -(\lambda N_s)\langle x \rangle_s \equiv -g\langle x \rangle_s \ . \tag{2.8b}$$

This introduces the second picture. What the cooling system does is to measure the average sample error and to apply a correcting kick, proportional to $\langle x \rangle_s$ to the test particle. Up to now the sample is defined with respect to a specific test particle; however, to the extent that any beam slice of length T_s has the same average error $\langle x \rangle_s$ our considerations apply to any test particle. This is true on a statistical basis, as will become clear later.

A word about notation. It has become customary to write $\lambda N_s = g$, and to call g the 'gain'. Remember that this g is proportional to the amplification (the electronic gain) of the system and proportional to N_s. As from Eqs. (2.8), $-g = \Delta x/\langle x \rangle_s$, a more precise (but longer) name is 'fraction of observed sample error corrected per turn'.

Now, we can again separate the coherent and incoherent effects and rewrite Eq. (2.6), by using the above notation:

$$\Delta x = \underbrace{-\frac{g}{N_s} x}_{\substack{\downarrow \\ \text{coherent} \\ \text{term} \\ \text{(cooling)}}} - \underbrace{g\langle x \rangle_{s'}}_{\substack{\downarrow \\ \text{incoherent} \\ \text{term} \\ \text{(heating)}}} \tag{2.9}$$

Clearly, the problem is how to treat the incoherent term. The following approximations will be discussed:

First approximation:	Neglect the incoherent term
Second approximation:	Treat it as a fluctuating random term
Third approximation:	Treat it as a fluctuating random term with some coherence due to imperfect mixing
Fourth approximation:	Include additional coherence due to 'feedback via the beam'

2.3.1 First approximation

Neglecting completely the incoherent term in Eq. (2.9) we get a best performance estimate

$$\Delta x = -\frac{g}{N_s} x \,. \tag{2.10}$$

We expect an exponential form, $x = x_0 e^{-t/\tau}$ for the amplitude of the test particle which gives the damping rate

$$\frac{1}{\tau} = -\frac{1}{x}\frac{dx}{dt} \simeq \frac{1}{x}\frac{\Delta x}{\Delta t} = \frac{1}{x}\frac{\Delta x \text{ per turn}}{T}. \tag{2.11}$$

Substituting into Eq. (2.11) from Eq. (2.10) gives

$$\frac{1}{\tau} = \frac{g}{T} N_s \,. \tag{2.12}$$

Interpreting g as the fractional correction, we intuitively accept that it is unhealthy to correct more than the observed sample error, i.e. we assume $g \leqslant 1$. Let us put $g = 1$ to make an estimate of the upper limit.

Finally it is convenient to express N_s in terms of the total number of particles, N, in the beam and by the system's bandwidth W, i.e. $N_s = N(T_s/T) = N/2WT)$ [see Eq. (2.2)]. We then obtain, a first useful approximation to the cooling rate:

$$\boxed{\frac{1}{\tau} = \frac{2W}{N}} \tag{2.13}$$

Amazingly enough, this simple relation overestimates the optimum cooling rate by only a factor of 2. However, to gain confidence, we have to justify some of our assumptions, especially the restriction of $g \leqslant 1$ and the neglect of the incoherent term. In fact, an evaluation of this term will clarify both assumptions and provide guidance on how to include other adverse effects such as amplifier noise.

2.3.2 Towards a better evaluation of the incoherent term

To be able to deal with the incoherent term, we make a detour into statistics to recall a few elementary 'sampling relations' [6]. Consider the following problem.

Given a beam of N particles characterized by an average $\langle x \rangle = 0$ and a variance $\langle x^2 \rangle \equiv x_{rms}^2$ of some error quantity x, suppose we take a random sample of N_s particles and do statistics on the sample population — rather than on the whole beam — to determine

i) the sample average $\langle x\rangle_s$;

ii) the sample variance $\langle x^2\rangle_s$;

iii) the square of the sample average $[\langle x\rangle_s]^2$, i.e. the square of (i).

What are the most probable values [the expectation values, denoted by $E(\langle x\rangle_s)$, etc.] of these sample characteristics?

For random samples the most probable values are:

i) sample average $\rightarrow$ beam average;

ii) sample variance $\rightarrow$ beam variance;

iii) square of sample average $\rightarrow$ beam variance/sample population.

Or, in more mathematical language,

$$E(\langle x\rangle_s) = \langle x\rangle = 0 \tag{2.14a}$$

$$E(\langle x^2\rangle_s) = \langle x^2\rangle = x_{rms}^2 \tag{2.14b}$$

$$E[(\langle x\rangle_s)^2] = x_{rms}^2/N_s\ . \tag{2.14c}$$

Results (2.14a) and (2.14b) are in agreement with common sense, which expects that the sample characteristics are true approximations of the corresponding population characteristics. This is the basis for sampling procedures. Equation (2.14c) is more subtle as it specifies the error to be expected when one replaces the population average by the sample average.

$$(\langle x\rangle_s)^2 - (\langle x\rangle)^2 \rightarrow x_{rms}^2/N_s$$

or symbolically

$$\langle x\rangle \approx \langle x\rangle_s \pm x_{rms}/\sqrt{N_s}\ .$$

In other words: the larger the beam variance and the smaller the sample size (N_s), the more imprecise is the sampling. In this form, Eqs. (2.14) are used in statistics to determine the required sample size for given accuracy and presupposed values for the beam variance x_{rms}^2.

A slightly different interpretation is useful in the present context: suppose we repeat the process of taking beam samples and working our $\langle x\rangle_s$ many times. Although the beam has zero $\langle x\rangle$, the sample average will in general have a finite (positive or negative) $\langle x_s\rangle$. The sequence of sample averages will fluctuate around zero (around $\langle x\rangle$ in general) with a mean-square deviation x_{rms}^2/N_s. This is the fluctuation (or, if you prefer, the noise) of the sample average due to the finite particle number.

A simple example to illustrate the sampling relations and to familiarize us further with $\langle x^2\rangle_s$ and $(\langle x\rangle_s)^2$ is given in Table 2. It is amusing to note that in this example 'the most probable values' 1/3 and 2/3 respectively [which agree with Eqs. (2.14)] never occur for any of the possible samples — just another instance of statistics dealing with averages and being unjust to the individual.

Table 2: An example of the sampling relations

Assume a discrete distribution such that the values $x = -1, 0, 1$ occur with equal probability. Hence, beam average: $\langle x \rangle = 0$, and beam variance: $\langle x^2 \rangle = x_{rms}^2 = 1/3[(-1)^2 + 0^2 + 1^2] = 2/3$. Consider samples of size $N_s = 2$. To work out the most probable values of the sample characteristics, write down all possible samples of size $N_s = 2$, determine $\langle x \rangle_s$, $(\langle x \rangle_s)^2$, and $\langle x^2 \rangle_s$, and take the average of these averages to find the expectations.

Sequence		Sample averages		
		$\langle x \rangle_s$	$(\langle x \rangle_s)^2$	$\langle x^2 \rangle_s$
−1	−1	−1	1	1
−1	0	−0.5	0.25	0.5
−1	1	0	0	1
0	−1	−0.5	0.25	0.5
0	0	0	0	0
0	1	0.5	0.25	0.5
1	−1	0	0	1
1	0	0.5	0.25	0.5
1	1	1	1	1
Expectation = average of above values		$0 = \langle x \rangle$	$1/3 = \langle x^2 \rangle / 2$	$2/3 = \langle x^2 \rangle$

To conclude our detour, let us mention that the sampling relations (2.14) are a consequence of the more general 'central limit theorem' [6] of statistics. For the present purpose we can quote this theorem as follows:

When a large number of random samples of size N_s *are taken from a population with statistics* $\langle x \rangle = 0$ *and* $\langle x^2 \rangle = x_{rms}^2$ *then the distribution of the sample averages is approximately Gaussian with a mean equal to the population mean and a standard deviation* $\sigma = x_{rms}/\sqrt{N_s}$.

2.3.3 A better approximation of the cooling rate – second approximation

Returning to Eq. (2.8a), but re-expressing $\langle x \rangle_s$ in full we have,

$$x_c = x - \frac{g}{N_s} \sum_s x_i \ . \tag{2.15}$$

In order to profit from the sampling relations, it is more useful to evaluate the change $\Delta(x^2) = x_c^2 - x^2$ of the squared error rather than Δx. Thus we obtain,

$$\Delta(x^2) = -2g \frac{x}{N_s} \sum_s x_i + \left(\frac{g}{N_s} \sum_s x_i \right)^2 \ . \tag{2.16}$$

The second term in Eq. (2.16) immediately gives

$$\left(\frac{g}{N_s} \sum_s x_i \right)^2 = g^2 (\langle x \rangle_s)^2 \to \frac{g^2}{N_s} x_{rms}^2 \ , \tag{2.17}$$

where we have used the sampling relation (2.14c) to express the expected variance of the sample average in terms of the beam variance x_{rms}^2. To work out the first term we separate the test particle (once again) from the sum and write

$$x \cdot \frac{1}{N_s} \sum_s x_i = \frac{x^2}{N_s} + \frac{x}{N_s} \sum_{s'} x_i \ .$$

Next we apply the sampling relation (2.14a) to the remaining sum, i.e. we take

$$E\left(\frac{1}{N_s - 1} \sum_{s'} x_i \right) = E\left(\langle x \rangle_{s'} \right) = 0$$

under the assumption that the sample (labelled s′) without the test particle is a random sample such that Eq. (2.14a) applies. Then

$$E\left(x \cdot \frac{1}{N_s} \sum_s x_i \right) = \frac{x^2}{N_s} \ . \tag{2.18}$$

Thus the first term in Eq. (2.16) has non-zero expectation. Clearly this is due to the fact that the x at the front 'coheres' with the corresponding term inside the sum.

Putting together the terms, the expected change is then

$$\Delta(x^2) \to -\frac{2g}{N_s} x^2 + \frac{g^2}{N_s} x_{rms}^2 \ . \tag{2.19}$$

Equation (2.19) applies to any test particle. Taking as typical a particle with an error equal to the beam r.m.s. we can write especially:

$$\frac{1}{x_{rms}^2} \Delta(x_{rms}^2) \to -\frac{1}{N_s}(2g - g^2) \ . \tag{2.20}$$

This gives the cooling rate (per second) for the beam variance:

$$\frac{1}{\tau_{x^2}} = -\frac{1}{T} \frac{\Delta(x_{rms}^2)}{x_{rms}^2} = \frac{1}{N_s T}(2g - g^2) = \frac{2W}{N}(2g - g^2) \ . \tag{2.21}$$

Clearly the term 2g presents the coherent effect already identified. The $-g^2$ term represents the incoherent heating by the other particles. The inclusion of this term is the improvement obtained in the statistical evaluation of this section.

It emerges quite naturally from Eq. (2.21) that g should not be too large! In fact, optimum cooling (maximum of $2g - g^2$) is obtained with $g = 1$, and antidamping occurs if $g > 2$ (see Fig. 10).

It should be remembered that Eq. (2.21) gives the cooling rate $1/\tau_x^2$ for x^2; the rate $1/\tau$ for x is half of this, as can be verified by comparing $x^2 \equiv x_0^2 \exp(-t/\tau_{x^2})$ and $x^2 = [x_0 \exp(-t/\tau)]^2$.

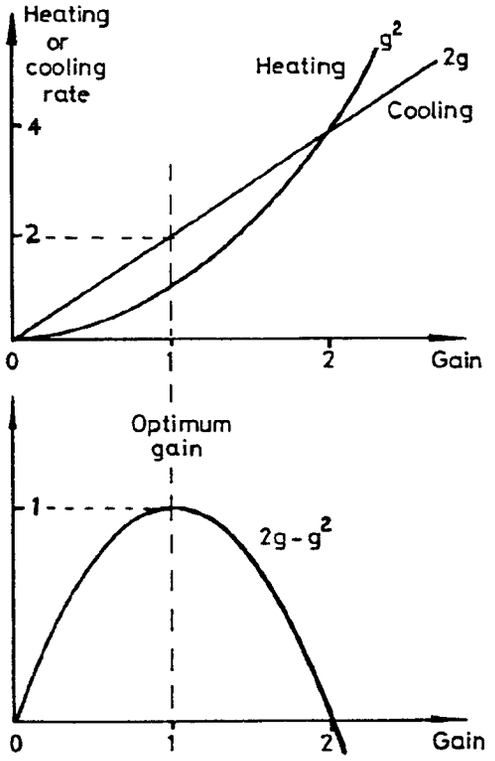


Figure 10: Cooling or heating rate when considering the incoherent term as a random fluctuation.

2.3.4 Alternative derivation

For those who were not pleased with the way in which we separated the test particle from its sample and regarded the remainder as a random sample of size $N_s - 1$, we give yet another derivation of Eq. (2.21) which is due to Hereward (unpublished notes 1976, see also Ref. [7]).

We restart from Eq. (2.16), which we write as

$$\Delta(x^2) = -2gx \cdot \langle x \rangle_s + g^2(\langle x \rangle_s)^2 \ . \tag{2.22}$$

This is the change for one test particle and one turn. We now take the average of this over the sample of the test particle (before, we took the average for one particle over many turns).

A slight complication arises from the fact that strictly speaking each particle defines its own sample, as sketched in Fig. 11. We can assume, however, that the long-term behaviour of any sample (i.e. any beam slice of length T_s) is the same, so that expectation values are independent of the choice of the sample.

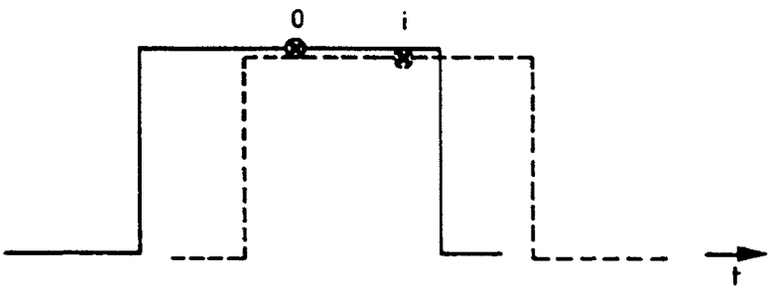


Figure 11: Sample of the original test-particle (0) and of a particle passing earlier (i). Working out the average $\langle x_i \langle x \rangle_s \rangle_s$ of $x_i \langle x \rangle_s$ each particle has to be associated with its own sample. To the extent that all beam samples have the same statistical properties, all long-term averages are the same: $\langle x_i \langle x \rangle_s \rangle_s \rightarrow (\langle x \rangle_s)^2$.

Then the only variable on the r.h.s. involved in averaging over the original sample is the x in the first term, and we obtain

$$\langle \Delta(x^2) \rangle_s \rightarrow -2g(\langle x \rangle_s)^2 + g^2(\langle x \rangle_s)^2 \ . \tag{2.23}$$

Next we use the sampling relations (2.14b) and (2.14c). We include the fact that the correction (2.23) is applied to all beam samples once per turn. Thus,

$$\langle \Delta(x^2) \rangle_s \rightarrow \Delta x^2_{rms} \ ,$$

$$(\langle x \rangle_s)^2 \rightarrow x^2_{rms}/N_s \ ,$$

and the expected correction of beam variance per turn is

$$\Delta(x^2_{rms}) = -\frac{1}{N_s} \ (2g - g^2) \ x^2_{rms}$$

i.e. exactly as assumed in Eq. (2.20).

This leads to the same cooling rate as that given by the previous approach, but the derivation lends itself to the following formulation of the 'sampling picture'.

Take a random beam sample of N_s particles. Measure and correct its average error $\langle x \rangle_s$ by giving a kick $-g\langle x \rangle_s$ to all particles. Owing to the finite particle number, the beam variance appears as a fluctuation with 'noise' $(\langle x \rangle_s)^2 \rightarrow x^2_{rms}/N_s$ of the centre of gravity $\langle x \rangle_s$. By correcting $\langle x \rangle_s$ to $(1 - g)$ of its value (i.e. to zero for full $g = 1$), one reduces the sample variance (on the average) by $1/N_s$ $(2g - g^2)$. Repeat N/N_s times per turn to reduce the beam variance by the same amount. Repeat for many turns.

Table 3: 'Simulation' of a one-turn correction (with g = 1) using the example of Table 2. We note down all possible samples of size $N_s = 2$ and reduce the sample errors to zero by applying the same correction to both sample members. This reduces the beam variance from 2/3 to 1/3, i.e. $\Delta x^2_{rms}/x^2_{rms} = 1/N_s = 1/2$

Before correction				After correction			
Sequence		Sample		Sequence		Sample	
		Average $\langle x_s \rangle_s$	Variance $\langle x^2 \rangle_s$			Average	Variance
−1	−1	−1	1	0	0	0	0
−1	0	−0.5	0.5	−0.5	0.5	0	0.25
−1	1	0	1	−1	1	0	1
0	−1	−0.5	0.5	0.5	0.5	0	0.25
0	0	0	0	0	0	0	0
0	1	0.5	0.5	−0.5	0.5	0	0.25
1	−1	0	1	1	−1	0	1
1	0	0.5	0.5	0.5	−0.5	0	0.25
1	1	1	1	0	0	0	0
'Beam variance' (average of all sample variances)			2/3				1/3

Thus, rather than treating single particles, one measures and corrects the centres of gravity of beam samples. It is amusing (but not too surprising) to note that the total number of measurements, namely the number of turns $n = N_s$ required for reasonable cooling multiplied by the number $\ell_s = N/N_s$ of samples per turn, is N, as if we treated the N particles individually.

It is easy to test this sampling prescription for simple distributions; in Table 3 we use the previous example (Table 2) to verify that the full correction (g = 1) reduces the variance by $1/N_s$ per turn. More generally, the sampling recipe can easily be simulated on a desk computer using a random number generator.

In the next two sections we will use the test particle and the sampling picture alternately to introduce two final ingredients, namely electronic noise of the amplifier and mixing of the samples due to the spread in revolution time.

2.3.5 A refinement to include system noise

A large amplification of the error signals detected by the pick-up is necessary to give the required kicks to the beam. Electronic noise in the preamplifiers then becomes important. In Table 4 we anticipate some typical numbers pertaining to transverse cooling of 10^9 antiprotons in LEAR. This example should convince us of the necessity to rewrite the basic equations to include noise. It is convenient [7] to represent noise by an equivalent sample error (denoted by x_n) as observed at the pick-up. We then regard the system sketched in Fig. 12 and write

$$x_c = x - g\langle x \rangle_s - g x_n \ . \tag{2.24}$$

Table 4: Signal, noise, and amplification of a cooling system; orders of magnitude for 10^9 particles and 50 s cooling time

Pick-up signal	50 nA
Preamplifier noise current	150 nA
Kicker voltage per turn	1 V
Corresponding current (into 50 Ω)	20mA
Power amplification	$\approx 2 \times 10^{10}$

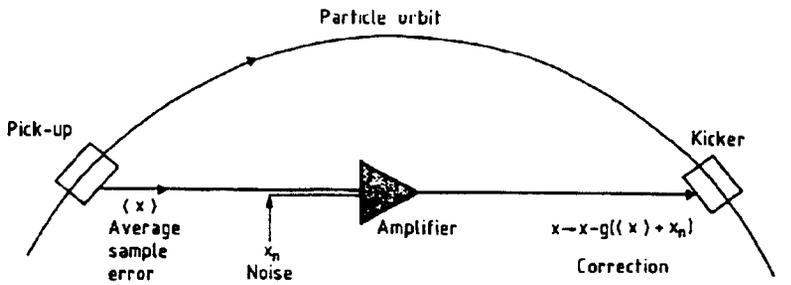


Figure 12: Cooling loop including system noise. The noise is represented as an equivalent sample error $x_n(t)$ as observed at the pick-up.

Going once again through our basic procedure, taking random noise uncorrelated with the particles we obtain the expected cooling rate

$$\frac{1}{\tau_{x^2}} = \frac{2W}{N}[2g - g^2(1 + U)] \ , \tag{2.25}$$

where $U = E(x_n^2)/E\langle x\rangle_s^2)$ is the ratio of the expected noise to the expected signal power**, called the 'noise-to-signal power ratio' or noise-to-signal ratio for brevity.

This introduces the noise into our pictures: it increases the incoherent term by $(1 + U)$. System noise and the distrubance caused by the other particles enter in much the same way; the latter is therefore also called particle noise.

Several things can be observed from Eq. (2.25). Cooling remains possible despite very poor signal-to-noise ratios ($1/U \ll 1$). All we have to do is to choose g small enough ($g < g_0 = 1/(1 + U) \approx 1/U$), which unavoidably means slow cooling ($\tau > NU/2W$). In other words, we have to be patient and give the system a chance to distil a signal out of the noise.

In the initial cooling experiment (ICE) [8] with 200 circulating antiprotons the system worked with signal-to-noise ratios as low as 10^{-6}.

Secondly, U has a tendency to increase as cooling proceeds: namely the noise tends to remain the same, whereas the signal decreases as the beam shrinks. This is the case

** Stochastic cooling of heavy ions is becoming very important so we should note that $U \rightarrow U/Z^2$ where Z is the charge number of the particle (i.e. the number of electrons stripped off) and U the noise-to-signal ratio calculated for singly charged particles.

unless the pick-up plates are mechanically moved to stay close to the beam edge — as will be done in the new antiproton collector ACOL [9] to be built at CERN.

With changing U, cooling is no longer exponential. Equation (2.25) gives a sort of instantaneous rate, and cooling stops completely ($1/\tau \rightarrow 0$) when U has increased such that $(1 + U) = 2/g$. In this situation, equilibrium is reached between heating by noise and the damping effect of the system. To avoid this 'saturation' it is sometimes advantageous to decrease g during cooling in order to work always close to the optimum gain [maximum of Eq. (2.25)] $g_0 = 1/(1 + U)$.

In all cases it is important to obtain a good signal-to-noise ratio. Frequently, this means having a large number of pick-ups as close as possible to the beam, as well as high quality, low-noise preamplifiers often working at cryogenic temperatures.

2.4 Mixing - third approximation

So far, all our considerations have been based on the assumption of random samples. This is a good hypothesis for an undisturbed beam. However, the cooling system is designed to correct the statistical error of the samples. Just after correction, samples will no longer be random. For full correction the centre of gravity $\langle x \rangle_s$ will be zero rather than $\sqrt{x^2_{rms}/N_s}$ as expected for random conditions. Cooling will then stop as no error signal is observable.

Fortunately, owing to momentum spread, particles in a storage ring go round at slightly different speeeds, and the faster ones continuously overtake the slower ones. Because of this mixing, the sample population changes and the sample error reappears, until ideally all particles have zero error. The dispersion of revolution time with momentum is governed by

$$\frac{\Delta T}{T} = -\eta \frac{\Delta p}{p} \tag{2.26}$$

where the off-momentum funtion [2] $\eta = \gamma^{-2} - \gamma_t^{-2}$ is given by the distance of the working energy (γ) from transition (γ_t).

If mixing is fast so that complete re-randomization has occured on the way from kicker to pick-up then the assumption of random samples made in the previous sections is valid. If however, mixing is incomplete, cooling is slower. In fact, if it takes M turns for a particle of typical momentum error to move by one sample length with respect to the nominal particle ($\Delta p/p = 0$), then intuitively one expects an M times slower cooling rate.

A slightly different way of looking at imperfect re-randomization suggests itself in the frame of the test particle picture: bad mixing means that a particle stays too long — namely M rather than 1 turns — together with the same noisy neighbours. This increases the incoherent heating by the other particles by a factor M.

We thus generalize the basic Eq. (2.25) (a rigorous derivation will be given later)

$$\frac{1}{\tau_{x^2}} = \frac{2W}{N}[2g - g^2(M + U)] \; . \tag{2.27}$$

and call $M \geqslant 1$ the mixing factor.

M is defined as the number of turns for a particle with one standard deviation in momentum to migrate by one sample length T_s.

Equation (2.27) has the optimum,

$$\begin{aligned} g &= g_0 = 1/(M+U) \ , \\ \tau &= \tau_0 = \frac{N}{2W}\,(M+U) \ . \end{aligned} \tag{2.28}$$

This underlines the importance of having good mixing — $M \to 1$ — on the way from correction to the next observation, but ...

What about mixing between observation and correction? Surely if the sample as observed is very different from the sample as corrected, then adverse effects can happen. Let us again resort to the test particle description and try to imagine how the coherent and the incoherent effects change. As to the latter, we expect that it is to first order not affected. We can just assume that the perturbing kicks are due to a new sample which has the same statistical properties as the original beam 'slice'.

The coherent effect will, however, change because the system will be adjusted in such a way that the correction pulse will be synchronous with the nominal particle ($\Delta p/p = 0$). Particles that are too slow or too fast on the way from pick-up to kicker will therefore slip with respect to their self-induced correction (Fig. 13). In fact, in the rectangular response model used above, the coherent effect will be completely zero if the particle slips by more than half the sample length ($|\Delta T_{PK}|) T_s/2$). At this stage, it is more realistic to use a parabolic response model of the form $1 - (\Delta T/T_c)^2$, where T_c, the useful width of the corection pulse, is about equal to the sample length T_s for a low-pass system. But T_c is shorter than T_s for a high-frequency band-pass system with $f_{min} \geqslant W$, with a response as sketched in Fig. 6(b); ΔT is the time-of-flight error of the particle between pick-up and kicker. Introducing the typical error ΔT_{PK} and calling $\Delta T_{PK}/T_c = 1/\widetilde{M}$, we can modify the coherent term $g \to g[1 - \widetilde{M}^{-2}]$ to account for unwanted mixing between observation and correction. In a regular lattice the flight time from pick-up to kicker is a fixed fraction of the time from kicker to pick-up, and the two mixing factors M and $\widetilde{M}$ are proportional to each other, $M = \alpha\,\widetilde{M}$, with α being the ratio of the corresponding distances — hence the interest in having a short beam path from pick-up to kicker.

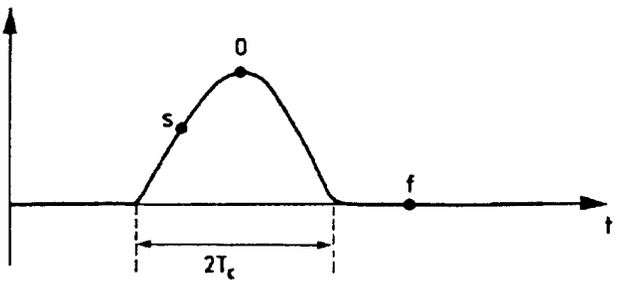


Figure 13: Synchronism between particles and their correcting pulse on their way from pick-up to kicker. The response of the cooling system to a particle (the 'coherent effect') is approximated by a 'parabola' $s(t) = 1 - (\Delta t/T_c)^2$ of width $\pm T_c$ instead of the 'rectangle' used in Figs. 7 and 8. A nominal particle (0) arrives at the kicker simultaneously with the correction kick. The particle f is much too fast and advances its correction pulse. The particle s is slightly too slow. Thus, the three particles receive full correction, no correction, or partial correction, respectively.

The reduced correction becomes,

$$\Delta x_c = -2g\left[1 - \left(\frac{\Delta T_{PK}}{T_c}\right)^2\right] x = -2g[1 - \widetilde{M}^{-2}]x \; . \tag{2.29}$$

This will give us a slightly different form for the basic equation,

$$\frac{1}{\tau_{x^2}} = \frac{2W}{N}[2g(1 - \widetilde{M}^{-2}) \; - \; g^2(M + U)] \; . \tag{2.30}$$

	Coherent	Incoherent
	(cooling)	(heating)

By a clever choice of the bending and focusing properties of the storage ring it is possible, in principle, to make $\Delta T_{PK} \to 0$ independent of momentum, and ΔT_{KP} large to approach the desired situation of $\widetilde{M}^{-2} = 0$ and $M = 1$. But this complicates the storage ring lattice. The compromise adopted in existing rings is to sacrifice some of the desired re-randomisation in order to avoid too much unwanted mixing.

Then the unwanted mixing imposes a limit on the bandwidth or, more precisely, on the upper frequency of the cooling band. In fact if the time of flight error ΔT_{PK} is bigger than T_c, cooling becomes ineffective. For a bandpass with flat response from f_{min} to f_{max} the time response (Fourier transform of the frequency characterist) has the form

$$\frac{1}{t}\left[\cos\left(\frac{\omega_{min} + \omega_{max}}{2}\, t\right) \times \sin\left(\frac{\omega_{max} - \omega_{min}}{2}\, t\right)\right] \; .$$

The useful width (given in this case by the first zero crossing) is

$$T_c = \frac{1}{2(f_{min} + f_{max})} \lesssim \frac{1}{2f_{max}} \; .$$

To relate this to the momentum spread, we express the trime of flight error:

$$\Delta T_{PK} = T_{PK}\, \eta_{PK}\, \frac{\Delta p}{p} = \alpha\, T\, \eta_{PK}\, \frac{\Delta p}{p} \; .$$

Here η_{PK} is the 'local off-momentum function PU to K' of the ring, T_{PK} the nominal local flight time, $T = 1/f_{rev}$ the revolution period and $\alpha = T_{PK}/T$. For a 'regular lattice' η_{PK} is equal to the $\eta = \left|\frac{1}{\gamma^2} - \frac{1}{\gamma_t^2}\right|$ of the whole ring.

Thus the condition $\Delta T_{PK} \lesssim T_c$ yields

$$\frac{1}{2f_{max}} \gtrsim \alpha T \eta_{PK}\, \frac{\Delta p}{p} \approx \alpha T \eta\, \frac{\Delta p}{p}$$

Taking as an example:
(distance PU to K/circumference) $= \alpha = 0.5$,

$$f_{rev} = \frac{1}{T} = 1.5 \text{ MHz}, \; \eta_{PK} = \eta = 0.01 \; , \qquad \frac{\Delta p}{p} = 2 \times 10^{-2}$$

(corresponding in round numbers to the Antiproton Accumulator Ring AA at CERN) we obtain

$$f_{max} \lesssim 7.5 \text{ GHz} \; .$$

This is close to the band limit used (f_{max} = 8 GHz) although the limit in the AA is not strictly applicable because the different cooling systems cover only parts of the aperture and do not 'see' the full $\Delta p/p$ of 2%. In any case a small value of $\alpha\eta_{PK}$ (short distance PU - K, operation close to transition energy and/or special optics) is essential to be able to work with large bandwidth.

We now return to the cooling rate which following convention we write for x rather than x^2 using $1/\tau = (1/2)\ 1/\tau_{x^2}$. Including both mixing effects as well as amplifier noise, we write

$$\boxed{\frac{1}{\tau} = \frac{W}{N}\left[2g(1-\widetilde{M}^{-2}) - g^2(M+U)\right]\ .} \tag{2.31}$$

(for factors of 2 see Appendix 3).

Equation (2.31) is the main result of our simple analysis. It exhibits some of the fundamental limitations of stochastic cooling. We first note that $1/\tau$ has a maximum characterized by

$$g_0 = \frac{1-\widetilde{M}^{-2}}{M+U}\ , \tag{2.32}$$

$$\frac{1}{\tau_0} = \frac{W}{N}\left(\frac{(1-\widetilde{M}^{-2})^2}{M+U}\right)\ . \tag{2.33}$$

As an example of relatively straightforward technology, we take W = 250 MHz. Then, in the best of all cases ($M \to 1,\ U \to 0,\ \widetilde{M}^{-2} \to 0$) this gives

$$1/\tau = W/N = 2.5 \times 10^8/N\ [\mathrm{sec}^{-1}] \tag{2.34}$$

i.e τ = 1 s at 2.5×10^8 p or $\tau \approx 1$ day at 10^{13} p.

To include mixing, we assume that the time-of-flight dispersion between pick-up and kicker and between kicker and pick-up and the system response are such that the unwanted mixing is one half of the wanted mixing, i.e. we put (as an example) $\widetilde{M}^{-1} = 0.5\ M^{-1}$. We further assume that the sensitivity and the number of pick-ups are such that U = 1 (little is gained in this example in going to more pick-ups, such that $U \ll 1$). Then the best cooling, obtained with $M \approx 1.4$, is

$$1/\tau \approx 0.32 W/N\ . \tag{2.35}$$

This is about three times slower than the rate (2.34) with $\widetilde{M}^{-2} \to 0,\ U \to 0$.

We retain that over a wide range of parameters $1/\tau = a_0\ W/N$. From Fig. 14 we conclude that 'classical' cooling systems follow a 'working line' with $1/\tau \approx 0.1$ to 0.3 W/N, i.e. a_0 = 0.1–0.3. A bandwidth of 250 to 500 MHz was standard in the 'first generation cooling experiments'; 1 to 8 GHz is used in the CERN-ACOL and the Fermilab antiproton sources. Bands of 10 GHz and higher have been contemplated for sources accumulating 10^{13} antiprotons in a few hours [10], as desirable for multi-TeV colliders (see Table 5). Most recent proposals of 'optical stochastic cooling' discuss a bandwidth of $10^{13} - 10^{14}$ Hz (!) but then the problem of mixing between observation and correction requires a time spread $\Delta T_{PK} < 10^{-13} - 10^{-14}$ sec. Even with special lattice insertions this is a formidable task.

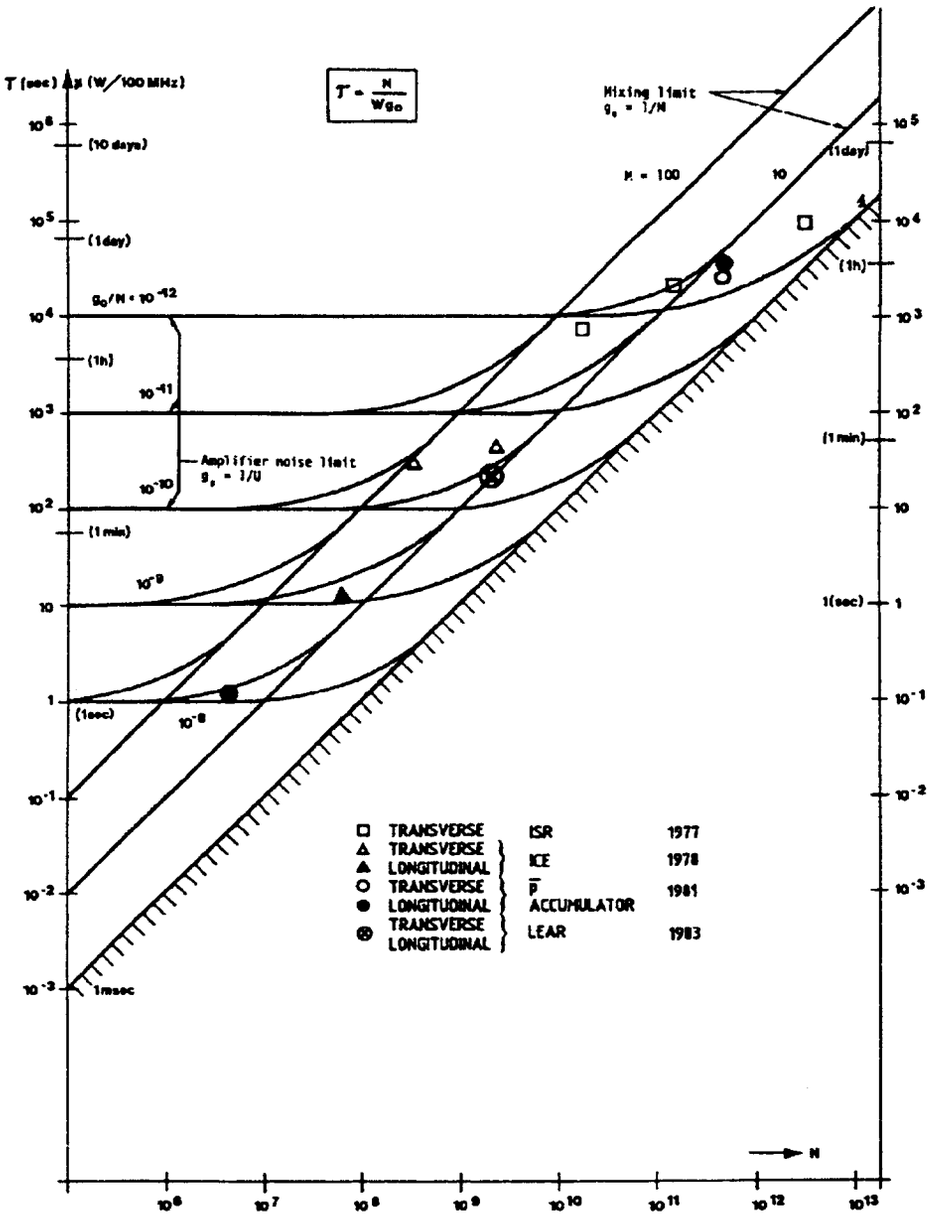


Figure 14: Normalized cooling time versus intensity. The inclined lines represent the mixing limit. For low intensity the cooling time levels off because of noise. The points represent initial cooling in various machines. These points roughly follow a line with $\tau \approx 10$ N/W. During cooling, noise and/or mixing tend to become more important and the cooling time longer. Note that the vertical scale is normalized for 100 MHz bandwidth.

Table 5: Parameters of past, present, and 'ultimate' cooling systems. The quantity a_0, defined by $1/\tau = a_0$ W/N describes the efficiency of solving the noise and mixing problems

Machine		Date	W GHz	a_0	τ[s]/N (theoretical)
First generation	ICE	1976	0.1	0.03	$1/(3 \times 10^6)$
	AA (precooling)	1980	0.25	0.1	$1/(2 \times 10^7)$
Second generation	Fermilab	1986	4	0.25	$1/(1 \times 10^9)$
	ACOL		4	0.25	$1/(1 \times 10^9)$
Present	Fermilab and ACOL upgrade	1990	8	0.25	$1/(2 \times 10^9)$
Ultimate		?	15	0.5	$1/(7 \times 10^9)$
'Futuristic'		?	10^5	0.5	$1/(5 \times 10^{13})$

2.5 Practical details

So far we have, in a general way, discussed a system for correcting 'some error x'.

In practice cooling is used to reduce the horizontal and/or vertical betatron oscillation and the momentum spread of the beam. Table 6 gives a summary of the corresponding hardware.

The simple time-domain approach can be directly applied to momentum cooling by the Palmer-Hereward method. This will be discussed in the next subsection. A discussion of the other momentum cooling methods and of betatron oscillation cooling will be deferred to later sections.

Table 6: Stochastic cooling systems in use or proposed

Type	Pick-up	Corrector
Betatron cooling, horizontal or vertical	Difference pick-up	Transverse kicker
Momentum cooling, Palmer-Hereward type	Horizontal difference pick-up	Longitudinal kicker
Momentum cooling, filter method	Longitudinal (sum) pick-up + comb filter	Longitudinal kicker
Momentum cooling, transit time method	Longitudinal pick-up + differentiator or two longitudinal pick-ups	Longitudinal kicker

2.6 Palmer cooling

A horizontal position pick-up is used to detect the horizontal orbit displacement $x = D\langle \Delta\ p/p\rangle_s$ concurrent with the momentum error of the sample; D (also denoted by α_p or x_p) is the value of the orbit 'dispersion function' at the pick-up as determined by the focusing properties of the storage ring. In addition to the momentum dependent displacement there are further contributions to the position error, especially the betatron oscillation (x_β) of the particles. We shall neglect this contribution, assuming that the pick-up is placed in a region of large dispersion so that $\langle D(\Delta p/p)\rangle_s$ dominates $\langle x_\beta\rangle_s$. We are then in a situation where momentum cooling as envisaged by R. Palmer (private communication to L. Thorndahl and H.G. Hereward in 1975) is possible. At the RF gap the particle receives a 'kick' of momentum and hence a change of x proportional to the detected error.

The basic one-passage equation (including noise) is written as

$$x_c = x - g[D\langle \Delta p/p\rangle_s + x_n]\ .$$

This is completely equivalent to Eq. (2.15), thus leading to the cooling rate [Eq. (2.16)] for the momentum deviation

$$\frac{1}{\tau_{\Delta p}} = \frac{W}{N}[2g(1-\widetilde{M}^{-2}) - g^2(M+U)]\ , \tag{2.36}$$

where $U = E(x_n^2)/E[(D\langle \Delta p/p\rangle_s)^2]$ is the noise-to-signal ratio; x_n is the system noise expressed as the equivalent pick-up signal $D(\Delta p/p)$, and $E(x_n^2)$ the expectation (i.e. the long-term average) of x_n^2. Above we assumed that the orbit dispersion $x = D\ (\Delta p/p)$ dominates at the pick-up so that the betatron oscillation x_β is negligible there. We also implied that at the kicker the dispersion function D_k, as well as its derivative, D'_k, are zero. Otherwise the momentum correction leads to an excitation of horizontal betatron oscillations. The effect is that the momentum kick introduces an abrupt change of the equilibrium orbit and the particle starts to oscillate around this new displaced orbit.

The more realistic case where both x and x_β are present at the pick-up and where D_k is non-zero at the kicker was analysed by Hereward. He showed [11] the mutual heating and at the same time the possibility of using the Palmer system for simulteneous longitudinal and horizontal cooling by a suitable choice of the pick-up to kicker distance.

3 A MORE DETAILED PRESENTATION OF BETATRON COOLING, FREQUENCY DOMAIN PICTURE

3.1 Betatron equation

Before entering into details, it is worth trying to establish a simple picture of betatron cooling in which the various phenomena can be identified.

Consider first the smooth sinusoidal approximation for the betatron motion [2] of a single particle (subscript i) in a storage ring, with forcing terms on the right-hand side arising from its proper motion, the motion of other particles (subscript j) and system noise.

$$\ddot{x}_i(t)+\Omega^2Q^2x_i(t) = \underbrace{G_{ii}x_i(t-t_p)}_{\text{Coherent effect}} + \underbrace{\sum_n G_{ij}x_j(t-t_p)}_{\text{Incoherent effect}} + \underbrace{\text{'system noise'}}_{\text{Additional incoherent effect}}$$

Coherent effect
−Mixing PU → K
−Betatron phase errors

Incoherent effect
−Mixing K → PU
−Signal shielding

Additional incoherent effect
−Enhancement of cooling

(3.1)

We interpret the left-hand side as the motion on entering the cooling kicker (K) and the forcing terms on the right-hand side as being derived from the motion seen earlier (i.e. at $t-t_p$) in the pick-up (PU). The characteristics of the pick-up, amplifier, transmission system and kicker enter into both the coefficient G_{ij} and the 'system noise', Ω is the revolution frequency and Q the tune of the storage ring.

3.2 Simplified coherent effect

If we neglect the incoherent terms in Eq. (3.1) and make G_{ii} = constant, we obtain a single-particle cooling equation,

$$\ddot{x}(t) + \omega_\beta^2 x(t) = G_{ii}\, x(t-t_p) \tag{3.2}$$

putting $\omega_\beta = \Omega Q$.

For a weak perturbation term, we can expect a solution of the form:

$$\begin{aligned} x(t) &= \tilde{x}\, e^{i\omega t} \\ x(t-t_p) &= \tilde{x}\, e^{i\omega t - i\mu}\ . \end{aligned}$$

Substituting into Eq. (3.2) gives,

$$\omega_\beta^2 - \omega^2 = G_{ii}\, e^{-i\mu}$$

$$-2\omega_\beta \Delta\omega \approx G_{ii}\, e^{-i\mu}$$

$$\Delta\omega = \frac{-G_{ii}}{2\omega_\beta}\, e^{-i\mu}\ . \tag{3.3}$$

This is the expected response of a feedback system. The real part of $\Delta\omega$ is the frequency shift of the perturbed oscillation and the imaginary part of $\Delta\omega$ gives the damping (or heating) of the oscillation.

$$\frac{1}{\tau} = \mathrm{Im}(\Delta\omega) = \mathrm{Re}\left[\frac{G_{ii}}{2\omega_\beta}\, e^{i(\mu-\pi/2)}\right]\ . \tag{3.4}$$

Equation (3.4) would be exact if the observation and feedback on the beam were continuous, and G_{ii} constant, which are manifestly not the case. We must now, therefore, investigate the effects of periodic observation and correction.

3.3 Orbit equation for a constant localized kick

The orbit in a storage ring with constant kicks can be regarded as a betatron oscillation which closes onto itself by virtue of the angular discontinuities at the kicks. The closed orbit is given for any distribution of kicks by the well-known equation [2],

$$x''(s) + K(s)x(s) = \left(\frac{\Delta B_z(s)}{B_0\rho_0} + \frac{E_x(s)}{p\beta}\right) = F(s) \tag{3.5}$$

where:

E_x is the transverse electric field [V/m]
ΔB_z is the transverse magnetic field [T] error (vertical field for horizontal orbit deformation and vice versa)
$B_0\rho_0$ is the magnetic rigidity [T·m] $= 3.3356 \cdot 10^{-9}$ p [eV/c]
p is the particle momentum [eV/c] and $\beta = v/c$
$s = \beta ct$ is the distance along the orbit.

The Eq. (3.5) is good for numerical calculations, but is inconvenient for analytical work, since it is defined piecewise in s as one proceeds around the machine through the various elements.

Using the well known Courant and Snyder transformation [2] the equation can be rewritten as a 'driven harmonic oscillator' with fixed frequency Q rather than with azimuthally varying K(s) as in Eq. (3.5):

$$\eta''(\phi) + Q^2\eta = Q^2\beta_x^{3/2}(s)F(s) \tag{3.6}$$

where $\eta = x(s)\beta_x^{-1/2}(s)$ is the normalised displacement, $d\phi = ds/Q\beta_x$ defines the Courant and Snyder angle which increase by 2π per turn, β_x is the betatron function of the storage ring and $'$ now indicates differentiation with respect to ϕ.

Equation (3.6) is quite general, but we are especially interested by a single narrow kick, which we can represent by a delta function, or rather by a periodic delta function ($\tilde{\delta}$) with the revolution frequency (see Fig. 15).

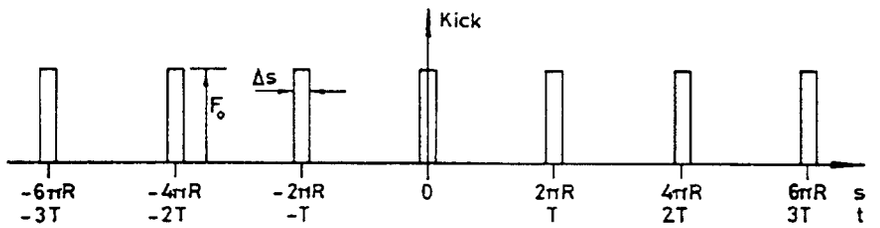


Figure 15: For a single narrow kick in a storage ring the particle sees a periodic influence.

The kick as seen by a circulating particle is represented by,

$$F(s) = (F_0\Delta s)\tilde{\delta}(s - n2\pi R) = (F_0\Delta s)\sum_{n=\infty}^{\infty}\delta(s - n2\pi R) \tag{3.7}$$

(accepting that high frequency components, i.e. $f > c/\Delta s$ are not required).

Equation (3.7) is a good representation of a short kicker with a constant kick, but analytically it would be easier to manipulate if we could replace the discontinuous delta functions with continuous functions. This can readily be done by making a Fourier analysis of the quantities $\beta_K^{3/2}$ F(s) which appears on the r.h.s. of Eq. (3.6).

$$\beta_x^{3/2}\,(s)F(s) = \sum_{\ell=-\infty}^{\infty} f_\ell(\phi)\, e^{i\ell\phi} \tag{3.8}$$

where,

$$f_\ell = \frac{1}{2\pi}\int_0^{2\pi} \beta_x^{3/2} F\, e^{-i\ell\phi}\, d\phi = \frac{1}{2\pi}\int_0^{2\pi} \beta_K^{3/2}\, F\, e^{-i\ell\phi(s)}\, \frac{ds}{Q\beta_x}\,.$$

Since F is a delta function the integral (3.8) simply leads to Fourier coefficients which are all equal. (This is the advantage of the complex formulation, which avoids the $f_0/2$ coefficient of the real expansion.)

$$f_0 = f_\ell = \frac{F_0 \Delta s}{2\pi Q}\beta_K^{1/2} = \frac{\Theta}{2\pi Q}\,\beta_K^{1/2} \tag{3.9}$$

where β_k is the value of the beta function at the kicker (s = 0) and

$$\Theta = \left(\frac{\Delta B_z}{B_0\rho_0} + \frac{E_x}{p\beta}\right)\Delta s$$

is the kick strength.

We can now rewrite Eq. (3.6) in terms of continuous functions

$$\eta''(\phi) + Q^2\eta = Q^2\frac{\sqrt{\beta_k}}{2\pi Q}\,\Theta \sum_{\ell=-\infty}^{\infty} e^{i\ell\phi} \tag{3.10}$$

This is in fact all we really need, but we can make two variable changes, which will make the equation more familiar. Firstly, since we prefer to think in terms of the time, we can introduce a 'time-like' variable ϕ/Ω i.e. the Courant and Synder normalised phase, ϕ, scaled by the revolution angular frequency, Ω. Rather loosely we will still refer to this variable as t. In fact, this lack of rigour is not too serious, since t will coincide with the true time at least once every revolution at the kicker (s = 0), which is the one point where true time is important. In any case, in most lattices ϕ/Ω will not stray far from true time at any point in the ring. Secondly, we like to think in the transverse deviation x, so we undo the normalisation of the variable η, but again we are only really interested in true deviations at the pickup, so we define a quasi-position variable $\eta\beta_{PU}^{1/2}$ which gives true position once per turn at the pick-up and again we loosely call it x

$$t = \frac{\phi}{\Omega} \simeq \text{true time}\,, \qquad x = \eta\beta_{PU}^{1/2} \simeq \text{true position}\,. \tag{3.11}$$

Using (3.11) in (3.10), we find,

$$\ddot{x}(t) + Q^2\Omega^2 x = \frac{Q^2\Omega^2\sqrt{\beta_K\beta_{PU}}}{2\pi Q}\,\Theta \sum_{\ell=-\infty}^{\infty} e^{i\ell\Omega t}\,. \tag{3.12}$$

Thus the betatron motion is driven by an infinite set of Fourier harmonics of equal amplitude and separated in frequency, one from the other, by the revolution frequency Ω (see Fig. 16).

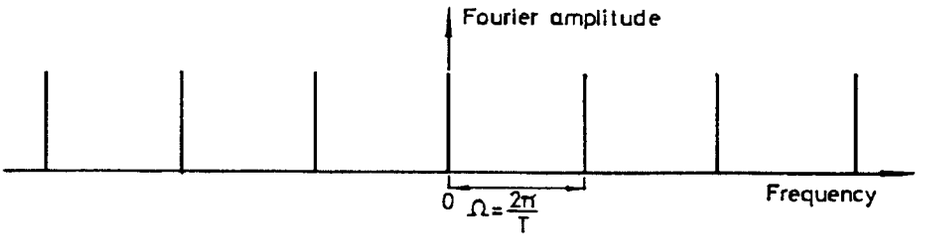


Figure 16: Fourier spectrum of the single constant kick of Fig. 15.

3.4 Transverse RF knockout

In the previous section we analysed the equation of betatron motion with a constant kick. If we now modulate this excitation we can still use the expansion (3.8) of the kick (3.7). We simply have to make the kick strength Θ of Eq. (3.9) a function of time. Just imagine F(s) in Eq. (3.7) multiplied by some modulation factor, say $e^{i\omega t}$. You can keep this factor separate, expand the rest as before and then multiply with the modulation factor. The result is that all Fourier coefficients f_ℓ (Eq. 3.9) are modulated by the same factor. This leads to the phenomenon known as transverse RF knockout. Let us rewrite Eq. (3.12) as:

$$\ddot{x}(t) + \omega_\beta^2 x = V_0 \, e^{i\omega t} \sum_{\ell=-\infty}^{\infty} e^{i\ell\Omega t} \tag{3.13}$$

where

$$V_0 = \frac{Q\Omega^2}{2\pi} \sqrt{\beta_K \beta_{PU}} \Theta_0 \, ,$$

Θ_0 being the amplitude of the kick, $\Theta = \Theta_0 \, e^{i\omega t}$.

Note that $V_0 \, e^{i\omega t}$ uses true time i.e. the time at the kicker. Equation (3.13) may also be written as

$$\ddot{x}(t) + \omega_\beta^2 x = V_0 \sum_{\ell=-\infty}^{\infty} e^{i(\ell\Omega+\omega)t} \, . \tag{3.14}$$

Equation (3.14) is expressed with negative and positive frequencies, which correspond to the slow and fast waves set up by the disturbing kick. Those not familiar with complex voltages and currents as used by electrical engineers may wonder about the significance of negative frequencies. In fact above we assumed a complex excitation $\theta = \theta_0 \, e^{i\omega t}$ of the kicker as this greatly simplifies the algebra. In real life we deal with cosine- rather than $e^{i\omega t}$-type of kicker fields. It is easy to go from the complex to the real world by taking the real parts of Eq. (3.14). Then the r.h.s. contains terms which can be written in the form cos (m $\Omega t + \omega t$), cos ($m_1 \, \Omega t - \omega t$) and — if $\omega > \Omega$ also cos ($\omega t - m_2 \Omega t$) — with positive frequencies only (m: any integer $\geqslant \omega/\Omega$ and m_2 any integer $0 < m_2 < \omega/\Omega$).

Thus the particle 'sees' the frequencies $m\Omega \pm \omega$, i.e. two sidebands spaced by the kicker excitation frequncy ω left and right of each revolution harmonic $m\Omega$. This is illustrated in Fig. 17 where the spectrum of the complex excitation [r.h.s. of Eq. (3.14)]

and its reflection into the real world are sketched. Taking the real part simply corresponds to reflecting the negative frequencies into the positive f-plane.

The revolution sidebands at $m\Omega \pm \omega$ are very similar to the sidebands at $\omega_{osc} \pm \omega_{mod}$ of an amplitude modulated oscillator. For this simpler example the complex and the real analysis are once again summarized in Table 7.

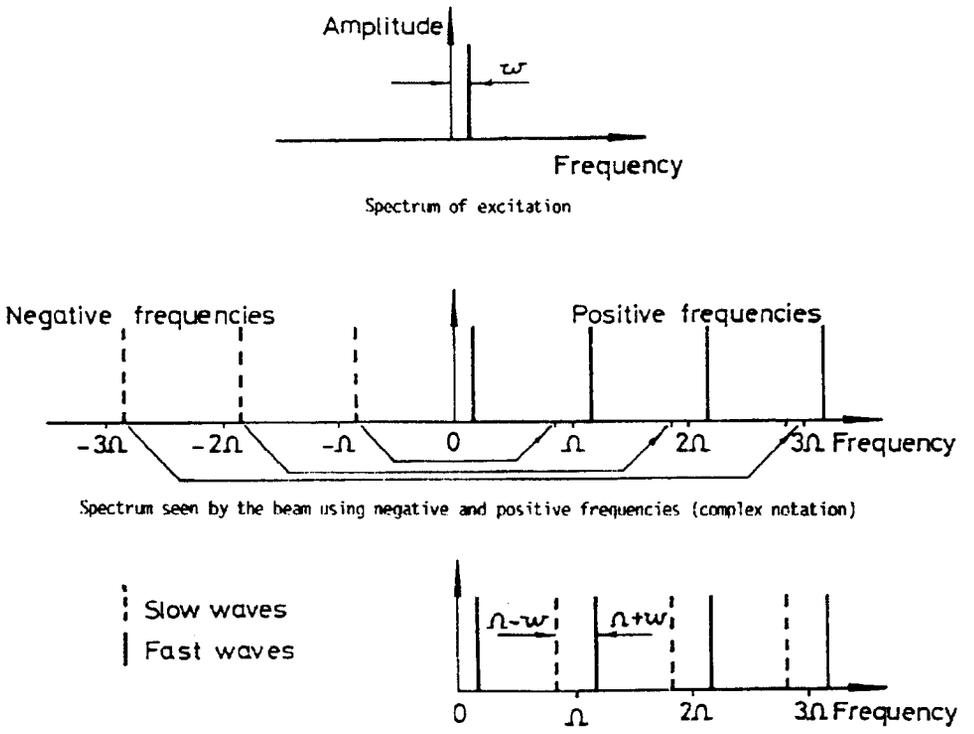


Figure 17: Simple harmonic excitation on a short kicker, spectrum of excitation waveform and of waves seen by a particle.

Table 7: The spectrum of an amplitude modulated oscillator $(\omega_{RF})\omega_m)$

	Complex notation	'Real world'
Modulated amplitude	$a = A\, e^{i\omega_m t}$	$a = A \cos \omega_m t$
Carrier	$c = 0.5(e^{i\omega_{RF} t} + e^{-i\omega_{RF} t})$	$c = \cos \omega_{RF} t$
Position of spectral lines of modulated signal a × c	$(\omega_{RF} + \omega_m)$ and $-(\omega_{RF} - \omega_m)$	$\omega_{RF} + \omega_m$ and $\omega_{RF} - \omega_m$

Returning to Eq. (3.14), clearly if the driving term has a harmonic at the natural frequency of response, ω_β, then the beam will behave resonantly.

For resonance, $(\ell\Omega \pm \omega)^2 = \omega_\beta^2$

i.e.

$$\omega_\pm = \ell\Omega \pm \omega_\beta = \underbrace{(\ell \pm \|Q\|}_{n} \pm q)\Omega \tag{3.15}$$

$$\omega_\pm = (n \pm q)\Omega \tag{3.16}$$

where q is the fractional part and $\|Q\|$ the integer part of the tune. Hence $n = \ell \pm \|Q\|$ is also an integer.

Equation (3.16) shows that the beam will respond resonantly at the 'betatron sidebands' $(n \pm q)\Omega$ centred on the revolution frequency harmonics. Take as an example LEAR at 600 MeV/c with $f_0 \simeq 1$ MHz, $q \simeq 0.3$. Resonant beam response ('RF - knockout') will occur when the kicker is excited at: 0.3 MHz [$= (0 + q)f_0$], or 0.7 MHz [$= (1 - q)f_0$] or 1.3 MHz ....

Figure 18 sketches how two voltages of different frequency (ω and $\omega + \Omega$ respectively) on a short kicker can produce the same series of kicks as seen by a particle. Note the analogy to an RF accelerating cavity which can in principle work at any revolution harmonic.

Having established the beam response to a kicker we shall next analyse the reciprocal problem of the signal response of the pick-up to the betatron oscillation of a particle.

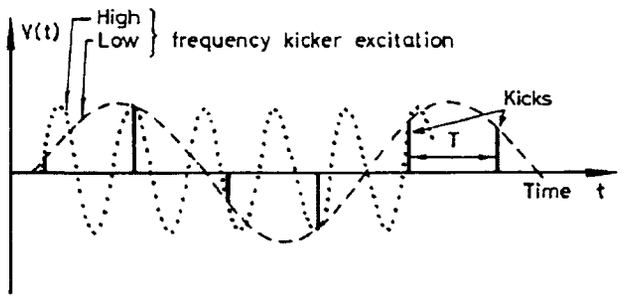


Figure 18: Example of how a beam can be driven in the same way by different frequencies applied from a short kicker. The bars are the kicker voltages at the moment the particle passes, i.e. the kick experienced. The low frequency (ω) and the higher frequency ($\omega + \ell\Omega$) excitation produce the same apparent kick.

3.5 Signals from a circulating particle

A circulating particle passes once per turn through the pick-up and induces a short pulse. This can be represented by a periodic delta function and we have this, now rather familiar, picture in Fig. 19.

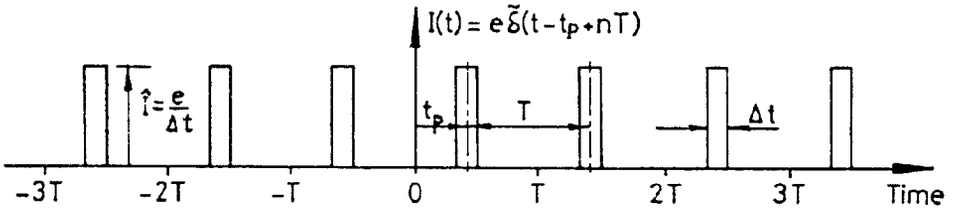


Figure 19: Representation of the pick-up signal from a single circulating particle.

The induced signal is given by,

$$I_s(t) = e\tilde{\delta}(t - t_p + nT) = e \sum_{n=-\infty}^{\infty} \delta(t - t_p + nT) \tag{3.17}$$

accepting that frequency components comparable to or higher than the inverse of the transit time through the pick-up, i.e. $f > 1/\Delta t$, are not required. Here e is the charge per turn $\int_0^T I(t)\, dt$ passing the pick-up, i.e. the charge of the particle.

As before we Fourier analyse the signal in order to replace the discontinuous delta functions by continuous functions

$$I_s(t) = \sum_{m=-\infty}^{\infty} \tilde{I}_m e^{im\Omega(t-t_p)} \tag{3.18}$$

where

$$\tilde{I}_m = \frac{e}{T} = e\frac{\Omega}{2\pi} .$$

The Fourier harmonic amplitudes are constant. (In practice as the harmonic frequency approaches $1/\Delta t$ the amplitudes will decrease.)

The pick-up we use for a betatron cooling system will need to be sensitive to the transverse beam position. This is achieved by placing electrodes on either side of the beam. Each plate will have a signal induced of the form of Eq. (3.18) and its amplitude will be proportional to the distance of the beam particle from the plate. From the difference signal between the two plates the particle's transverse position (x) is obtained. We write the difference signal as $I(t) = I_s \cdot S_p \cdot x/h$, where $I_s(t)$ is the sum signal Eq. (3.18), h is the half aperture of the pick-up and S_p a factor of the order of unity. This factor as well as the assumed linearity depend very much on the construction of the pick-up. For a particle performing betatron oscillations we obtain an induced signal modulated by the transverse betatron motion

$$x = \tilde{x}\, e^{i(Q\Omega t - \mu_0)} . \tag{3.19}$$

This signal is:

$$I_s(t) = \sum_{m=-\infty}^{\infty} S_p \left(\frac{x}{h}\right) \tilde{I}_m\, e^{im\Omega(t-t_p)} \tag{3.20}$$

$$I_s(t) = S_p\, e\frac{\Omega}{2\pi}\frac{\tilde{x}}{h} \sum_{m=-\infty}^{\infty} e^{i\Omega Qt}\, e^{im\Omega(t-t_p)-i\mu_0} .$$

Since only the real part of the betatron motion (3.19) and hence of the current (3.20) interests us, we way write Eq. (3.20) as

$$I_s(t) = S_p \, e\frac{\Omega}{2\pi} \frac{\tilde{x}}{h} \sum_{m=0}^{\infty} e^{i(m\pm Q)\Omega t - im\Omega t_p \pm i\mu_0} \, . \tag{3.21}$$

Looking at the exponent we find again that the particle induces signals at the sidebands,

$$\omega_{PU} = (m \pm \|Q\| \pm q)\Omega = (n \pm q)\Omega \, . \tag{3.22}$$

These are the same as the beam response frequencies (3.16). Thus the beam 'responds' and 'talks' at the same frequencies.

Figure 20 shows the time and frequency domain picture of the pick-up signal of a single particle. This signal [Eq. (3.20)] will be used in calculating the coherent effect of the test particle upon itself. The incoherent effect due to the other particles can be obtained by adding up their currents (3.20) with a proper distribution in amplitude and phase.

In a coasting beam of N particles with random initial betatron phase and random time of arrival these induced currents add in square to give noise like signals at the frequencies (3.22).

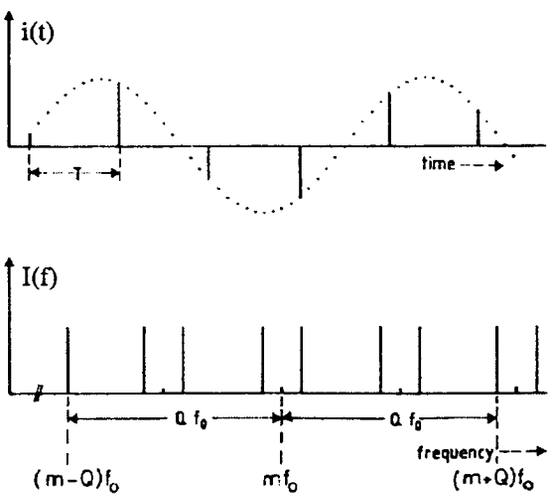


Figure 20: Time and frequency domain signal of a particle performing a betatron oscillation. A position sensitive pick-up records a short pulse at each traversal modulated in amplitude by the betatron oscillation. The frequency spectrum contains lines at the two sideband frequencies $(m \pm Q)f_0$ of each revolution harmonic mf_0.

As different particles have slightly different revolution and betatron frequencies, these signals occur in bands with a spectral power density

$$\frac{dI_{sc}^2}{d\omega} = \frac{1}{2}\left(S_p \frac{e}{T} \frac{x_{rms}}{h}\right)^2 \frac{dN}{d\omega} \tag{3.23}$$

where $dN/d\omega$ is the fraction of particles with sideband frequencies in a range of width $d\omega$ around $\omega = (n \pm q)\Omega$. These are the Schottky noise bands discussed e.g. in [3]. With a

dispersion of revolution and betatron frequencies ($\Delta\Omega$ and Δq respectively) the width of the band at $(n \pm q)\Omega$ is $(n \pm q)\Delta\Omega \pm \Omega\Delta q$. A spectrum analyser usually records current, i.e. the square root of the signal (3.23). A practical example is given in Fig. 21, where the spectrum analyser picture of the signal from a horizontal pick-up is shown. The frequency band of this 'Schottky scan' is centered around a revolution harmonic nf_0 and contains the two sidebands $(n \pm q)f_0$. Note that the height of these sidebands [root of Eq. (3.23)] is proportional to the r.m.s. betatron amplitude x_{rms} and thus decreases during cooling.

All pick-up currents discussed here are the 'induced currents'. To obtain the true output signal one has to include the response functions of the pick-up structure and the acquisition system. Usually one aims at making these response functions as flat as possible.

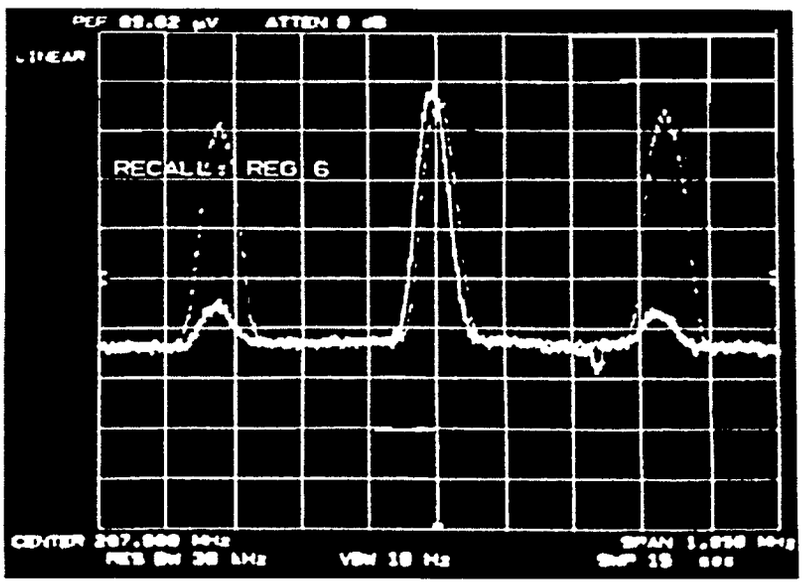


Figure 21: Example of a horizontal Schottky scan in LEAR at 600 MeV/c. The central band, the harmonic n = 100 of the revolution frequency, is visible as the beam is not completely centred at the position pick-up. The right and left bands are the sidebands $(98 + Q)f_0$ and $(103-Q)f_0$ where $Q \approx 2.3$. During emittance cooling the sidebands decrease. The difference between the base line of the trace and the bottom line (zero signal) is given by the noise of the pick-up system. The span covers (approximately) half a revolution interval f_0. During horizontal cooling the height of the sidebands decreases.

3.6 Coherent effect

We can now have a fresh look at the motion of particle i, retaining for the moment only its 'self-terms'. We take a kicker voltage on the r.h.s. of Eq. (3.13) which is proportional to the pick-up signal [Eq. (3.20)] of the circulating particle. As before we refer to all signals at their time of arrival at the kicker.

The particle takes a nominal time t_p and the cooling signal a time t_c to travel from pick-up to kicker. The electronic delay t_c is in general frequency dependent. We include this by a phase factor $\varphi(\omega_m) = m\Omega t_c(\omega_m)$ in the development (3.20) of the pick-up signal. The frequencies ω_ℓ contained in this signal are the $(n \pm q)\Omega$ betatron sidebands (3.22).

In our complex notation (including positive and negative frequencies) they simply appear as

$$\omega_m = (m + Q)\Omega\ ; \quad m = -\infty \text{ to } \infty\ . \tag{3.24}$$

The point to retain is that all signal transmission occurs at the frequencies of Eq. (3.24).

In addition to the initial phase μ_0 [Eq. (3.19)], we have to include the phase-advance $\mu = \omega_\beta t_p$ of the betatron oscillation of the particle on its way from pick-up to kicker. Hence the exponential factor in Eq. (3.20) referred to the kicker is written as,

$$e^{i[\omega_\beta t - \mu + \mu_0 + \varphi(\omega_m) + m\Omega(t - t_p)]}\ .$$

To complete the driving term in the betatron equation (3.13) we introduce a transfer function $G(\omega_m)$. It has to include the pick-up response $Z_p(\omega)$ [i.e. the voltage output for the induced current (3.21)], the transfer function of the cooling loop between pick-up and kicker (with cables, amplifiers, filters, etc. taken into account) as well as the kicker response. Let us for simplicity also absorb the constant factors

$$\frac{Q^2\Omega}{2\pi}\sqrt{\beta_k \beta_{PU}}\theta_0$$

of Eq. (2.13) and the factor

$$S_p\, e \frac{\Omega}{2\pi}\, \frac{1}{h}$$

of Eq. (3.20) into this transfer function but keep $x_i = \tilde{x}_i\, e^{i(\omega_\beta t + \mu_0)}$ separate. Hence we rewrite Eq. (3.14) as

$$\ddot{x}_i + \omega_\beta^2 x_i = x_i \sum_{m=-\infty}^{\infty} G_{ii}(\omega_m) \cdot e^{-i\mu + im\Omega t + i\varphi(\omega_m) - im\Omega t_p} \cdot \sum_{\ell=-\infty}^{\infty} e^{i\ell\Omega t} \tag{3.25}$$

We take $G_{ii}(\omega_m)$ as entirely real and include all phase shifts in $\varphi(\omega_m)$. Note that the second sum in Eq. (3.25) is the 'sampling term' — appearing already in Eq. (3.13) — due to the fact that the particle passes the short kicker once per turn. The first sum clearly is due to the localised nature of the pick-up.

Equation (3.25) is almost the same as Eq. (3.2) except that we include frequency dependent 'gain' $G_{ii}(\omega) \cdot e^{i\varphi(\omega)}$ and localised pick-up and kicker. The product of the two sums can readily be converted into a double sum

$$\sum_m e^{im\Omega t} \cdot \sum_\ell e^{i\ell\Omega t} = \sum_m \sum_\ell e^{i(m+\ell)\Omega t}\ , \tag{3.26}$$

noting that in general

$$\sum a_m \cdot \sum b_\ell = \sum_\ell a_1 \cdot b_\ell + \sum_\ell a_2 \cdot b_\ell \ldots = \sum_m \sum_\ell a_m \cdot b_\ell\ .$$

Equation (3.25) may now be interpreted as an oscillator with a frequency shift that varies in time. An approximate solution to such equations is obtained by taking the time

average of the frequency shift only, i.e. if we retain terms with $\ell = -m$ in Eq. (3.26) and drop the rapidly oscillating frequency shifts. Using this approximation (3.25) becomes:

$$\ddot{x}_i + \omega_\beta^2 x_i = x_i \sum_{m=-\infty}^{\infty} G_{ii}(\omega_m) e^{-i\mu + i\varphi(\omega_m) - im\Omega t_p} \ . \tag{3.27}$$

This defines a change of betatron frequency:

$$\Delta\omega_\beta \approx \frac{1}{2\omega_\beta}\Delta(\omega_\beta^2) = \frac{-1}{2\omega_\beta} \sum_{m=-\infty}^{\infty} G_{ii}(\omega_m) e^{-i\mu + \varphi(\omega_m) - im\Omega t_p} \ .$$

As we assumed $x = \tilde{x}e^{i\omega_\beta t}$ the damping rate is

$$1/\tau = \mathrm{Im}(\Delta\omega_\beta) = \frac{-1}{2\omega_\beta}\mathrm{Im} \sum_{m=-\infty}^{\infty} G_{ii}(\omega_m) \cdot e^{-i\varphi(\omega_m) - i\mu - m\Omega t_p} \ . \tag{3.28}$$

Optimum cooling is given by

$$1/\tau = \frac{1}{2\omega_\beta} \sum_{m=-\infty}^{\infty} |G_{ii}(\omega_m)| \tag{3.28a}$$

and obtained for all m, if the phase factor is properly chosen such that $-\mathrm{Im}(e^{i\cdots}) = 1$. This requires

$$\varphi(\omega_m) - \mu - m\Omega t_p = -\frac{\pi}{2}/\text{modulo } 2\pi \ . \tag{3.29}$$

Usually Eq. (3.29) is satisfied by putting the kicker at the 'proper' betatron phase advance from the pick-up, i.e.

$$\mu = \frac{\pi}{2} \ , \qquad \text{or} \qquad \frac{3\pi}{2} \qquad \text{with signal inversion,} \qquad \text{or} \qquad \frac{5}{2}\pi ...$$

and designing $\varphi(\omega_m)$ to be as close as possible to $m\Omega t_p$:

$$\varphi(\omega_m) - m\Omega t_p = m\Omega[t_c - t_p] = 0 \ . \tag{3.30}$$

Ideally this requires a signal delay of the cooling loop, $t_c(\omega_m)$, equal to t_p, the particle travelling time pick-up to kicker, independent of frequency. If the optimum spacing pick-up to kicker ($\mu = \pi/2$ mod. π) is not possible one can in principle include filters (as proposed by Thorndahl) with a time delay characteristic $t_c(\omega)$ such that Eq. (3.29) is still satisfied. This requires however 'steep' filters with a phase delay $\varphi(\omega)$ varying by $2(\mu - \pi/2)$ from the n + q to the n − q betatron band. In fact denoting the deviation from the optimum phase advance $\mu = \frac{\pi}{2}$ by $\delta\mu$ we can write Eq. (3.29) in the form

$$\varphi(\omega_m) - m\Omega t_c = \left(\mu - \frac{\pi}{2}\right) = \delta\mu \ .$$

As for any network $\varphi(-\omega) = -\varphi(\omega)$ we need for positive m (corresponding to the n + q sidebands):

$$\varphi(\omega) - m\Omega t_p = \delta\mu$$

and for negative m (corresponding to the n − q bands)

$$\varphi(\omega) - |m|\Omega t_p = -\delta\mu$$

thus requiring a phase difference $2\delta\mu$ between neighbouring bands.
Other effects can be identified from Eq. (3.28):

- The time-of-flight error Δt_p of a particle ('mixing between pick-up and kicker') as well as improper delay Δt_c of the cooling loop or improper pick-up to kicker betatron phase advance $\delta\mu = \mu - \pi/2$ appear as a phase factor in Eq. (3.28) which may be rewritten as

$$\frac{1}{\tau} = \frac{1}{2\omega_\beta} \sum_{\ell=-\infty}^{\infty} G_{ii}(\omega_m) \cdot \cos(\delta\varphi_m - \delta\mu) \qquad (3.28b)$$

where $\delta\varphi_m = \varphi(\omega_m) - m\Omega t_p = m\Omega(t_c - t_p)$.

- The fall-off of the pick-up-current spectrum at high frequency can be included in the expansion (3.21) and absorbed into the transfer function. The similar effect of the finite kicker length can be included in much the same manner via the expansion of the r.h.s. of Eq. (3.13). Finally we remark that Eq. (3.28) can be written in various other forms involving sums over positive m only which clearly reveal the (n ± q) bands. This is left as an exercise to those interested.

Having established the interaction of the particle with itself we next include the noise due to the other particles and the electronic system.

3.7 Noise

Noise is treated in detail in G. Dôme's chapter in the Proceedings of the Oxford course (CERN 87–03 p. 110), see also [12]. For convenience we repeat the essentials here, which are useful to include the incoherent effect in the frequency-domain analysis of stochastic cooling.

Look at an oscilloscope picture like Fig. 22 which displays a pick-up signal u(t) when 'no beam is in the machine', i.e. the electronic noise of the system. It is customary to represent the mean square (averaged over a long enough time T) of such noisy voltages by a pseudo Fourier transformation

$$\overline{u^2(t)} = \int_{-\infty}^{\infty} \phi(\omega) d\omega \, .$$

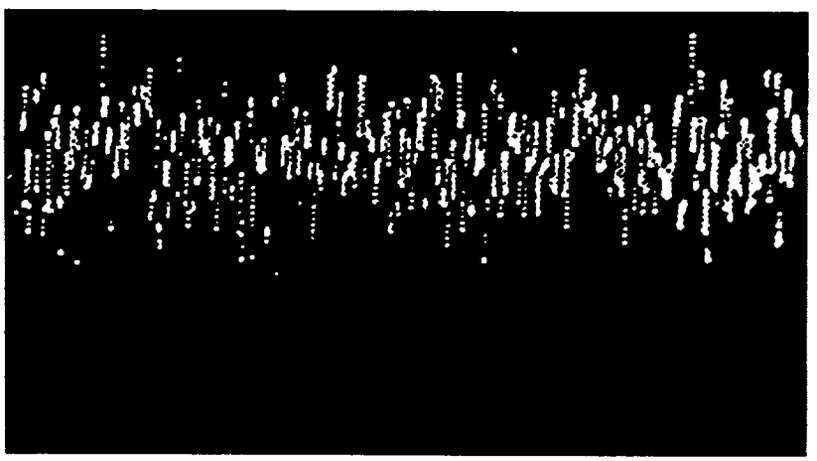


Figure 22: Noise signal on a pick-up.

The 'spectral power density function' $\phi(\omega)$ is closely related to a Fourier development of u(t). In all practical applications the noisy voltage has been 'switched on' at some time t = 0 and we regard it up to t = T. Outside this range the waveform is irrelevant, so, for the purpose of computation we can periodically continue it (Fig. 23). We then deal with a periodic function $u(t \pm nT) = u(t)$ which we can Fourier-expand in the usual way

$$u(t) = \sum_{m=-\infty}^{\infty} \tilde{u}_m e^{im\omega_0 t} ; \qquad \omega_0 = 2\pi/T . \tag{3.31}$$

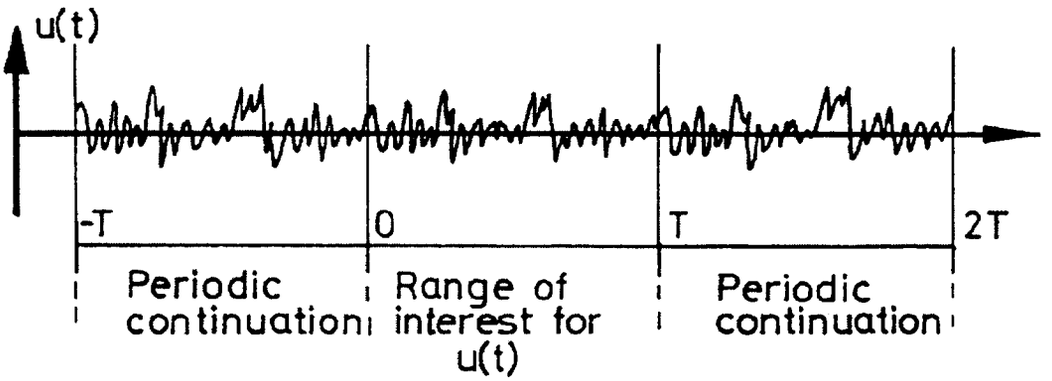


Figure 23: A noisy voltage u(t) observed from time t = 0 to t = T and its periodic continuation outside this range to permit a Fourier development.

The Fourier amplitudes:

$$\tilde{u}_m = \frac{1}{T}\int_0^T u(t)e^{-im\omega_0 t}dt \tag{3.32}$$

are in general complex but for real u(t) u_{-m} is the conjugate of u_m. The mean square of Eq. (3.31) over the observation time T is by definition

$$\overline{u^2(t)} = \frac{1}{T}\int_0^T u^2(t)dt \tag{3.33}$$

which yields after some calculation (transforming the square of the sum into a double sum similar to the analysis in conjunction with Eq. (3.26) and noting that averaged over a period all $e^{ik\omega_0 t}$ terms vanish except for k = 0)

$$\overline{u^2(t)} = \sum \tilde{u}_m \cdot \tilde{u}_{-m} = \sum_{m=-\infty}^{\infty} |\tilde{u}_m|^2 . \tag{3.34}$$

This is known as Parseval's equation in the theory of Fourier series; it applies to any Fourier development! Equation (3.34) presents the 'average noise power' $u^2(t)$ as the sum of its spectral contributions at frequencies $\omega_0 = n2\pi/T$. Analysed over shorter or longer time T the spectra are as sketched in Fig. 24.

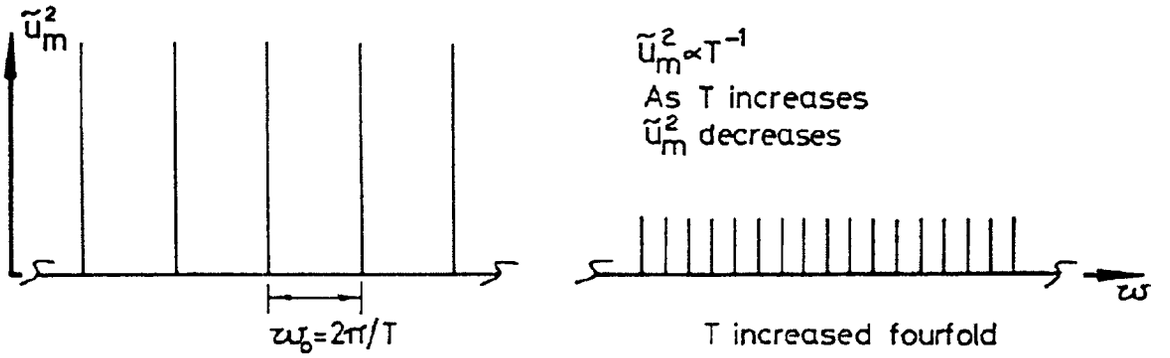


Figure 24: The power spectrum $|u^2(\omega)|$ of a noisy voltage which is observed for a time interval T and periodically continued outside this interval. Increasing T the height of the spectral lines decrease proportionally to their spacing $\omega_0 = 2\pi/T$ so that the quantity $|u^2(\omega)|/\omega_0$ remains the same. In the limit $T \to \infty$ one has a continuous spectrum where $u^2(\omega)/\omega_0 \to \phi(\omega)$ is the spectral power density function of Eq. (3.30).

As the sum of the rays ('height x number') is $\overline{u^2(t)}$ in both cases their height scales proportional to their spacing (for large T), i.e. $u_m^2/\omega_0 = \text{const}$.

For very large observation time $T \to \infty$ the spectrum is practically continuous and the sum Eq. (3.34) is approximated by an integral of the form of Eq. (3.30):

$$\overline{u^2_{(t)}} = \sum_{m=-\infty}^{\infty} |\tilde{u}_m|^2 = \sum \frac{|u_m|^2}{\omega_0}\Delta\omega \to \int_{-\infty}^{\infty} \frac{|\tilde{u}_{(\omega)}|^2}{\omega_0} d\omega \,. \tag{3.35}$$

Hence we identify for $T \to \infty$

$$\frac{|\tilde{u}_{(\omega)}|^2}{\omega_0} \to \phi(\omega) \,. \tag{3.36}$$

This interpretation permits us to calculate (at least in simple cases) $\phi(\omega)$ and to establish the following important theorem:
When noise with a spectral power density $\phi_1(\omega)$ is transmitted through a linear system with a (complex) transfer function $H(\omega)$ then the power spectrum at the output is

$$\phi_2(\omega) = |H(\omega)|^2 \phi_1(\omega) \,. \tag{3.37}$$

This follows immediately from the preceding noting that each of the components of the r.h.s. of Eq. (3.31) when transmitted through the network transforms according to Eq. (3.36):

$$\tilde{u}_{m_2} = \tilde{H}(m\omega_0) \cdot \tilde{u}_{m_1} \,.$$

An example of the theorem (3.37) is the transformation of 'broadband noise' into band limited noise by a band pass filter (Fig. 25).

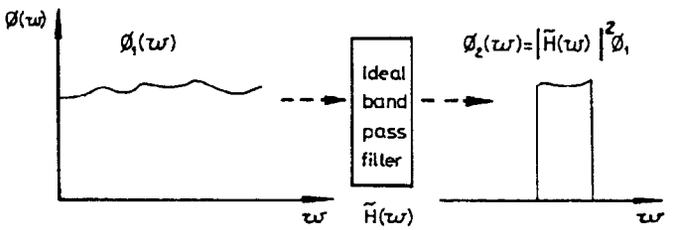


Figure 25: An example of the theorem [Eq. (3.37)]: broadband noise transmitted through an ideal band pass filter is converted into a band-limited noise.

This is as much as we need about noise for the purpose of this chapter.

3.8 Beam response to a noisy kicker

Consider a linear oscillator driven by a noisy excitation u(t) with power density $\phi(\omega)$:

$$\ddot{x} + \omega_\beta^2 x = u(t) \ . \tag{3.38}$$

This problem was treated (in a more general context) in a classical paper by Hereward and Johnsen [12].

Their result, the 'Hereward–Johnsen theorem' may — in our present case — be stated as follows:

For a particle injected at $t = 0$, the square of the amplitude $\tilde{x}$ of x, Eq. (3.38) expected (Fig. 26) at time t (t large) is

$$\tilde{x}^2 = \frac{2\pi}{\omega_\beta^2} \cdot \phi(\omega_\beta) \cdot t \ . \tag{3.39}$$

In words: the amplitude grows in a diffusion-like manner ($\tilde{x} \propto \sqrt{t}$) at a rate which is determinated by the spectral density of the noise *at the resonance frequency* ω_β.

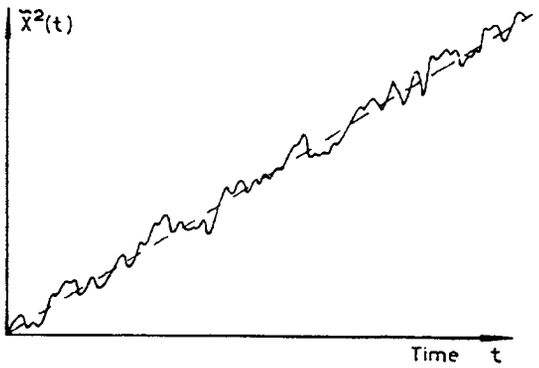


Figure 26: Amplitude $\tilde{x}$ of betatron oscillation of a particle driven by a noisy kicker. The expectation value (i.e. the average) of x^2 grows linearly in time at a rate given by the spectral density of the noise at the resonance frequency. In addition to this average growth there is a fluctuating motion which is of little importance for the long term behaviour.

Equation (3.39) is for a simple harmonic oscillator. If we inject noise into the cooling loop (or directly onto the kicker) we have again to include the 'sampling factor' $\sum e^{i\ell\Omega t}$ of Eq. (3.13) because we use a short localized kicker. Thus we use

$$\ddot{x} + \omega_\beta^2 x = u(t) \sum_{\ell=-\infty}^{\infty} e^{i\ell\Omega(t-t_p)} \ . \tag{3.40}$$

The effect of each component, $e^{i\ell\Omega t}$, is to 'shift' the frequency content of the driving force $\omega \to \omega + \ell\Omega$. In this sense we may interpret the r.h.s. of Eq. (3.40) as a sum of noisy driving forces with frequencies

$$\omega_d = \omega + \ell\Omega \ .$$

We can apply the Hereward–Johnsen result to each of these bands (noting that in working out $x^2(t)$ cross-terms between bands average to zero), hence the response of Eq. (3.40) to a noise u(t) is:

$$\tilde{x}^2 = \frac{2\pi}{\omega_\beta^2} t \sum_{\ell=-\infty}^{\infty} \phi(\omega_\beta + \ell\Omega) \ . \tag{3.41}$$

Equivalently if you prefer to work with positive frequencies only you may write [using $\phi(-\omega) = \phi(\omega)$], Eq. (3.41) as

$$\tilde{x}^2 = \frac{2\pi}{\omega_\beta^2} t \left[\sum_{\ell=0}^{\infty} \phi(\ell + q)\Omega + \sum_{\ell=1}^{\infty} \phi[\ell - q)\omega] \right] \ . \tag{3.42}$$

This clearly presents the amplitude growth in terms of the spectral density of the noise at the betatron sidebands.

In working out the long term average of $x^2(t)$ leading to Eq. (3.41) and Eq. (3.42) we have assumed that cross terms between different bands ℓ average to zero and also that $\phi(-\omega) = \phi(\omega)$. A justification [which can be carried through, e.g. using an expansion of the type (3.31) for u(t) and the definition (3.36) of $\phi(\omega)$] is left as an exercise to those interested.

3.9 Back to beam Schottky noise and amplifier noise

Returning to a test-particle: apart from its self-term it will experience the 'beam Schottky noise' due to the presence of the other particles and the electronic noise of the preamplifier etc. The spectral density $\phi(\omega) = dI_{sc}^2/d\omega$ inside a Schottky band is determined by Eq. (3.23) and may be rewritten here (see D. Boussard's chapter on Schottky noise [3]):

$$\begin{aligned} \phi_{sc}(\omega) &= \phi_0 M(\omega_\ell) \\ \omega_\ell &= (\ell \pm q)\Omega \ , \qquad \ell = 1, 2, 3... \\ \phi_0 &= \frac{N\, e^2 \Omega\, x_{rms}^2\, S_p^2}{(2\pi)^2\, h^2} \end{aligned} \tag{3.43}$$

$$M(\omega_\ell) = \begin{cases} \frac{\Omega}{2N} \frac{dN}{d\omega_\ell} \approx \frac{\Omega}{2\Delta\omega_\ell} & \text{if all bands are separated, i.e. } \Delta\omega_\ell < 2q\Omega \\ 1 & \text{for complete band overlap, i.e. } \Delta\omega_\ell > \Omega \ . \end{cases}$$

Here x_{rms}^2 is the mean-square betatron amplitude of particles with a sideband frequency near ω_ℓ, $dN/d\omega_\ell$ is the number of particles with frequency in a range of width $d\omega$ near ω_ℓ.

The dimensionless quantity $M(\omega)$ is closely related to mixing as will become clear later. It is customary to approximate $dN/d\omega$ by a rectangular distribution $N/\Delta\omega$ of total width

$$\Delta\omega(\ell) = \Delta[(|\ell| \pm q)\Omega] = (|\ell| \pm q)\Delta\Omega \pm \Omega\Delta q \approx |\ell|\Delta\Omega \tag{3.44}$$

where $\Delta\omega(\ell)$ is the width of the sidebands at the ℓ-th harmonic. This leads to the approximations for $M(\omega_\ell)$ indicated under Eq. (3.43) and sketched in Fig. 27.

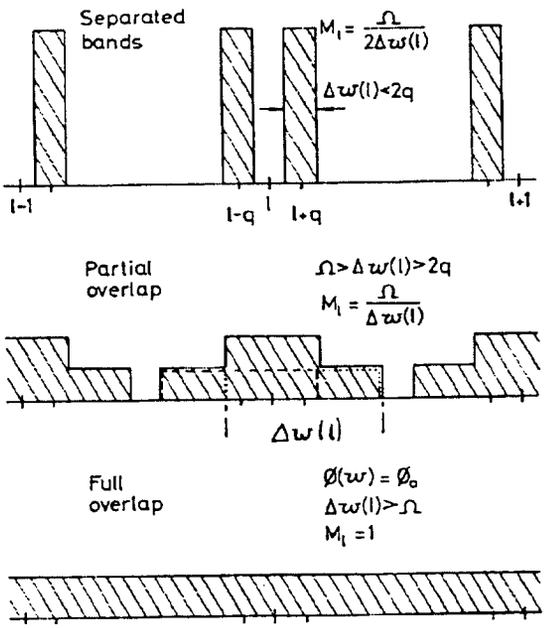


Figure 27: Transverse beam Schottky noise for separated bands (low frequency), partial overlap (intermediate frequency) and complete band overlap.

For high frequencies, bands completely overlap and the noise has a continuous power spectrum with density $\phi(\omega) = \phi_0$ as given by Eq. (3.43) with $M = 1$. Note that in this limit Eq. (3.43) is just the classical Schottky formula for the noise of a 'DC-current' I_0. Defining the noise density $\phi = dI^2/d\omega$ with respect to angular frequency ω as we consistently do in the present chapter, this formula writes as:

$$\phi(\omega) = e\ I_0/\pi\ . \tag{3.45}$$

In our case the circulating current is $I_0 = Ne\Omega/2\pi$ and $0.5\ (S_p x_{rms}/h)^2$ enters as we take the betatron oscillation signal from a difference pick-up.
The noise is transmitted through the cooling loop in the same way as the 'self-signal' of the test-particle. Hence the transfer function is the same except for phase factors due to the different arrival times of particles and due to different betatron phase. We denote this transfer function, which has the same modulus as $\tilde{G}_{ii}(\omega)$ and $\tilde{G}_{ij}(\omega)$, by $G(\omega)$.

By virtue of Eq. (3.37) the noise density on the kicker is therefore

$$\phi_p(\omega) = |G(\omega)|^2 \phi_0'(\omega) M(\omega) \tag{3.46}$$

$$\phi_0' = \frac{N}{\Omega} x_{rms}^2\ .$$

Here ϕ_0' is the same as ϕ_0, Eq. (3.43) except for the factor $(S_p e\Omega/2\pi h)^2$ which was absorbed into $G^2(\omega)$ as before, see the discussion preceding Eq. (3.25).

We now turn to the electronic noise and assume that — referred to at the entrance of the cooling loop (exit of the pick-up) — it has a power spectrum $\phi_a(\omega)$. Let the transfer function from this point to the kicker be $H(\omega)$. Clearly this $H(\omega)$ is the same as $G(\omega)$ except for the pick-up response function. The noise seen by the test particle (j) (Fig. 28) is then

$$\phi(\omega_j) = \underbrace{|G(\omega_j)|^2 \phi_0' M(\omega_j)}_{\substack{\text{Schottky noise} \\ \text{seen by particle}}} + \underbrace{|H(\omega j)|^2 \phi_a'(\omega_j)}_{\substack{\text{amplifier noise} \\ \text{seen by particle}}} \tag{3.47}$$

Here $\phi_a' = \phi_a/(S_p e\Omega/2\pi h)^2$ is the electronic noise reduced by the same factor as ϕ_0 for consistency. To recover previous results it is useful to rewrite Eq. (3.47) as

$$\phi(\omega) = |G(\omega)|^2 \phi_0' [M(\omega) + U(\omega)] \tag{3.47a}$$

$$U_{(\omega)} = \frac{|H_{(\omega)}|^2}{|G(\omega)|^2} \frac{\phi_a(\omega)}{\phi_0} .$$

Clearly $U(\omega)$ is the ratio of amplifier to beam noise (the latter in the high frequency limit where bands overlap) including the ratio of the two transfer functions $|\, H(\omega)\,|^2 \,/\, |\, G(\omega)\,|^2$. Using Eq. (3.41) we can write down the expectation value for the amplitude of the test particle as driven by the noise Eq. (3.47)

$$\tilde{x}_j^2(t) = \frac{2\pi}{\omega_\beta^2} t\, \phi_0' \sum_{\ell=-\infty}^{\infty} |G(\omega_\ell)|^2 [M(\omega_\ell) + U(\omega_\ell)] \tag{3.48}$$

with ϕ_0' as given under Eq. (3.46), $M(\omega)$ as given under Eq. (3.43), $U(\omega_j)$ as given under Eq. (3.47a) and $\omega_\ell = (\ell + q_j)\Omega_j$ the sideband frequency of particle j.

Equation (3.48) determines the 'incoherent effect' as experienced by the test particle.

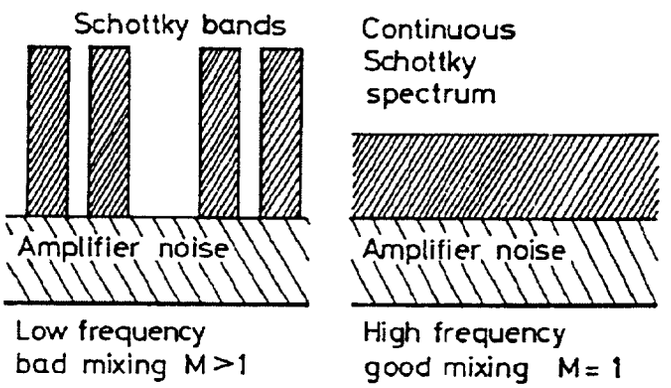


Figure 28: The noise seen by a test particle is the Schottky noise due to the other particles and the electronic noise of the amplifier etc. At low frequency the Schottky noise occurs in bands with a density M times higher than in the situation of complete overlap. This increase of noise density corresponds to enhanced 'heating of the test particle due to bad mixing'.

3.10 Cooling rate

We can now calculate the expected amplitude $\tilde{x}^2 j(t)$ of the test particle (j) by adding up the coherent and the incoherent effects. To use Eq. (3.28) for $d(\tilde{x}^2)/dt$ we note that in general

$$\frac{1}{x^2}\frac{d(x^2)}{dt} = \frac{2}{x}\frac{dx}{dt} = -\frac{2}{\tau} .$$

Hence we have from Eq. (3.28) for the coherent (damping) effect:

$$\left.\frac{d(\tilde{x}_j^2)}{dt}\right|_{coh} = -\tilde{x}_j^2 \frac{1}{\omega_\beta} \sum_{\ell=-\infty}^{\infty} \mathrm{Re}\,[G(\omega_\ell)\ \underbrace{e^{i\delta\varphi(\omega_\ell)}}_{1}\ \underbrace{^{-i\delta\mu}]}_{2} \tag{3.49}$$

For the incoherent (heating) effect we rewrite Eq. (3.48) substituting ϕ_0' form Eq. (3.46):

$$\left.\frac{d(\tilde{x}_j^2)}{dt}\right|_{incoh} = \tilde{x}_{rms}^2 \frac{\pi}{\omega_\beta^2}\frac{N}{\Omega} \sum_{\ell=-\infty}^{\infty} |G(\omega_\ell)|^2\ \underbrace{[M(\omega_\ell)}_{3}\ +\ \underbrace{U(\omega_\ell)]}_{4} . \tag{3.50}$$

The resultant cooling equation is:

$$d(x_j^2)/dt = d(x_j^2)/dt|_{coh} + d(x_j^2)/dt|_{incoh} . \tag{3.51}$$

Equations (3.49) to (3.51) represent cooling as a sum of the contributions at the sideband frequency $\omega_\ell = (\ell + q_j)\Omega_j$ of the particle. In this form all frequency characteristics of the cooling loop can readily be included. This is especially handy for those who like to measure and calculate in the frequency domain. In addition we can rediscover and re-interpret the effects discussed before:

1. The influence of imperfect synchronisation of particle and cooling signal ('mixing pick-up to kicker'), $\delta\varphi \neq 0$. It enters as a phase error in the coherent term.

2. The influence of betatron phase errors (imperfect spacing) pick-up to kicker, $\delta\mu \neq 0$. It enters as another phase error in the coherent term.

3. Imperfect mixing on the way kicker to pick-up; expressed here as enhancement ($M > 1$) of the heating by Schottky noise which is concentrated in bands and hence increased in density. Good mixing ($M = 1$) corresponds to overlap of Schottky bands.

4. Amplifier (and other electronic) noise, $U \geqslant 0$.

Equation (3.49) so far is for any test particle. To obtain the damping rate for the mean square amplitude x_{rms}^2 we have to average Eq. (3.49) and Eq. (3.50) over the frequency distribution of the beam particles. In the simple case of perfect $\delta\varphi = 0, \delta\mu = 0$, $M = 1$ and constant $G(\omega) = G, U(\omega) = U$ inside the passband you can rediscover the familiar

$$-\frac{1}{x_{rms}^2}\frac{d(x_{rms}^2)}{dt} = \frac{W}{N}[2g - g^2(1+U)] \tag{3.52}$$

by calling

$$\frac{G \cdot N}{\omega_\beta \Omega/2\pi} = g . \tag{3.53}$$

To work out the sums over ℓ, note that with a passband of width $\Delta f = W$ in the positive frequency plane the number of betatron lines contributing is $2 \cdot 2\pi W/\Omega$ as sketched in Fig. 29, namely $2\pi W/\Omega$ for positive ℓ (the n + q bands) and the same number again for negative ℓ (the n–q bands).

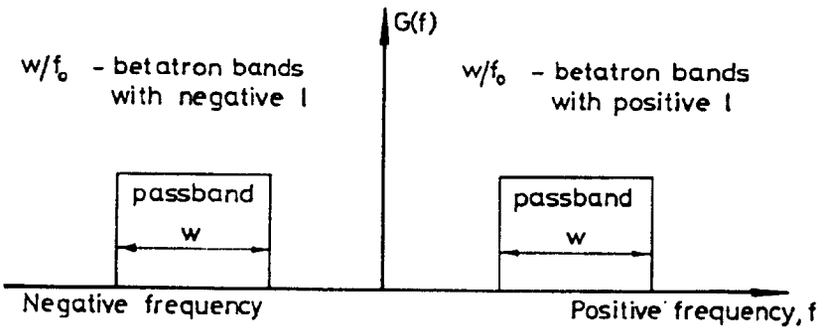


Figure 29: Passband of cooling system in the $f = -\infty$ to $f = \infty$ frequency plane. There are $2\pi W/\Omega = W/f_0$ lines $\omega_\ell = (\ell + g)\Omega$ in the passband at negative frequencies (negative ℓ) and the same number in the passband at positive frequency (positive ℓ).

You can generalise Eq. (3.52) to include mixing factors, betatron phase errors and frequency dependence in G and U by interpreting the sums in Eq. (3.49) as averages over the passband. With this interpretation you may write Eq. (3.52) in various different forms useful for comparison with previous results, for instance:

$$\frac{d(\tilde{x}^2)}{dt} = \frac{W}{N}\left(-2\tilde{x}^2\langle \mathrm{Re}[g(\omega)e^{i(\delta\varphi-\delta\mu)}]\rangle_{\text{passband}} + \tilde{x}^2_{\text{rms}}\langle|g^2(\omega)|[M(\omega)+U(\omega)]\rangle_{\text{passband}}\right) \tag{3.52a}$$

with

$$g(\omega) \equiv \frac{N \cdot G(\omega)}{\omega_\beta \Omega/2\pi}$$

$$\langle \ldots \rangle_{\text{passband}} = \frac{\Omega/2\pi}{2W}\sum_{\ell=-\infty}^{\infty} \ldots$$

3.11 Feedback via the beam and signal shielding

We shall now attempt to introduce a final ingredient of cooling theory known as 'feedback via the beam' or 'signal shielding'. Although this refinement will change our previous results by at most a factor of 2, the change of the beam Schottky signals when the cooling loop is closed has become an important diagnostics tool [13].

Where did we miss out this effect in our treatment so far? In fact considering the test-particle equation

$$\ddot{x}_j + \omega_\beta^2 x_j = \underbrace{G_{jj}x_j}_{\text{effect of particle upon itself : coherent term}} + \underbrace{\sum_i G_{ij}x_i}_{\text{effect of other particles : Schottky noise, fluctuating term with zero average}} + \underbrace{\text{'system noise'}}_{\text{amplifier noise etc. fluctuating with zero average}} \qquad (3.54)$$

O.K. Not O.K. O.K.

We have described the effect of the other particles — $\sum G_{ij}x_i$ — as Schottky noise of an undisturbed beam, i.e. as a fluctuating term with zero time average. This assumption is not generally correct. F. Sacherer [14] has pointed out that — in the case of poor mixing — $\sum G_{ij}x_i$ does lead to a coherent oscillation with finite average amplitude. The fluctuation occurs around this average amplitude and not around zero as it would be case in an undisturbed beam. The reason is that part of the 'modulation' imposed at the kicker is still present at the pick-up and re-enters the loop as sketched in Fig. 30.

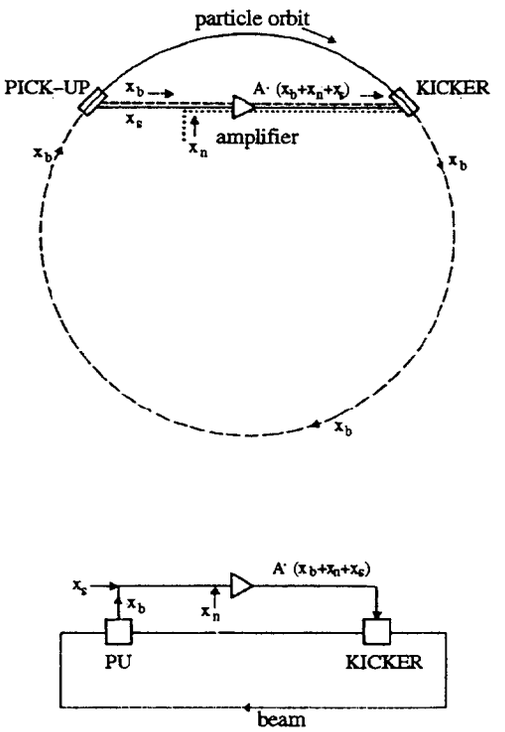


Figure 30: Cooling system including the coherent beam modulation x_b imposed at the kicker and partially preserved up to the pick-up due to imperfect mixing. The lower diagram shows Sacherer's equivalent feedback loop. Amplifier noise (x_n) and Schottky noise (x_s) are random noises whereas the coherent modulation is fed back via the beam from kicker to the pick-up. This feedback changes the open loop response to $x_n + x_s$ by a complex transfer function $T(\omega)$ which depends on the amplification (cooling strength) and the degree of mixing between kicker and pick-up.

Thus the noise on a beam subject to cooling is different from the free beam noise. The feedback of the cooling signals via the beam changes all ingredients of the analysis, namely beam noise as well as the influence of the coherent term and the amplifier noise.

Fortunately F. Sacherer has also shown the road to rescue our previous results. The way out he provides is a beautiful piece of accelerator theory.

As a pre-exercise: consider a system of N oscillators with a harmonic driving force and a 'collective force' proportional to the average displacement of the oscillators. Take for the g-th particle — sorry — oscillator

$$\ddot{x}_j + \omega_j^2 x_j = \underbrace{\tilde{V}(\omega)e^{i\omega t}}_{\substack{\text{harmonic} \\ \text{driving force}}} + \underbrace{\langle \tilde{G}_k(\omega)x_k \rangle}_{\text{collective force}} \tag{3.55}$$

Here the term $(\tilde{G}_k \tilde{x}_k)$ may be interpreted as the weighted contribution of particle k to the average

$$\langle \tilde{G}_k(\omega)x_k \rangle = \frac{1}{N} \sum \tilde{G}_k(\omega)x_k \ .$$

A 'mechanical' and an 'electrical' analog of Eq. (3.55) are sketched in Figs. 31 and 32.

To solve Eq. (3.55) we insert a trial solution

$$x_j = \tilde{x}_j e^{i\omega t} \ ; \qquad \langle \tilde{G}_k(\omega)x_k \rangle = \langle \tilde{G}_k(\omega)\tilde{x}_k \rangle e^{i\omega t} \ .$$

To find

$$\tilde{x}_j = \frac{1}{\omega_j^2 - \omega^2}[\tilde{V}(\omega) + \langle \tilde{G}_k \tilde{x}_k \rangle] \tag{3.56}$$

multiply with $\tilde{G}_j$ and average both sides. Call:

$$\langle \frac{\tilde{G}_j}{\omega_j^2 - \omega^2} \rangle = \tilde{S}(\omega) \tag{3.57}$$

and solve for the average

$$\langle \tilde{G}_j \tilde{x}_j \rangle = \frac{1}{N} \sum_{j=1}^{N} \tilde{G}_j \tilde{x}_j = \langle \tilde{G}_k \tilde{x}_k \rangle$$

to find

$$\langle \tilde{G}_k \tilde{x}_k \rangle = \frac{\tilde{S}(\omega)}{1 - \tilde{S}} \cdot \tilde{V}(\omega) \ . \tag{3.58}$$

Thus we do have a finite coherent amplitude $\langle \tilde{x} \rangle$. We can now use Eq. (3.58) to eliminate the 'collective force' term from Eq. (3.55). We find

$$\ddot{x}_j + \omega_j^2 x_j = \tilde{V}(\omega)\ e^{i\omega t} \underbrace{\left[1 + \frac{\tilde{S}(\omega)}{1 - \tilde{S}(\omega)}\right]}_{\text{shielding factor}} \ . \tag{3.55a}$$

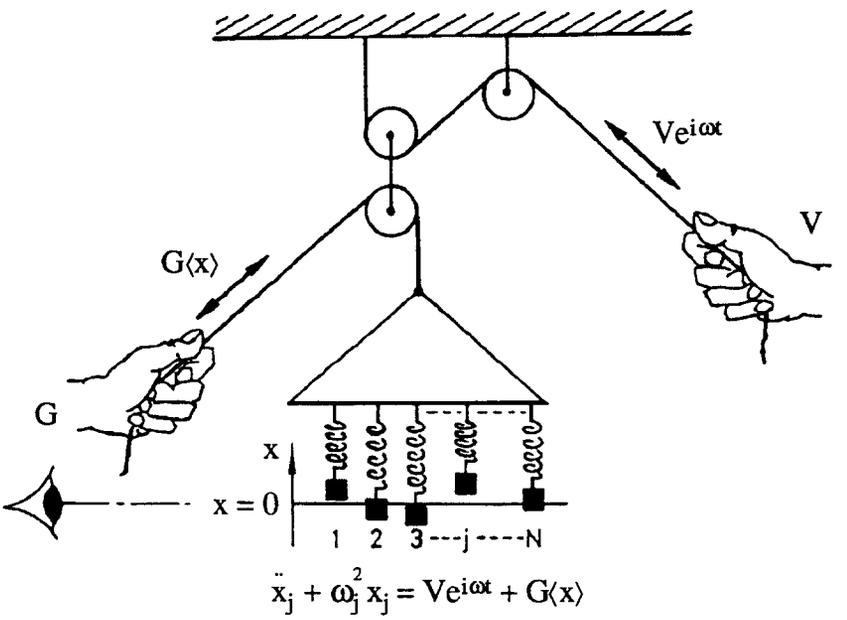


Figure 31: A 'mechanical analog' of Eq. (3.55). Person V tries to excite a system of oscillators (masses on springs) by shaking their point of suspension. Person G tries to damp the motion by observing the average displacement $\langle x \rangle$ of the oscillators and applying a damping force $G\langle x \rangle$.

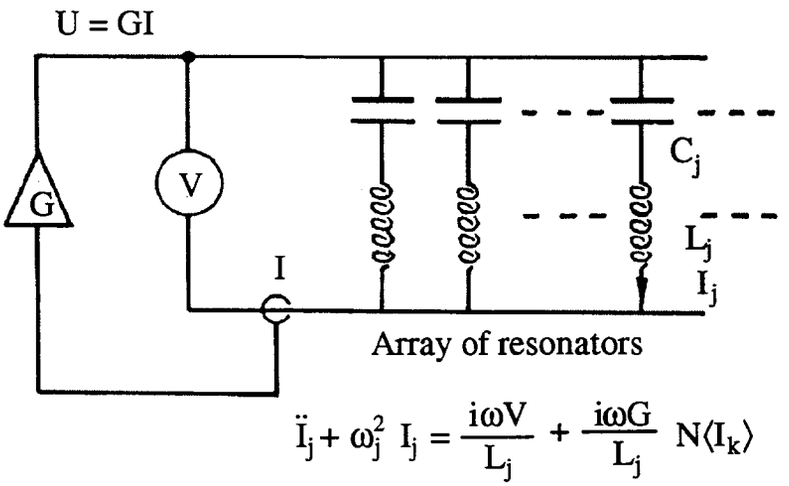


Figure 32: An electrical analog of Eq. (3.55). A group of LC-resonators is driven by a voltage $V\, e^{i\omega t}$. The sum $I = \sum I_k$ of the currents through the resonators is fed back through an amplifier with gain G to add an input voltage $G \sum I_k$.

Instead of treating the original Eq. (3.55) with a

$$\text{r.h.s.} = [\text{driving force}] + [\text{weighted average displacement}]$$

we can therefore treat the same equation with the more convenient:

$$\text{r.h.s.} = [\text{driving force}] \times [\text{shielding factor}] \ .$$

This is the essence of 'Sacherer's trick'. In the cooling equation we shall want to replace for each of the betatron bands involved:

$$\text{r.h.s.} = [\text{coherent term}] + \text{N} \cdot [\text{weighted average displacement}] + [\text{Schottky noise}] + [\text{amplifier noise}]$$

by

$$\text{r.h.s.} = [\text{coherent term} + \text{Schottky noise} + \text{amplifier noise}] \times [\text{shielding factor}] \ .$$

A quantity of key importance is the 'dispersion function' $\tilde{\text{S}}(\omega)$ entering into the shielding factor

$$\tilde{\text{T}}(\omega) \equiv \left[1 + \frac{\tilde{\text{S}}}{1 - \tilde{\text{S}}}\right] = \left[\frac{1}{1 - \tilde{\text{S}}}\right] \ . \tag{3.59}$$

For large N we have

$$\tilde{\text{S}}(\omega) = \frac{1}{\text{N}} \sum \frac{\tilde{\text{G}}(\omega_\text{j})}{\omega_\text{j}^2 - \omega^2} \rightarrow \int \frac{\tilde{\text{G}}(\omega_\text{j}) \text{n}(\omega_\text{j})}{\omega_\text{j}^2 - \omega^2} \text{d}\omega_\text{j} \approx \frac{1}{2\omega} \int \frac{\tilde{\text{G}}(\omega_\text{j}) \text{n}(\omega_\text{j})}{\omega_\text{j} - \omega} \text{d}\omega_\text{j} \ . \tag{3.60}$$

Here $\text{n}(\omega_\text{j})\text{d}\omega_\text{j}$ is the fraction of particles with eigenfrequencies in a band of width $\text{d}\omega_\text{j}$ near ω_j.

Dispersion integrals of the type (3.60) are treated in H.G. Hereward's chapter on Landau damping [15]. For convenience some features are repeated in Appendix 2. Due to the pole, the integral has an imaginary part even if $\text{G}(\omega_\text{j})$ is real. Details depend on the distribution $\text{n}(\omega_\text{j})$ of eigenfrequencies and on $\text{G}(\omega)$. A typical behaviour of $\tilde{\text{S}}(\omega)$ is sketched in Fig. 33.

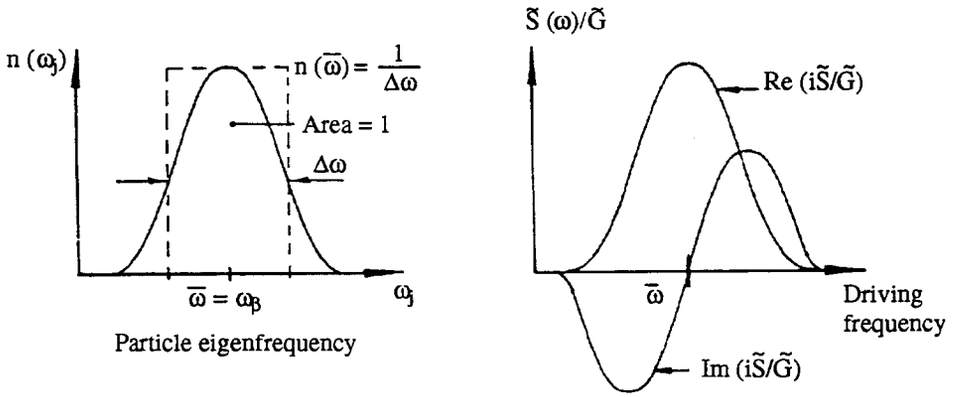


Figure 33: Frequency distribution and typical behaviour of the dispersion function Eq. (3.60) for a given function $\tilde{\text{G}}(\omega)$ which is constant (or slowly varying) near the beam response frequency ω_β. This behaviour of $\tilde{\text{G}}(\omega)$ is required for betatron cooling.

A useful approximation is

$$\begin{aligned} \tilde{S}(\omega) &\approx \frac{\pi}{2i\omega_\beta}\left[1 - i\frac{\omega-\omega_\beta}{\Delta\omega}\right] n(\omega)\tilde{G}(\omega) \\ n(\omega) &\approx \begin{cases} 1/\Delta\omega & \text{for } |\omega-\omega_\beta| < \Delta\omega/2 \\ 0 & \text{for all other values} \end{cases} \\ \omega_\beta &= \langle\omega_j\rangle = \frac{1}{N}\sum_1^N \omega_j \qquad \text{average eigenfrequency} \end{aligned} \tag{3.61}$$

To go one step further we now analyse a problem which is of some practical importance namely beam excitation by a single harmonic driving force on a kicker when the cooling loop is closed. We write the equation of motion of particle j as

$$\ddot{x}_j + \omega_j^2 x_j = \left[\sum_k (G(\omega)\sum_m e^{im\Omega_k(t-t_k)-i\mu_k} x_k\, e^{i\varphi_m(\omega)} + V\, e^{i\omega t}\right] \cdot \sum_{\ell=-\infty}^{\infty} e^{-i\ell\Omega_j(t-t_j-t_p)} \,. \tag{3.62}$$

Here the first sum (k) is over the N beam particles, the sum over m is the 'sampling term' due to the localized pick-up and the sum over ℓ represents the harmonics of the localized kick, t_k is the arrival time of particle k at the pick-up, $x_k \to x_k\, e^{-i\mu_k}$ presents the transformation of its oscillation from pick-up to kicker, $\varphi_m(\omega) = m\Omega t_d(\omega)$ is the signal delay of the cooling loop, t_p is the travelling time of particle j from pick-up to kicker, hence its arrival time at the kicker is $t_j + t_p$, $V\, e^{i\omega t}$ is the external driving force, the term proportional to $G(\omega)$ is the corresponding 'driving force' given by the response of the cooling loop to the beam oscillation.

Once again we drop all rapidly varying 'frequency shifts', i.e. we only take harmonics with $m = \ell$ in the first term on the r.h.s. of Eq. (3.62).

In the second term we only retain frequencies $\omega \pm \ell\Omega \approx \omega_\beta \approx \omega_j$ close to resonance. We assume that all bands are well separated so that only one ℓ leads to resonance. Thus we simplify Eq. (3.62) to

$$\ddot{x}_j + \omega_j^2 x_j = \left[\sum_k x_k G(\omega) e^{i\ell\Omega_k(t-t_k)-i\mu_k} e^{i\varphi(\omega)} + V e^{i\omega t}\right] \cdot e^{-i\ell\Omega_j(t-t_j-t_p)} \,. \tag{3.62a}$$

As response to the driving term $V\, e^{i\omega t}$ we expect a solution of the form

$$x_j = \tilde{x}_j\, e^{i(\omega-\ell\omega_j)t + i\ell\Omega_j(t_j+t_p)}$$

for any particle j.

Upon substitution [using the corresponding expression for x_k on the r.h.s. of Eq. (3.62a)] we find:

$$\tilde{x}_j = \frac{1}{\omega_j^2 - (\omega-\ell\Omega_j)^2}\,(N\cdot\langle\tilde{x}\tilde{G}\rangle + V) \tag{3.63}$$

where we define

$$\begin{aligned} \langle\tilde{x}\tilde{G}\rangle &= \frac{1}{N}\sum_{k=1}^N \tilde{x}_k \tilde{G}_k(\omega) \\ \tilde{G}_k(\omega) &\equiv G(\omega) e^{-i\mu_k + \delta\varphi_k} \end{aligned}$$

The quantity $\delta\varphi_k = \varphi(\omega) - \ell\Omega_k t_p$ is the synchronisation error between particle k and the cooling signal. Let us denote the resonant driving frequency — as introduced already in Eq. (3.24) — by

$$\omega \approx \omega_j + \ell\Omega_j = (Q_j + \ell)\Omega_j = \omega_\ell$$

and use $\omega_j^2 - (\omega - \ell\Omega)^2 \approx 2\omega_j[\omega_j - (\omega - \ell\Omega)]$.

From the preceding analysis we can now define a shielding factor for the present situation

$$\begin{aligned} \tilde{T}(\omega_\ell) &= \left[\frac{1}{1 - N \cdot \tilde{S}(\omega_\ell)}\right] \\ \tilde{S} &= \frac{1}{2\omega_\beta}\int \frac{\tilde{G}(\omega_\ell)n(\omega_\ell)}{(\omega_j + \ell\Omega_j) - \omega} d\omega_\ell \\ \omega_\beta &= \langle\omega_j\rangle : \text{ beam average of betatron frequency.} \end{aligned} \tag{3.64}$$

Using this shielding factor we can rewrite Eq. (3.62a) as

$$\ddot{x}_j + \omega_j^2 x_j = \tilde{T}(\omega_\ell) \cdot V \cdot e^{i\omega_\ell t - i\alpha_\ell} \tag{3.62b}$$

where we use $\alpha_\ell = -\ell\Omega_j(t_j + t_p)$ to denote the phase factor due to the arrival time of particle j at the kicker. Thus when the cooling loop is closed, the response of Eq. (3.62) to $V\, e^{i\omega t}$ changes by $\tilde{T}(\omega)$. In this way $\tilde{T}(\omega)$ can be observed and $\tilde{G}(\omega)$ can be deduced from it. Usually these measurements are done using a network analyser to display the beam response to a swept sine wave (beam transfer function measurement) as sketched in Fig. 34. This permits us to adjust the characteristics of the cooling loop band by band.

To complete our analysis we return to Eq. (3.62) but now assume a general driving force represented by a Fourier series (or a Fourier integral) with a spectral density function $V(\omega)$. We invoke superposition and resonant behaviour of the betatron equation at the frequencies ω_ℓ. Thus we rewrite Eq. (3.62b) as

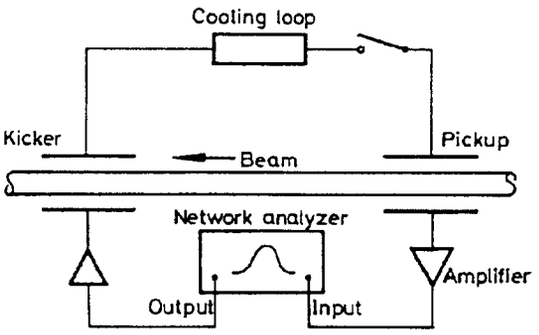


Figure 34: Arrangement to measure beam transfer function. The frequency sweep of the network analyser is set to cover one or several betatron sidebands. The difference in beam response with cooling loop open and closed can be used to optimize the loop again.

$$\ddot{x}_j + \omega_j^2 x_j = \sum_{\ell=-\infty}^{\infty} \tilde{T}(\omega_\ell) V(\omega_\ell) e^{i\omega_\ell t - i\alpha_\ell} \tag{3.62b}$$

which presents the effect as the sum of the interaction at the sidebands $\omega_\ell = (\ell + Q)\Omega$. As a consequence of the beam feedback each band now has its proper shielding factor $\tilde{T}(\omega_\ell)$, Eq. (3.64) (well separated band, i.e. poor mixing assumed). The effect of the shielding factor is fully equivalent to introducing a transfer function $\tilde{T}(\omega)$ between the driver and the kicker.

We can now generalize the cooling rate in Eqs. (3.49) and Eq. (3.50) to include shielding. We can interpret V(t), Eq. (3.65) as the cooling signals discussed before (namely the self-effect of the test particle, the Schottky noise due to the other particles and the amplifier noise). Since the beam feedback acts like a transfer function we simply include this into Eqs. (3.49) and Eq. (3.50) by substituting

$$\tilde{G}(\omega_\ell) \rightarrow \tilde{T}(\omega_\ell)\tilde{G}(\omega_\ell) \; . \tag{3.65}$$

A typical behaviour of the shielding function is sketched in Fig. 35. Note that for small 'gain' (N.G. small) and small $S(\omega)$ the shielding factor is close to 1.

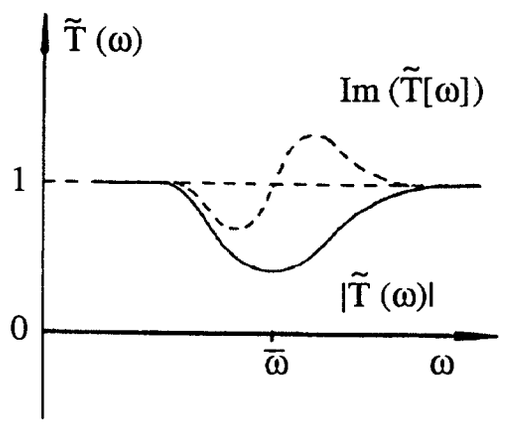


Figure 35: Typical behaviour of the shielding factor $T(\omega)$ near a resonance frequency of the beam.

To gain further insight we only look at particles near the centre of the distribution ($\omega_\ell \approx \langle \omega_j + \ell\Omega_j \rangle = \omega_\beta + \ell\Omega$) and assume perfect betatron phase and perfect signal delay pick-up to kicker ($\delta_\mu = 0, \delta\varphi(\omega_\ell) = 0$). Then the gain function $G(\omega_\ell) = |G(\omega_\ell)|e^{-i\pi/2-\delta_\mu+\delta\varphi\rightarrow -i|G(\omega_\ell)|}$ becomes purely imaginary and $S(\omega_\ell)$ and $T(\omega_\ell)$ real in the centre of the band. Let us introduce the 'reduced gain'

$$g(\omega_\ell) = \frac{|G(\omega_\ell)| \cdot N}{\omega_\beta \Omega/2\pi} \equiv g_\ell \tag{3.53b}$$

in analogy to Eq. (3.53) and recall the definition of the mixing factor $M_\ell \approx \Omega/2\Delta\omega_\ell$ for well separated bands [see under Eq. (3.43)].

Using the simplification Eq. (3.61) for the dispersion integral and Eq. (3.64) for $T(\omega_\ell)$ we have

$$\tilde{S}(\omega_\ell) \approx \frac{\pi}{2i\omega_\beta} \frac{\tilde{G}(\omega_\ell)}{\Delta\omega_\ell} = \frac{-1}{N} g_\ell \, M_\ell/2 \tag{3.66}$$

$$T(\omega_\ell) = \frac{1}{1 + g_\ell\, M_\ell/2}\ .$$

The cooling rate equation for any particle is obtained from the expressions of section 3.10 by replacing $\tilde{G}(\omega) \rightarrow \tilde{T}(\omega)G(\omega)$, Eq. (3.65). We obtain in the present case:

$$\frac{1}{\tau} = \frac{\Omega/2\pi}{2N} \sum_{\substack{\ell=-\infty \\ \text{passband}}}^{\infty} \left[\frac{2g_\ell}{(1 + g_\ell M_\ell/2)} - \frac{g_\ell^2}{(1 + g_\ell M_\ell/2)^2}(M_\ell + U_\ell)\right]\ . \tag{3.67}$$

This is formally the same as Eq. (3.52) if we substitute

$$g_\ell \rightarrow T_\ell g_\ell = \frac{g_\ell}{1 + g_\ell M_\ell/2}\ .$$

Optimum cooling is obtained from Eq. (3.67) when:

$$T_\ell g_\ell (M_\ell + U_\ell) = 1$$

i.e. when for all bands

$$g_\ell = \frac{1}{M_\ell/2 + U_\ell}\ ; \qquad g_\ell \rightarrow \frac{2}{M_\ell}\ . \tag{3.68}$$

The limiting case ($\rightarrow$) is for negligible amplifier noise, $U_\ell \ll M_\ell/2$. The optimum shielding factor corresponding to Eq. (3.68) is:

$$T_\ell = \frac{1}{1 + M_\ell/(M_\ell + 2U_\ell)} \rightarrow \frac{1}{2} \tag{3.69}$$

and the optimum damping rate

$$\frac{1}{\tau} = \frac{\Omega/2\pi}{2N} \sum_\ell \frac{1}{M_\ell + U_\ell}\ . \tag{3.70}$$

Thus in the situation of negligible amplifier noise, optimum cooling is obtained when the gain (at all bands involved) leads to signal reduction by a factor of about 2. By comparing open and closed loop signals (either Schottky noise or driven-beam response) the gain can thus be optimized band by band. An example of Schottky signal shielding of a band is given in Fig. 36. Note that the optimum gain Eq. (3.68) for $U = 0$ is twice the optimum Eq. (2.28) calculated without beam feedback. When the amplifier noise becomes important ($U \gg M$) then Eq. (3.68) and Eq. (3.69) yield the optimum $g \rightarrow 1/U$, $T \rightarrow 1$ as in the case without shielding.

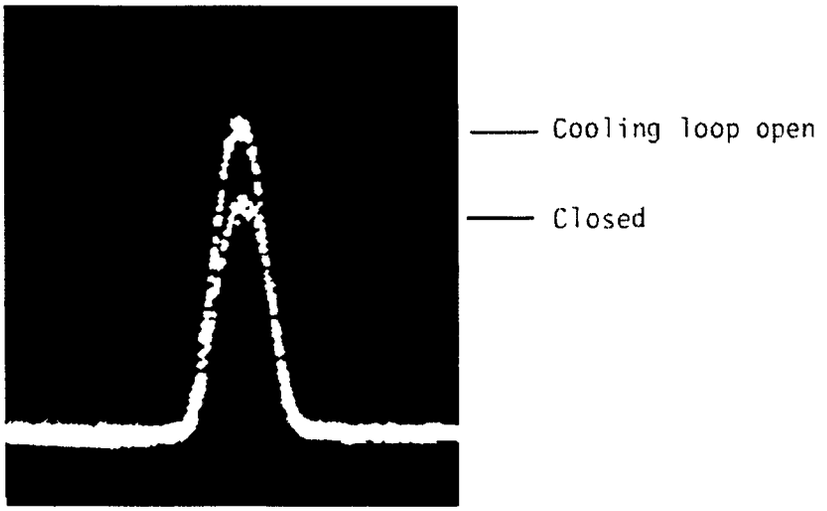


Figure 36: Reduction of a Schottky noise band when the cooling loop is closed. With negligible amplifier noise and well separated bands optimum gain of the cooling loop corresponds to a signal amplitude reduction by about 2 in the centre of the bands.

Thus the inclusion of beam shielding (which was done in an approximate manner here) leads to an improved expression for the cooling rate and — more importantly — to an adjustment criterion for the cooling system.

The analysis done here for betatron cooling can be repeated for momentum spread damping where similar gain adjustment criteria apply.

4 DISTRIBUTION FUNCTION EQUATIONS (FOKKER–PLANCK) AND MOMENTUM SCALING

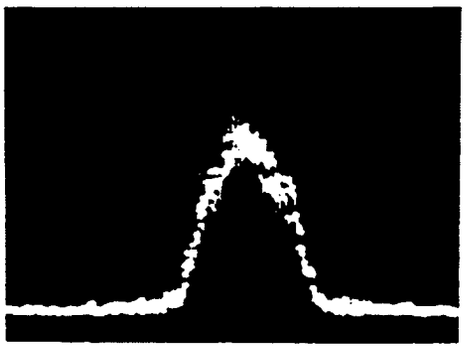
4.1 Distribution functions and particle flux

To follow the details of the cooling process, we (may) want to know more than the evaluation of the mean-square beam size and the r.m.s momentum spread — the only quantities used up to now to characterize cooling. In fact, a beam profile monitor records the particle distribution with respect to transverse position (see Fig. 37 as an example), and a longitudinal Schottky scan such as Fig. 2 gives the (square root of the) momentum distribution. These pictures are rich in fine information on peak densities, densities in the tails, asymmetries, and other practical details which are overlooked if only r.m.s. is regarded.

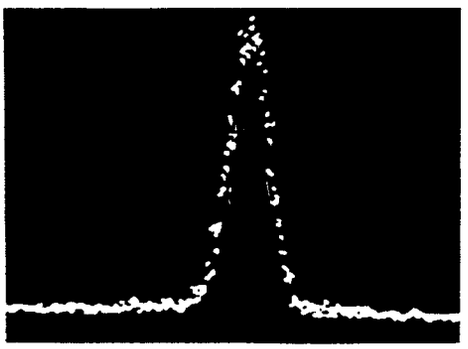
It is therefore challenging to find an equation which describes all that can be observed and that is of practical importance. Such an equation does in fact exist!

For stochastic cooling the problem was (to my knowledge) first tackled by Thorndahl [16] who already in 1976 worked with a Fokker–Planck type of equation for the particle density. This line was followed by virtually all subsequent workers [17], and computer codes for solving the distribution function equations are extensively used in the design of stochastic cooling and stacking systems.

The basic ideas behind this 'distribution function analysis' are simple, so that also the beginner can get — hopefully without too much pain — some first degree of familiarity with this powerful tool of cooling theory. I will first give the recipe and then try to justify it.



a)



b)

Figure 37: Evolution of beam profile (number of particles vs. vertical position) during stochastic cooling test in 'ICE'. The scans were obtained with a profile monitor which records the position of electrons liberated by beam particles through collisions with the residual gas. a) Before cooling; b) after 4 min of cooling.

Let $\psi(x)$ (Fig. 38) be the particle distribution with respect to the error x (e.g. $x = \Delta p/p$). Define $\psi(x) = dN/dx$ so that $\psi(x)\,dx$ gives the number of particles with an error in the range x to $x + dx$. During cooling we find different distributions $\psi(x)$, taking snapshots at different times (see Fig. 2 as an example). We characterize this by letting $\psi = \psi(x,t)$ be a function of time also. The partial differential equation which describes the dynamics of $\psi(x,t)$ can be written in the following form:

$$\frac{\partial \psi}{\partial t} = \frac{\partial}{\partial x}\left(-F\psi + D\frac{\partial \psi}{\partial x}\right) . \tag{4.1}$$

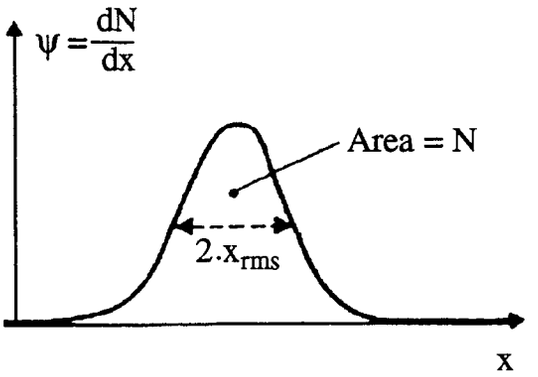


Figure 38: A particle distribution function $\psi(x)$ defining the number of particles $dN = \psi(x)dx$ with an error in the interval from x to x + dx.

The cooling process is completely characterized by the two coefficients F and D (which describe the cooling system) and the initial conditions $\psi(x,t = 0)$ (which describe the distribution at the start). Particle loss due to walls or influx during stacking can be included via appropriate boundary conditions $\psi(x_1) = 0, (\partial\psi/\partial x)(x_1) = \text{const.}$, etc. Two representative examples of results obtainable with Eq. (4.1) are given in Fig. 39, taken from Ref. [18], and Fig. 20 from Ref. [13].

To analyse a given system we have to find its coefficient F and D. These quantities are closely related to the coherent and incoherent effect, respectively, which we have identified before. In fact

$$F/f_0 = \langle \Delta x \rangle_t \tag{4.2}$$

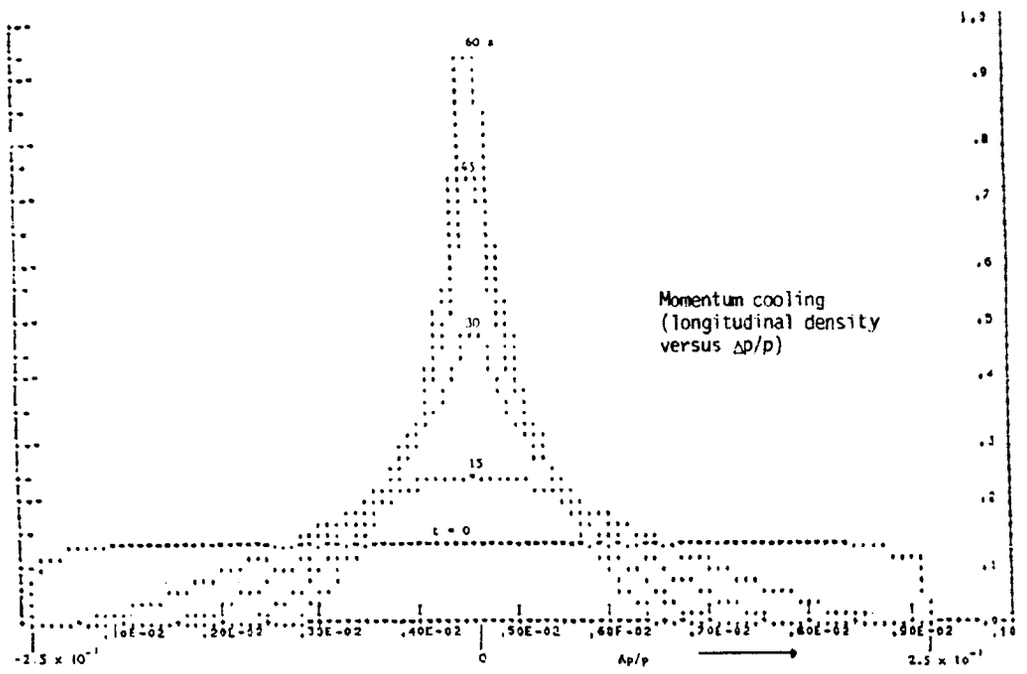


Figure 39: Momentum cooling at 600 MeV/c in LEAR computed using Eq. (8.1). (Curves taken from Ref. [18]).

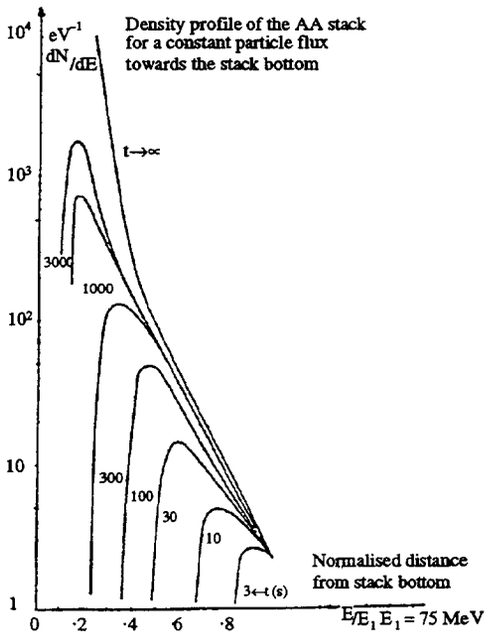


Figure 40: Evolution of the stack in the AA during stochastic accumulation. Curves computed using the distribution function equation with the boundary condition of constant particle influx simulating the new $\bar{p}$ added every 2.6 sec. (Curves taken from Ref. [19]).

is the expectation value (long-term average) of the coherent change Δx per turn of the error, and

$$2D/f_0 = \langle(\Delta x)^2\rangle_t \tag{4.3}$$

is the expectation of the square of this change. The quantities F and 2D alone are corresponding average changes per second. Note the difference between $(\Delta x)^2 = (x_c - x)^2$ used here, and $\Delta(x^2) = x_c^2 - x^2$ as frequently used before!

The important thing is that a distribution function Eq. (4.1) — similar to the Fokker–Planck equation used in a variety of fields — exists and that relatively simple prescriptions (4.2) and (4.3) permit us to establish the two coefficients F and D for any given stochastic cooling system. Incidentally, an equation similar to Eq. (4.1) had long been used (before 1976!) by the Novosibirsk Group to study the dynamics of electron cooling. Also the kinetic equations in plasma physics closely resemble our distribution equation.

Let us now try to follow a simple derivation of Eqs. (4.1)–(4.3). This derivation is due to Thorndahl [16]. It proceeds along the lines used in textbooks to derive the diffusion — or heat transfer — equations which resemble Eq. (4.1). Imagine a distribution function $\psi(x)$ and calculate, for a particular value x_1 of x, the number of particles per turn which are transferred from x-values below x_1 to values above x_1 (Fig. 16). If the correction per turn at the kicker is Δx, then particles with an error between x_1 and $x_0 = x_1 - \Delta x$ (cross-hatched area in Fig. 41) pass through x_1. Their number is

$$\Delta N = \int_{x_0}^{x_1} \psi(x)dx \ . \tag{4.4}$$

Expanding ψ at x_1,

$$\psi \approx \psi(x_1) + \frac{\partial\psi(x_1)}{\partial x}(x - x_1) \ , \tag{4.5}$$

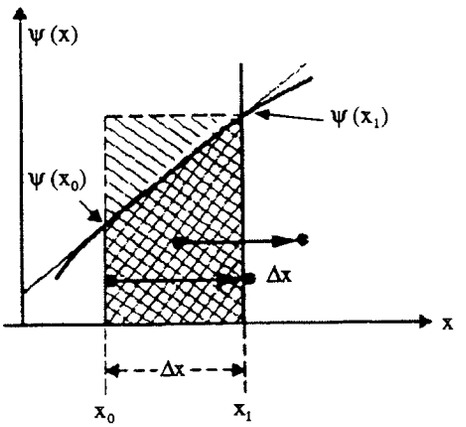


Figure 41: A look at the distribution function Fig. 38 through a magnifying glass. When the error for particles with a value near x_1 is changed by Δx, particles in the dark shaded area have the error value changed from values below to values above x_1, Eq. (4.6) expresses this area as the difference between the rectangle and the triangle sketched in the figure.

the integration yields

$$\Delta N = \psi(x_1) \cdot \Delta x - \frac{1}{2}\frac{\partial\psi(x_1)}{\partial x}(\Delta x)^2 \ . \tag{4.6}$$

The first and second terms can be interpreted as the area of the rectangle and the triangle, respectively, sketched in Fig. 41.

We now define the (average) particle flux

$$\phi = f_0\langle\Delta N\rangle_t$$

as the expected number of particles per second passing a given error value. Clearly, then, from Eq. (4.6), the instantaneous flux is:

$$\phi(x) = \underbrace{f_0\langle\Delta x\rangle_t}_{F}\ \psi(x) - \underbrace{\frac{f_0}{2}\langle(\Delta x)^2\rangle_t}_{D}\ \frac{\partial\psi}{\partial x} \ .$$

This gives the flux in terms of F and D as defined by Eqs. (4.2) and (4.3). The assumption has tacitely been made that the change Δx per turn at the kicker is small and $\psi(x)$ smooth, so that higher expansion terms in Eq. (4.5) can be neglected.

Having found the flux we can immediately obtain Eq. (4.1) from the continuity equation

$$\frac{\partial\phi}{\partial x} + \frac{\partial\psi}{\partial t} = 0 \ . \tag{4.7}$$

It states that the change per second of the density is given by the 'gradient'$-\partial\phi/\partial x$ of the flux. This is similar to continuity considerations in other fields like, for instance, the

charge conservation law of electrodynamics:

$$\frac{\partial \mathrm{j}}{\partial \mathrm{x}} + \frac{\partial \rho}{\partial \mathrm{t}} = 0 \ , \tag{4.8}$$

relating current density j and charge density ρ.

Like other continuity equations, Eq. (4.7) can be obtained by looking at the flux going into and coming out of an element of width dx in ψ, x-space (Fig. 42):

Incoming flux per second: ϕ_1

Outgoing flux per second: $\phi_2 \approx \phi_1 + \frac{\partial \phi}{\partial \mathrm{x}}\mathrm{dx}$

Surplus per second: $\Delta\phi = \phi_1 - \phi_2 = -\frac{\partial \phi}{\partial \mathrm{x}}\mathrm{dx}$.

The resulting densisty increase (per second) in the element is thus

$$\frac{\Delta\phi}{\mathrm{dx}} = -\frac{\partial \phi}{\partial \mathrm{x}} \ ,$$

and conservation of the particle number requires a $\partial\phi/\partial \mathrm{t}$ equal to this.

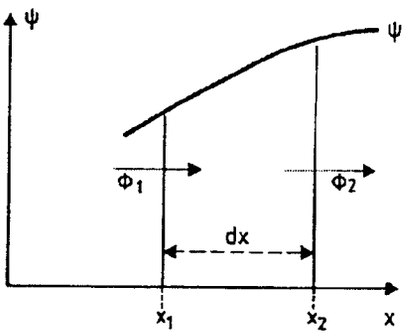


Figure 42: The flux into and out of a narrow element of width dx in ψ-x space. An excess of incoming over outgoing flux leads to an increase with time of the density $\psi = \Delta\mathrm{N}/\mathrm{dx}$ of particles in the element.

This completes the derivation. The resulting equation (4.1) agrees with observations made in the ISR and all subsequent machines using stochastic cooling. The reader who might have had some difficulty in appreciating the derivation may now be pleased to learn that the exact form of Eq. (4.1) has been a subject of discussion for quite some time. Looking at the derivation of the Fokker–Planck equation in textbooks [20], one is tempted to put the coefficient D under the second derivative as is correct for a variety of other stochastic processes. In 1977 a machine experiment [21] was performed at the ISR to clear up this question for cooling and diffusion problems in storage rings. The experiment clearly indicated that in the present case the diffusion term should be $(\partial/\partial \mathrm{x})[\mathrm{D}(\partial\psi/\partial \mathrm{x})]$ as in Eq. (4.1) and not $(\partial^2/\partial \mathrm{x}^2)(\mathrm{D}\psi)$.

4.2 Example of asymptotic distributions and Palmer cooling

We may conclude from the preceding sections that it is relatively simple to determine the distribution equation pertaining to a given cooling problem. It is usually much more

difficult to solve the equation. This is because in general the coefficients F and D are functions of x, t, and ψ itself. Analytical solutions have therefore only been obtained in a few simple cases.

As an example, let us briefly look at Palmer cooling with the following simplifying assumption: No unwanted mixing, and Schottky noise negligible compared with amplifier noise. Denoting $x = (\Delta p/p)$, the correction per turn is

$$\Delta x = -g[\langle x\rangle_s + x_n]$$

as given by Eq. (2.24) in Section 2. In analogy to Eq. (2.18) in Section 2.3.3, we assume that the long-term average of $\langle x\rangle_s = (1/N_s)\sum x_i$ is zero except for the contribution x/N_s of the test particle upon itself. The noise x_n has zero average. Hence

$$\langle \Delta x\rangle_t = -g\frac{x}{N_s} = -g\frac{2W}{Nf_0}\cdot x\ .$$

In a similar way (using the assumption that $\langle x_n^2\rangle_t \gg \langle\langle x\rangle_s^2\rangle_t$), i.e. amplifier noise dominating over Schottky noise)

$$\langle(\Delta x)^2\rangle_t = g^2\langle x_n^2\rangle_t = g^2 x_{n,rms}^2 = \text{const.}$$

Hence in this simple case $F = F_0' x$ and $D = D^0$, where $F_0' = (2W/N)g$ and $2D_0 = f_0 g^2 x_{n,rms}^2$ are constants. In this case, Eq. (4.1) is amenable to an analytic solution. Try

$$\psi = \frac{N}{\sqrt{2\pi}\sigma(t)}\exp[-x^2/2\sigma(t)^2]$$

i.e. a Gaussian with σ changing in time. Upon substitution, one obtains an ordinary differential equation for the width, σ, of the Gaussian:

$$\dot{\sigma}/\sigma = -F_0' + D_0/\sigma^2\ .$$

Special cases:

$$\begin{aligned} D_0 &= 0: \quad \sigma^2 = \sigma_0^2 e^{-2F_0' t} \qquad \text{(continuous cooling)}\ , \\ F_0' &= 0: \quad \sigma^2 = \sigma_0^2 + 2D_0 t \qquad \text{(diffusion)}\ . \end{aligned}$$

General solution:

$$\sigma^2 = \sigma_1^2 e^{-2F_0' t} + D_0/F_0'\ .$$

This describes cooling towards an asymptotic (Gaussian) distribution with $\sigma_\infty = \sqrt{D_0/F_0'}$. In this situation an equilibrium between heating and cooling is reached. A similar result is arrived at from the simple cooling equations [e.g. Eq. (25), Section 2] which suggest $1/\tau \to 0$ when the signal $(\langle x\rangle_s)^2$ has decreased so much that $gU = g[x^2/(\langle x\rangle_s)^2] \to 2$. The new information obtained from Eq. (4.1) is that the asymptotic ψ is Gaussian in the simple case considered.

The existence of asymptotic equilibrium distribution is a common feature also in more complicated cases of Eq. (4.1). The final distribution ψ_∞ can be obtained putting $\partial\psi/\partial t = 0$, which converts Eq. (4.1) into a simpler ordinary differential equation:

$$-F\omega_\infty + D\frac{d\psi_\infty}{dx} = \text{const.} \tag{4.9}$$

The constant is frequently zero (e.g. when $F(x) = 0$ and $\partial\psi/\partial x = 0$ for $x = 0$ as can often be inferred from the symmetry of the problem). Equation (4.9) is important as it indicates the limiting density which can be reached.

4.3 Momentum cooling by filter and transit time methods

These methods measure the revolution frequency of particles or the time of flight between pick-up and kicker in order to detect the momentum error. The filter method of Carron and Thorndahl [22] (Fig. 43) uses a notch filter between the preamplifier and the power amplifier, with notches at all revolution harmonics in the passband (Fig. 44). In the simplest case the filter is a transmission line shorted at the far end (Fig. 44), with a length corresponding to half the revolution time of the particles in the storage ring. The notches are produced by $\lambda/2$ resonances, where ideally the input impedance is zero and the phase changes sign. Because of these phase and amplitude characteristics, particles with a wrong revolution frequency are accelerated or decelerated until ideally all have 'fallen into the notches'. The filter method is important for the cooling of low-intensity beams, and in fact the whole antiproton complex at CERN would probably not have worked with stochastic cooling had this technique not been invented in due time. Sum pick-ups are used, and these produce a much larger signal than the difference devices that are necessary with other methods. The filter reduces not only the particle signals but also the preamplifier noise at the critical frequencies. This feature is important for fast cooling at low intensity. The price to pay for this is that all the Schottky bands used have to be well separated, so that particles 'know' the notches into which they have to fall. This means unavoidably imperfect mixing. However, this slight disadvantage could probably be circumvented by using the signal from a second pick-up — rather than the reflection of the previous turn pulse via a cable — to cancel signals of a particle with the correct time of flight between the two pick-ups and to accelerate/decelerate others [23].

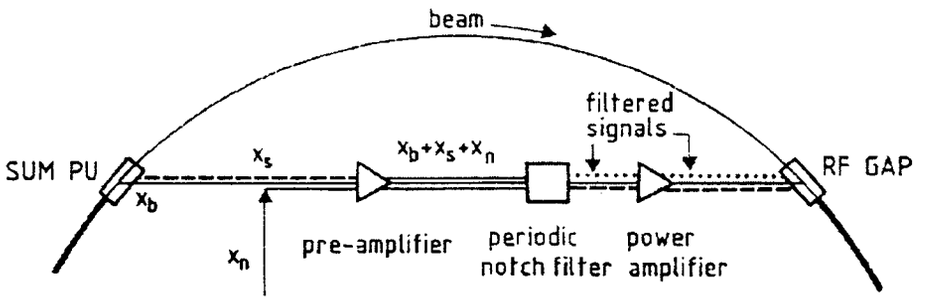


Figure 43: The basic set up for momentum cooling by the filter method. An advantage of this method is that a sum pick-up is used which is sensitive even to small beam signals. Secondly, Schottky and preamplifier noise are reduced by the filter.

Yes another time-of-flight method has been discussed at Fermilab [24]. Essentially, the idea is to differentiate the pick-up pulse and apply this signal on the kicker with a delay so that particles with the correct time of flight between pick-up and kicker are not affected, whereas slow or fast ones get a correction.

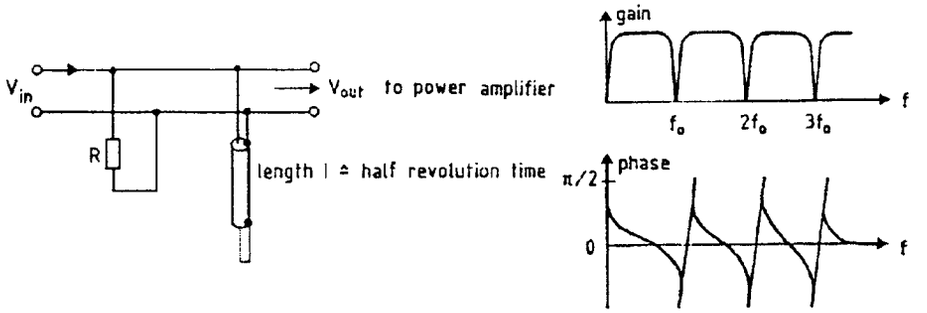


Figure 44: A simple periodic notch filter namely a half-wave, low-loss transmission line (used as a stub resonator). The (idealized) gain and phase characteristics are given by the half-wave resonances at the multiples of the revolution frequency. Additional elements are usually added to reduce the gain between the harmonics.

Both variants of the filter method are less efficient in noise suppression and have, therefore, not found applications so far.

We shall return to the time domain for a short moment to suggest slightly different explanations of the filter method: the pulse sent into the cooling system by a particle of nominal frequency will be cancelled by its pulse from the previous revolution reflected at the end of the line. For particles that are too slow or too fast, the cancellation is imperfect and acceleration or deceleration will result.

The filter method is usually analysed using the distribution Eq. (4.1). The coefficients F and D can be worked out theoretically and/or by measurements on the system. Usually, measurements and calculations are done harmonic by harmonic, including various ingredients such as imperfect mixing and signal suppression. All we want to do here is to write down the general form of the relevant coefficients F and D which, expanding up to second order in the error quantity $x = \Delta E/E$ take the following form:

$$\begin{aligned} F &= -G_0 x \\ D &= G_1^2\, \psi(x^2 + \kappa_0) + G_2^2\, \mu(E^2 + \kappa_1)\ , \end{aligned}$$

where x is the relative energy error; G_0 (proportional to the 'gain' g), $G_1^2(\propto g^2)$, and $G_2^2(\propto g^2)$ are given by the ideal filter, κ_0 and κ_1 by the losses; and μ relates to the amplifier noise. The first term of D (which is proportional to the density ψ) gives the Schottky noise filtered by the notches, and the second term the filtered preamplifier noise. For more details, the reader should consult the specialized literature.

References

[1] See for example Bibliography, 1977:1, 1978:3 and 13, 1980:1, 1981:1, 3 and 12, 1982:1 and 16, 1983:7, 1984: 1,2 and 3, 1986:11.
H. Poth, Electron cooling, Proc. CERN Accelerator School, Advanced Accelerator Physics, Oxford, 1985, CERN 87–03 (1987).
K. Hübner, Radiation damping, Proc. CERN Accelerator School, General Accelerator Physics, Gif-sur-Yvette, 1984, CERN 85–19 (1985).
K. Hübner, Synchrotron radiation, ibid, p. 239.

[2] H. Bruck, Accélérateurs circulaires de particules, Bibliothèque des Sciences Nucléaires, Paris (1966).
M. Sands, The physics of electron storage rings, SLAC report 121 (1970).

[3] See Bibliography, 1981:12, 1984:4, 1984:7, and 1986:2.

[4] W. Meyer Eppler, Grundlagen und Anwendungen der Informationstheorie, Springer Verlag (1969).

[5] See Bibliography, 1980:9 and 10.

[6] M.R. Spiegel, Statistics (Schaum's Outline Series) (McGraw-Hill, New York, 1972), Chapter 8.
L. Maisel, Probability, statistics and random processes (Simon and Schuster Inc., New York, 1971), Chapter 6.

[7] See Bibliography, 1977:1.

[8] M. Bregman et al., Measurement of antiproton lifetime using the ICE storage ring, Phys. Lett. **78B** (1978) 174.
See also Bibliography, 1980:1.

[9] See Bibliography, 1983:19 and 28, 1985:3 and 5.

[10] See Bibliography, 1983:2, 1984:9, 10, 11 and 12.

[11] See Bibliography, 1977:1, 1980:1.

[12] H.G. Hereward and K. Johnsen, The effect of radio frequency noise, CERN 60–38 (1960).

[13] See Bibliography, 1980:1 and 1982:18.

[14] See Bibliography, 1978:3.

[15] H.G. Hereward, Landau damping, Proceedings of the CERN Accelerator School, Oxford 1985, CERN 87–03, p. 255, and: The elementary theory of Landau damping, CERN 65–20 (1965).

[16] See Bibliography, 1977:8.

[17] See Bibliography, 1977:10, 1978:3, 1979:1, 2 and 3, 1980:3 and 13, 1981:1.
S. van der Meer and L. Thorndahl, Computer programs for solving Eq. (4.1), private communication.

[18] See Bibliography, 1980:10.

[19] See Bibliography, 1978:15.

[20] S. Chandrasekar, Stochastic problems in Physics and Astronomy, in: Noise and stochastic processes, ed. N. Wok (Dover Press, New York, 1954).
G. Ishimaru, Basic principles of plasma physics (W.A. Benjamin Inc., Reading, Mass., 1973).

[21] L.Thorndahl et al., Diffusion in momentum caused by filtered noise, CERN internal report ISR-RF-TH Machine Performance report, 19 August 1977.

[22] See Bibliography, 1978:2.

[23] See Bibliography, 1984:3.

[24] See Bibliography, 1980:4, p. 777.

BIBLIOGRAPHY

Chronological list of internal reports, conference presentations, and other publications related to stochastic cooling

1838
1. J. Liouville, J. Math. Pures et Appl. 3, 348.

1918
1. W. Schottky, Über spontane Schwankungserscheinungen in verschiedenen Elektrizität-shalbleitern, Ann. Phys. **57**, 541.

1956
1. D.B. Lichtenberg, P. Stahle and K.R. Symon, Modification of Liouville's theorem required by the presence of dissipative forces, Midwestern Univ. Research Assoc. report MURA-DBL-PS-KRS1.

1958
1. F. Mills and A.M. Sessler, Liouville's theorem for a continouous medium..., report MURA-433.

1972
1. S. Van der Meer, Stochastic damping of betatron oscillations, internal report CERN/ISR PO/72–31.
2. W. Schnell, About the feasibility of stochastic damping in the ISR, internal report CERN/ISR RF/72–46.

1973
1. R.B. Palmer, stochastic cooling, Brookhaven Nat.- Lab. report BNL 18395.

1974
1. J. Borer, P. Bramham, H.G. Hereward, K. Hübner, W. Schnell and L. Thorndahl, Non-destructive diagnostics of coasting beams with Schottky noise, Proc. 9th Int. Conf. on High Energy Accelerators, Stanford (USAEC Conf. 740522, Washington, 1974), p. 53.

1975
1. P. Bramham, G. Carron, H.G. Hereward, K. Hübner, W. Schnell and L. Thorndahl, stochastic cooling of a stored protron beam, Nucl. Instrum. Methods **125**, 201.
2. L. Thorndahl, stochastic cooling of momentum spread and betatron oscillations for low-intensity stacks, internal report CERN/ISR-RF/75–55.

1976
1. P. Strolin, L. Thorndahl and D. Möhl, stochastic cooling of antiprotons for ISR physics, internal report CERN/EP 76–05.

2. D. Cline, P. McIntyre, F. Mills and C. Rubbia, Collecting antiprotons in the Fermilab booster and very high energy proton-antiproton interactions, Fermilab internal report TM 689.
3. C. Rubbia, P. McIntyre and D. Cline, Producing massive neutral intermediate vector bosons with existing accelerators, Proc. Int. Neutrino Conf., Aachen, 1976 (Vieweg Verlag, Braunschweig, 1977), p. 683.
4. K. Hübner, D. Möhl, L. Thorndahl and P. Strolin, Estimates of ISR luminosities with cooled beams, CERN/PS/DL Note 76–27.
5. G. Carron, L. Faltin, W. Schnell and L. Thorndahl, stochastic cooling of betatron oscillations and momentum spread, Proc. Vth All-Union Particle Accelerator Conf., Dubna (USSR Acad. Sci., Moscow, 1977), p. 241.
6. G. Carron and L. Thorndahl, stochastic cooling of vertical betatron oscillations in the frquency range 80–340 MHz, CERN ISR Perf. Report, ISR-RF/LTS/ps (RUN 775).
7. W. Hardt, Augmentation of damping rate for stochastic cooling by the additional use of non-linear elements, CERN/PS/DL 76–10.
8. E.D. Courant, Workshop on phase-space cooling, in Proc. ISABELLE Workshop, Brookhaven (Report BNL 50611, Upton, 1976), p. 241.

1977

1. H.G. Hereward, Statistical phenomena – Theory, Proc. 1st Course Int. School of Particle Accelerators, Erice (Report CERN 77–13, Geneva, 1977), p. 281.
2. W. Schnell, Statistical phenomena – Experimental results, ibid., p. 290.
3. L. Faltin, RF fields due to Schottky noise in a coasting particle beam, Nucl. Instrum. Methods **145**, 261.
4. L. Thorndahl, Stochastic cooling of betatron oscillations in ICE, CERN/ISR Technical Note ISR/RF/LT/ps.
5. G. Carron, L. Faltin, W. Schnell and L. Thorndahl, Experiments with stochastic cooling in the ISR, Proc. Particle Accelerator Conf., Chicago, 1977 [IEEE Trans. Nucl Sci. **NS–24** (3), 1977], p. 1402.
6. G. Carron, L. Faltin, W. Schnell and L. Thorndahl, Recent results with stochastic cooling in the ISR, Proc. 10th Int. Conf. on High-Energy Accelerators, Serpukhov (IHEP, Serpukhov, 1977), vol. I, p. 523.
7. S. van der Meer, Influence of bad mixing on stochastic acceleration, internal note CERN/SPS/DI/pp 77–8.
8. L. Thorndahl, A differential equation for stochastic cooling of momentum spread with the filter method, technical note ISR-RF/LT/ps.
9. S. van der Meer, Normalized solution for linear momentum cooling of a square distribution, unpublished document.
10. L. Jackson Laslett, Evolution of the amplitude distribution function for a beam subjected to stochastic cooling, Berkeley report LBL-6459.

1978

1. ICE Team, Initial Cooling Experiment progress reports Nos. 1 and 2, CERN-EP Div.
2. G. Carron and L. Thorndahl, Stochastic cooling of momentum spread with filter techniques, internal report CERN/ISR-RF/78–12 and ISR-RF/Note LT/ps.
3. F. Sacherer, Stochastic cooling theory, internal report CERN ISR-TH 78–11.

4. G. Carron et al., Stochastic cooling tests in ICE; Phys. Lett. **77B**, 353.
5. Design Study Team, Design study of a proton-antiproton colliding beam facility, internal report CERN/PS/AA 78–3.
6. F. Bonaudi et al., Antiprotons in the SPS, internal report CERN DG 2.
7. S. van der Meer, Stochastic stacking in the Antiproton Accumulator, internal report CERN/PS/AA 78–22.
8. S. van der Meer, Precooling in the Antiproton Accumulator, internal report CERN/PS/AA 78–26.
9. D. Möhl, Stochastic cooling, Proc. Workshop in Possibilities and Limitations of Accelerators and Detectors, Batavia (FNAL, Batavia, 1978), p. 145.
10. L. Faltin, Slot-type pick-up and kicker for stochastic beam cooling, Nucl. Instrum. Methods **148**, 449.
11. A.G. Ruggiero, Are we beating Liouville's theorem, in Proc. Workshop on Producing High Luminosity high Energy Proton-Antiproton Collisions, Berkeley (LBL, Berkeley, 1978).
12. A.G. Ruggiero, Stochastic cooling with noise and good mixing, ibid.
13. S. van der Meer, Stochastic cooling theory and devices, ibid.
14. H. Herr and D. Möhl, Bunched beam stochastic cooling, Proc. Workshop on the Cooling of High Energy Beams, Madison, 1978 (Univ. Wisconsin, Madison, 1979) and internal report CERN/PS/DL/Note 79–3 (1979) and CERN/EP/Note 79–34 (1979).
15. Design study of a proton-antiproton colliding beam facility, internal report CERN/PS/AA 78–3.

1979

1. J. Bisognano, Transverse stochastic cooling, Berkeley (LBL) internal report BECON-9.
2. J. Bisognano, Vertical transverse stochastic cooling, Berkeley (LBL) internal report BECON-10, LBIB-119.
3. J. Bisognano, Kinetic equations for longitudinal stochastic cooling, Berkeley (LBL) internal report BECON-11, LBID-140.
4. Ya.S. Derbenev and S.A. Kheifets, On stochastic cooling, Particle Accelerators, **9**, 237.
5. Ya.S. Derbenev and S.A. Kheifets, Damping of incoherent motion by dissipative elements in a storage ring, Soviet Phys. Techn. Phys. **24(2)**, 203.
6. Ya.S. Derbenev and S.A. Kheifets, Stochastic cooling, Sov. Phys. Tech. Phys. **24(2)**, 209.
7. G. Carron et al., Experiments on stochastic cooling in ICE, Proc. Particle Accelerator Conf., San Francisco, 1979 [IEEE Trans. Nucl. Sci **NS–26** (3), 1979], p. 3456.
8. S. van der Meer, Debunched $\bar{p}$-p operation of the SPS, internal report CERN/PS/AA 79–42.
9. D. Möhl, Possibilities and limits with cooling in LEAR, Proc. Joint CERN-KfK Workshop on Physics with Cooled Low-Energy Antiprotons, Karlsruhe (KfK report 2836, Karlsruhe, 1979), p. 27.
10. B. Leskovar and C.C. Lo, Low-noise wide-band amplifier system for stochastic beam cooling experiments, Proc. Nuclear Science Symposium, San Francisco, 1979 [IEEE Nucl. Sci **NS–27** (1), 1980], p. 292 (preprint LBL 9841).

1980
1. D. Möhl, G. Petrucci, L. Thorndahl and S. van der Meer, Physics and technique of stochastic cooling, Phys. Rep. **58**, 75.
2. S. van der Meer, A different formulation of the longitudinal and transverse beam response, internal report CERN/PS/AA 80–4.
3. J. Bisognano, Kinetic equations for longitudinal stochastic cooling, 11th Int. Conf. on High Energy Accelerators, Geneva (Birkhäuser, Basel, 1980), p. 772.
4. W. Kells, Filterless fast momentum cooling, ibid., p. 777.
5. G. Lambertson et al., Stochastic cooling of 200 MeV protons, ibid., p. 794.
6. E.N. Demet'ev et al., Measurement of the thermal noise of a proton beam in the NAP-M storage ring, Sov. Phys. Tech. Phys. **25(8)**, 1001.
7. F. Krienen, Initial cooling experiments (ICE) at CERN, 11th Int. Conf. on High Energy Accelerators, Geneva (Birkhäuser, Basel 1980), p. 781.
8. D.E. Young, Progress on beam cooling at Fermilab, ibid., p. 800.
9. P. Lefevre et al., The CERN low energy antiproton ring (LEAR) project, ibid., p. 819.
10. Design study of a facility for experiments with low energy antiprotons (LEAR), internal report CERN/PS/DL 80–1.
11. V.V. Parkhomchuk and D.V. Pestrikov, Thermal noise in an intense beam in a storage ring, Sov. Phys. Tech. Phys. **25(7)**, 818.

1981
1. F.T. Cole and F.E. Mills, Increasing the phase-space density of high-energy particle beams, Ann. Rev. Nucl. Part. Sci. **31**, 295.
2. J. Bisognano and S. Chattopadhyay, Bunched beam stochastic cooling, Berkeley (LBL) internal report BECON-18.
3. A.N. Skrinski and V.V. Parkhomchuk, Methods of cooling beams of charged particles, Sov. J. part. Nucl. **12** (3), 223.
4. S. van der Meer, Stochastic cooling in the CERN Antiproton Accumulator, Proc Particle Accelerator Conference, Washington, DC [IEEE Trans. Nucl. Sci. **NS–28** (3), 1984], p. 1994.
5. J. Hardek et al., ANL stochastic cooling experiments using the FNAL 200 MeV cooling ring, ibid., p. 2455.
6. J. Bisognano and S. Chattophadhyay, Stochastic cooling of bunched beams, ibid., p. 2462.
7. G. Lamberton et al., Experiments on stochastic cooling of 200 MeV protons, ibid., p. 2471.
8. T. Linnecar and W. Scandale. A transverse Schottky noise detector for bunched proton beams, ibid., p. 2147.
9. A.G. Ruggiero, Pickup loop analysis, Fermilab internal $\bar{p}$ note 148.
10. A.G. Ruggiero, Stochastic cooling – a comparison with bandwidth and lattice functions, Fermilab internal $\bar{p}$ note 171.
11. E.N. Demet'ev et al., Experimental study of stochastic cooling of protons in NAP-M, Novosibirsk preprint 81–57 [translated at CERN as internal report CERN/PS/AA 82–3 (1982)].

12. J. Bisognano and C. Leeman, Stochastic cooling, in: The physics of high energy accelerators (Proc. Batavia Summer School on High Energy Particle Accelerators) (AIP Conf. Proc. No 87, New York, 1982), p. 583.

1982

1. D. Möhl, Phase-space cooling techniques and their combination in LEAR, Proc. Workshop on Physics at LEAR with Low-Energy Cooled Antiprotons, Erice (Plenum Press, London, 1983), p. 27.
2. A.G. Ruggiero, Theory of signal suppression for stochastic cooling with multiple systems, Fermilab internal $\bar{p}$ note 193.
3. C. Kim, Design options for the fast betatron precooling systems in the debuncher or on the injection orbit, Berkeley (LBL) internal report BECON-25.
4. S. Chattopadhyay, Stochastic cooling of bunched beams from fluctuation and kinetic theory, Thesis Univ. of Berkeley, Calif. internal report LBL 14826.
5. C.S. Taylor, Stochastic cooling measurements – the CERN AA, Proc. Beam Cooling Workshop, Stoughton (Univ. Wisconsin, Madison, 1982).
6. S. Kramer, LBL/ANL cooling experiments, ibid.
7. L. Tecchio, Comparison of electron and stochastic cooling for intermediate energy range, ibid.
8. F. Voelker, Electrodes for stochastic cooling of the FNAL antiproton source, ibid.
9. B. Autin, Fast betatron cooling in an antiproton collector, ibid.
10. B. Leskovar and C.C. Lo, Low-noise wide-band amplifiers for stochastic beam cooling experiments, ibid.
11. K. Takayama, Effects of non-linear electrode on transverse cooling, ibid.
12. A.G. Ruggiero, 2–4 GHz tail stochastic cooling system with filters, ibid.
13. J. Marriner, Stochastic stacking without filters and accumulator gain profile, ibid.
14. J. McCarthy, Superconducting filters for stochastic system performance measurements, ibid.
15. C.R. Holt, A normal mode analysis of coupled betatron oscillations including cooling, ibid.
16. D. Möhl and K. Kilian, Phase-space cooling of ion beams, Proc. Symposium on Detectors in Heavy-Ion Reactions, Berlin (Vol. 178 of Lecture Notes in Physics, Springer Verlag, Berlin, 1982), p. 220, preprint CERN–EP/82–214.
17. A. Ando and K. Takayama, RF stacking and tail cooling in the antiproton accumulator, Fermilab internal report TM 1103.
18. S. van der Meer, Gain adjustment criterion for betatron cooling in the presence of amplifier noise, internal report CERN/PS/AA Note 82–2.
19. B. Leskovar, Reliability considerations of travelling-wave tube and gallium-arsenide field effect transistor amplifiers, internal report LBL 1381.

1983

1. G. Brianti, Experience with the CERN $\bar{p}p$ complex, Proc. Particle Accelerator Conf., Santa Fe, 1983 [IEEE Trans. Nucl. Sci. **NS–30** (4), 1983], p. 1950.
2. G.R. Lambertson and C.W. Leeman, Intense antiproton source for a 20 TeV collider, ibid., p. 2025.

3. G.R. Lambertson, K.J. Kim and F.V. Völker, The slotted coax as a beam electrode, ibid., p. 2158.
4. B. Leskovar and C.C. Lo, Low-noise gallium arsenide field-effect transistor preamplifiers, ibid., p. 2259 (preprint LBL 15122).
5. F. Völker, T. Henderson and J. Johnson, An array of 1–2 GHz electrodes for stochastic cooling, ibid., p. 2262.
6. S. Chattopadhyay, On stochastic cooling of bunches in the colliding beam mode in high-energy $\bar{p}p$ storage rings, ibid., p. 2334.
7. J. Bisognano, Stochastic cooling: recent theoretical directions, ibid., p. 2393.
8. E. Peschardt and M. Studer, Stochastic cooling in the CERN ISR during $\bar{p}p$ colliding beam physics, ibid., p. 2584.
9. G. Carron, R. Johnson, S. van der Meer, C. Taylor and L. Thorndahl, Recent experience with antiproton cooling, ibid., p. 2587.
10. W. Kells, Stability and signal suppression of Schottky signals from stochastically cooled beams, ibid., p. 2590.
11. B. Autin, J. Marriner, A. Ruggiero and K. Takayama, Fast betatron cooling in the debuncher ring for the Fermilab Tevatron I project, ibid., p. 2593.
12. A. Ruggiero and J. Simpson, Momentum precooling in the debuncher ring for the Fermilab Tevatron I project, ibid., p. 2596.
13. A. Ruggiero, Signal suppression analysis for momentum stochastic cooling with a multiple system, ibid., p. 2599.
14. A. Ando and K. Takayama, Effects of rf-stacking on tail cooling in the Fermilab antiproton accumulator, ibid., p. 2601.
15. S. Chattopadhyay, Vlasov theory of signal suppression for bunched beams interacting with a stochastic cooling feedback loop, ibid., p. 2646.
16. S. Chattopadhyay, Theory of bunched beam stochastic cooling, ibid., p. 2649.
17. S. Chattopadhyay, A formulation of transversely coupled betatron stochastic cooling of coasting beams, ibid., p. 2652.
18. B. Leskovar and C.C. Lo, Travelling-wave tube amplifier characteristics study for stochastic beam cooling experiments, ibid., p. 3423, preprint LBL 14142.
19. R. Billinge and E. Jones, The CERN antiproton source, Proc. 12th Int. Conf. on High Energy Accelerators, Batavia (FNAL, Batavia, 1984), p. 14.
20. B. Autin, Technical developments for an antiproton collector at CERN, ibid., p. 393.
21. S.L. Kramer, J. Simpson, R. Konecny and D. Suddeth, Relativistic beam pickup test facility, ibid., p. 258.
22. N. Tokuda, H. Yonehara, T. Hattori, T. Katayama, A. Noda and M. Yoshizawa, Stochastic cooling of 7 MeV protons at TARN, ibid., p. 386.
23. J. Marriner, The Fermilab antiproton stack tail system, ibid., p. 579.
24. R.E. Shafer, The Fermilab antiproton debuncher betatron cooling system, ibid., p. 581.
25. R.J. Pasquinelli, Superconducting notch filters for the Fermilab antiproton source, ibid., p. 584.
26. S. van der Meer, Optimum gain and phase for stochastic cooling systems, internal report CERN/PS/AA 83–48.
27. Design report Tevatron I project, FNAL internal report (1983).
28. Design study of an antiproton collector for the Antiproton Accumulator, CERN 83–10.

1984

1. S. van der Meer, Stochastic cooling and the accumulation of antiprotons (Nobel lecture 1984), CERN preprint PS 84–32 (AA) published in Rev. Mod. Phys. **57** 1985, p. 689.
2. S. van der Meer, An introduction to stochastic cooling, lecture at the 1984 US Summer School on High Energy Accelerators, CERN preprint PS 84–33 (AA).
3. D. Möhl, Stochastic cooling for beginners, in Proc. CERN Accelerator School, 'Antiprotons for colliding beam facilities', CERN 84–15 (1984), p. 97.
4. C.S. Taylor, Stochastic cooling hardware, ibid., p. 163.
5. S. van der Meer, Optimum gain and phase for stochastic cooling, ibid., p. 183.
6. D. Boussard et al., Feasibility study of stochastic cooling of bunches in the SPS, ibid., p. 197.
7. J. Borer and R. Jung, Diagnostics, ibid., p. 385.
8. P. Bryant, Antiprotons in the CERN intersecting storage rings, ibid., p. 509.
9. B. Autin, The CERN antiproton collector, ibid., p. 525.
10. B. Autin, Antiproton betatron cooling for the SSC, in Proc. DFP workshop on $\bar{p}p$ options for the supercollider (J.E. Pilcher and R. White, ed.) Chicago 1984, p. 311.
11. D. Möhl, Fast stochastic cooling of batches of 10^9 particles, ibid., p. 315.
12. R.P. Johnson, Proton cooling for the SSC, ibid., p. 321.
13. R.P. Johnson and J. Simpson, $\bar{p}$ source group summary, ibid., p. 301.
14. D. Möhl, A comparison between electron cooling and stochastic cooling, Proc. of Workshop on electron cooling, ECOOL 84, Karlsruhe, September 1984, KfK report 3846 (H. Poth, ed.) p. 293 (preprint PS-LEA 85–8).
15. C.C. Lo and B. Leskovar, Cryogenically cooled broad band GaAs field effect transistor preamplifiers, IEEE Trans. Nucl. Sci. **NS–31** (1984), p. 475 and LBL 15983.
16. H.G. Jackson and T.T. Shimzu, A low noise 2–4 GHz preamplifier, LBL 17811.

1985

1. R. Billinge, Introduction to CERN's antiproton facilities for the 1990's, in Proc. 3rd LEAR Workshop, Physics with cooled low energy antiprotons, Tignes, 1985 (Ed. Frontières, Gif-sur-Yvette, 1985), p. 13, preprint CERN/PS/DL/85–10.
2. G. Carron et al., Status and future possibilities of the stochastic cooling system for LEAR, ibid., p. 27, preprint CERN/PS/LEA 85–12.
3. E. Jones, Progress on ACOL, ibid., p. 9, preprint CERN/PS/AA/85–14 (1985).
4. G. Carron et al., Development of power amplifier modules for the ACOL stochastic cooling systems, CERN 85–01 (1985).
5. B. Autin, The new generation of antiproton sources, CERN/PS/AA/85–43 (1985).
6. E. Brambilla, A microwave Cerenkov pick-up for stochastic cooling, CERN/PS/AA/85–45 (1985).
7. N. Tokuda et al., Stochastic cooling of a low energy beam at TARN, IEEE Trans. Nucl. Sci. **NS–32** 5 (1985), p. 2415.
8. F. Caspers, Planar slot line pick-ups for stochastic cooling, CERN/PS/AA/85–48 (1985).
9. C.C. Lo, Stochastic beam cooling amplifier system front end components characteristics, in Proc. 1985 Part. Acc. Conf. Vancouver 1985 (IEEE Trans. Nucl. Sci. **39** 1985), p. 2174.

10. D.A. Goldberg et al., Measurement of frequency response of LBL cooling arrays for TeV-1 storage rings, ibid., p. 2168.
11. J.K. Johnson and R. Nemetz, Power combiners/dividers for loop pickup and kicker array for FNAL stochastic cooling ring, ibid., p. 2171.
12. W.C. Barry, Suppression of propagating TE modes in the FNAL antiproton source stochastic beam cooling system, ibid., p. 2424.
13. P. Lebrun, S. Milner and A. Poncet, Cryogenic design of the cooling pickups for the CERN antiproton collector (ACOL), Adv. cryog. eng. 31 (1986) 543, preprint CERN/LEP/MA 85–29.
14. H.G. Jackson et al., Preliminary measurements of gamma ray effects on characteristics of GaAs field effect transistor preamplifiers, LBL Berkeley report PUB-5141, LBL-21546.

1986

1. D. Boussard and G. Di Massa, High frequency slow wave pick-ups, CERN/SPS/ARF/86–4 (1986).
2. D. Boussard, Schottky noise and beam transfer function diagnostics, CERN Accelerator School, Oxford, 1985, CERN 87–03, p. 416, preprint CERN/SPS/ARF/86–11 (1986).
3. N. Tokuda, A Helix coupler for a pick-up of a low velocity beam, CERN/PS/LEA/Note 86–5 (1986).
4. D. Möhl, Perspectives of ion cooling rings, preprint CERN-PS/LEA 86–32, published in Proc. 1986 International Cyclotron Conference, Tokyo Oct. 1986, p. 116.
5. D. Möhl, Principle and technology of beam cooling, CERN-PS/LEA 86–31, published in Proc. of the RNCP-Kickuchi summer school on accelerator technology, Osaka Oct. 1986.
6. E.J.N. Wilson, Antiproton production and accumulation, CERN/PS/AA 86–22 (1986).
7. N. Beverini et al., stochastic cooling of charged particles in a Penning trap, Europhys. Lett. 1, p. 435 (1986).
8. S. Mtingwa, stochastic cooling of antiprotons at the Tevatron, Fermilab internal report $\bar{p}$ Note 445 (January 1986).
9. F.E. Mills, Cooling of stored beams, Fermilab internal report, $\bar{p}$ Note 463 (10 June 1986).
10. J. Marriner, Stochastic cooling at Fermilab, preprint Fermilab-Conf-86/124 (August 1986) (Presented at the XIII International Conference on High Energy Accelerators, Novosibirsk, USSR, 7–11 August 1986).
11. V.V. Parkhomchuk, The cooling of heavy particles, presented at the XIII International Conference on High Energy Accelerators Novosibirsk, USSR, 7–11 August 1986.

1987

1. D. Möhl, Stochastic cooling, CERN Accelerator School, Oxford, 1985, CERN 87–03, p. 453. (the present report is a preprint with history and references updated).
2. F. Caspers, Planar slot line pickup for stochastic cooling, preprint CERN PS-AA 86–48, Proc. 1987 IEEE Part. Acc. Conf., Washington, DC., p. 1866.
3. B. Autin, G. Carron, F. Caspers, ACOL stochastic cooling systems, ibidem, p. 1549, preprint CERN PS 87–24 AA.

4. G. Lambertson, Dynamic Devices – Pickups and Kickers, in Proc. of the US Particle Accelerator School, (1984/85) (AIP Conf. Proc. 153, M. Month and M. Dienes eds.), (1987) p. 1415.
5. N. Tokuda, Formulation of a flat-helix coupler for a pick-up of a low velocity beam, CERN PS/LEA/Internal Note 87–05, March 1987.
6. J. Marriner, Stochastic cooling in the debuncher, FERMILAB $\bar{p}$, Note 436.
7. B. Autin, G. Carron, F. Gaspers, Applications of microwaves to antiproton control, Proceedings of the 17th European Microwave Conference, Rome, 1987, Alta Freq. 56 (1987) p. 381.
8. G. Carron, F. Caspers, L. Thorndahl, Microwave power amplifiers for the ACOL stochastic cooling systems at CERN, July 1987, Mikrowellen Magazin. preprint CERN PS 87–59 AA.
9. S.K. Mtingwa, Fast 4–8 GHz debuncher betatron stochastic cooling for the Tevatron upgrade, FNAL report TM 1469 (1987).

1988

1. F. Nolden, G. Berthes, D. Böhne, B. Franzke, A. Schwinn, Stochastic precooling of heavy fragments in the ESR, Proc. of the European Part. Acc. Conf. (EPAC 88) Rome 1988, p. 579.
2. J. Bosser, P. Bramham, C. Metzger, Design and performance of the AAC stack core cooling system, Rome 1988, ibidem, p. 1421, preprint CERN PS 88–25 AR.
3. B. Autin, G. Carron, F. Caspers, Performance of the CERN antiproton accumulator complex, ibidem, p. 392, preprint CERN PS 88–43 AR.
4. B. Autin, M. Battiaz, M. Bell, The CERN antiproton collector ring, ibidem, p. 308, preprint CERN PS 88–43 AR.
5. E. Jones, The CERN $p\bar{p}$ facility, ibidem p. 215, preprint CERN PS 89–07 DI.
6. D. Boussard, Advanced cooling techniques: stochastic cooling of bunches in high-energy colliders, Proceedings of the Joint US-CERN school on particle accelerators – topical course on frontiers of particle beams, South Padre Island, 1986, Proceedings edited by M. Month and S. Turner, Springer, Berlin (1988) p. 269, preprint CERN-SPS 87–29 (RF).
7. L. Evans, The proton-antiproton collider, CERN 88–01.
8. P. Lefèvre, LEAR, present status, future and developments, Proceedings of the 4th LEAR workshop, Villars, 1987, Physics at LEAR with low-energy antiprotons, edited by C. Amsler et al., Harwood: Chur (1988), p. 19.
9. S. Baird, J. Bengtsson, M. Chanel, Machine aspects for LEAR operation with jetset, ibidem, p. 91.
10. N. Tokuda, Travelling wave pickups and kickers for stochastic cooling at low beam velocity, ibidem, p. 135.
11. S. Baird, D. Möhl, N. Tokuda, The Lear stochastic cooling system: present and future, ibidem, p. 175.
12. E. Jones, Initial performance of AAC, ibidem, p. 7, preprint CERN PS 87–93 AA.
13. L. Evans, E. Jones, H. Koziol, The CERN $p\bar{p}$ collider, 'World Scientific series on Directions in High Energy Physics', Vol. 4: 'Proton Antiproton Collider Physics', World Scientific Publishing Co, Singapore 1988.

14. N. Beverini et al., Stochastic cooling in Penning traps, Phys. Rev. **A38**, p. 107 and Hyperfine Inter., 44, p. 247. **1989**
1. J. Petter, D. McGinnis and J. Marriner, Novel stochastic cooling pickups/kickers, Proc. Part. Acc. Conf. Chicago, Ill (1989) p. 636 .
2. D. Krämer, The Heidelberg heavy ion cooler storage ring TSR, ibidem, p. 49.
3. D. McGinnis, J. Petter, J. Marriner, J. Misek, and S. Y. Hsueh, Frequency response of 4–8 GHz stochastic cooling electrodes, ibidem, p. 639.
4. D.A. Goldberg and G.R. Lambertson, Behaviour of power-limited stochastic cooling systems, ibidem, p. 642.
5. S. Baird et al., Stochastic cooling and storage test with oxygen ions in LEAR, ibidem, p. 645.
6. J. Petter et al., New 4–8 GHz core cooling pickups and kicker for the Fermilab accumulator, ibidem, p. 648.
7. J. Kewell, AAC studies, CERN PS 89–46 AR.
8. Fermilab team: Accelerator upgrade, phase I, FNAL report April 1989.

1990
1. F. Caspers, Recent developments in stochastic cooling, in: Proc. European Part. Acc. Conf. (EPAC 90) Nice, France, p. 169.
2. T. Tanabe et al., Electron Cooling Experiments at TRAN II, ibidem, p. 1571.
3. M. Chanel, R. Macaferri, J.C. Perrier, Momentum stochastic cooling with digital notches, ibidem, p. 1574.
4. J. Marriner et al., Bunched beam cooling in the FNAL antiproton accumulator, ibidem, p. 1577.
5. K. Bongart, S. Martin, D. Prasuhn, H. Stockhorst, Theoretical analysis of transverse stochastic cooling in the Cooler-Synchrotron COSY, ibidem, p. 1583.
6. E. Jaeschke, Heavy ion storage rings, ibidem, p. 77.
7. B. Autin, G. Carron, F. Caspers, Antiproton Accumulation Complex (AAC) performance, ibidem, p. 614.
8. D.A. Goldberg and G.R. Lambertson, Improving the performance of power-limited transverse stochastic cooling systems, Proc. 14th Internat. Conf. on High Energy Acc., Tsukuba, Japan, Part. Acc. 27, (1990) p. 7/965.
9. D.A. Golberg and G.R. Lambertson, Successful observation of Schottky signals at the Tevatron Collider, ibidem, p. 1/959.
10. V. Lagomarsino, G. Manuzio, G. Testera, Stochastic cooling of the radial motion in a Penning trap, preprint INFN (Genova) BE 90–12, Phys. Rev. **A44**, p. 5173.

1991
1. J. Bosser, Overview of stochastic and electron cooling at LEAR, 1991, Conf. on Low energy antiproton physics: Proceedings edited by P. Carlson, A. Kerek and S. Szilagyi, World Sci., Singapore (1991) p. 486.
2. A.G. Ruggiero, Beam cooling for the ADRIA project, in Proc. of the Workshop on electron cooling and new cooling technologies (R. Calabrese and L. Technicho, eds.) World Sci., Singapore (1991).
3. G. Carron, The CERN Antiproton Accumulator Complex (AAC): current status and operation for the nineties, in Proc. 1991 IEEE Particle Accelerator Conf. (San Francisco) p. 106.

4. P. Brittner et al., The stochastic cooling system for COSY-Jülich, ibidem, p. 1425.
5. D. Goldberg, G. Lambertson, Dynamic devices: a primer on pickups and kickers, LBL-31664.
6. T. Katayama and A. Noda (eds.), Cooler rings and their applications, Proceedings of the INS Symposium Tokyo, World Scientific Publishing Company, Singapore (1991).
Status reports of cooling rings are given in the following contributions to this symposium.

R.E. Pollock, The IUCF cooler after three years, p. 4.
E. Jaeschke, The Heidelberg Heavy Ion Cooler Ring TSR.
H. Eickhoff, K. Beckert, B. Franczak, B. Franzke, F. Nolden, H. Poth, U. Schaaf, H. Schulte, P. Spädtke and M. Steck, storage and cooling of Heavy Ions in the ESR up to 200 MeV/u, p. 11.
A. Johansson and D. Reistad, Present Status of CELSIUS, p. 15.
T. Katayama, Cooler Synchrotron TARN II, present and future, p. 21.
P.A. Chamouard and J. M. Lagniel, SATURNE II and MIMAS status report, p. 31.
C.J. Herrlander, CRYRING, a low energy Heavy Ion facility, p. 40.
A.V. Dolinsky, The Ukrainian (INR, Kiev's) storage ring, p. 48.
R. Maier et al., Status of the COSY-Jülich Project, p. 52.

Amongst them TARN, ESR, and COSY apply or plan to apply stochastic cooling.

1992

1. H. Geissel et al., First storage and cooling of secondary heavy-ion beams at relativistic energies, GSI report 92–20.
2. P. Brittner et al., The stochastic cooling system of COSY, Proceeding of the Third European Part. Acc. Conf. (EPAC 92), Berlin, p. 842.
3. T. Tanabe et al., Heavy-Ion cooling and related applications in the cooler-ring TARN II at INS, ibidem, p. 830.
4. J. Wei, Stochastic cooling with a Double RF system, ibidem, p. 833.
5. J. Wei, Bunched Beam stochastic cooling, Proc. 15th Internat. Conf. on High Energy Acc. (HEACC '92), Hamburg, p. 1028.

1993

1. A.A. Mikhailichenko, M.S. Zolotorev, Optical stochastic cooling, preprint SLAC PUB 6272, Phys. Rev. Lett. **71** (1993) p. 4146.
2. M. Church, J. Marriner, The antiproton sources design and operation, Annual Rev. Nucl. Part. Sci. **43**, p. 43.
3. G. Jackson et al., Bunched beam stochastic cooling in the Fermilab Tevatron collider, Proceedings of the 1993 IEEE Part. Acc. Conf. (Washington DC), p. 3533.
4. K. Beckert et al., Heavy ion beam accumulation, cooling and experiments at the ESR, ibidem, p. 1641.

Status reports on cooling rings in the same Proceedings (1993 IEEE Part. Acc. Conf.) are:

S.P. Møller, The Aarhus storage ring for ions and electrons ASTRID, p. 1741.

D. Reistad, Recent developments at the Gustaf Werner Cyclotron and CELSIUS, p. 1744.

M. Grieser, et al., ion beam acceleration and new operation modes at the TSR Heidelberg, p. 1747.

5. D. Möhl, Cooling of Particle Beams, in: Advances of Accelerator Physics and Technology, Directions in High Energy Physics Vol. 12, H. Schopper ed. World Scientific Publishing Company, Singapore, p. 358.

1994

1. J. Bosser (Ed.): Proceedings of the Workshop on beam cooling and related topics, Montreux October 93, CERN 94–03. Papers dealing with stochastic cooling and related questions are:

A.M. Sessler and R.L. Mills, Liouville's theorem and phase-space cooling, p. 3.

J. Marriner, Theory, technology, and technique of stochastic cooling, p. 14.

G. Jackson, Bunched beam stochastic cooling in the Fermilab Tevatron Collider, p. 127.

J. Wei, Stochastic cooling and intra-beam scattering in RHIC, p. 132.

F. Nolden et al., Stochastic cooling of fragment beams at GSI in Darmstadt: concepts and status, p. 137.

F. Caspers, M. Chanel, and J.C. Perrier, Upgrading of the LEAR stochastic cooling systems, p. 207.

A. Dainelli, A.G. Ruggiero, and L. Tecchio, The beam cooling system of the ADRIA proposal, p. 212.

A. Ruggiero, Derivation of a Fokker-Planck equation for bunched beams, p. 219.

T. Katayama, and M. Chanel, Report of the working session on stochastic cooling, p. 436.

2. M.S. Zolotorev, A.A. Zholents, Transit-time method of optical stochastic cooling, preprint SLAC PUB 6476, LBL 35469, Phys. Rev. **E50**, p. 3084.

3. A.M. Sessler, The cooling of particle beams, preprint LBL 36302, invited talk at the Tamura Symposium on Accelerator Physics, Austin, TX, USA, November 1994.

4. P. Brittner et al. Stochastic cooling system in COSY, Proc. 1994, European Part. Acc. Conf. (EPAC 94) London, p. 1203.

5. F. Caspers, G. Dôme, Reciprocity between pick-up and kicker structures including the far-field zone, ibidem, p. 1208.

6. A.A. Mikhailichenko, Damping ring for testing the optical stochastic cooling method, ibidem, p. 1214.

References on selected topics

Workshops on beam cooling (devoted mostly or partly to stochastic cooling)

Workshop on Phase-Space Cooling, Brookhaven, 1976. (Summary report edited by E. Courant, published in Proc. 1976 ISABELLE Workshops, Brookhaven Nat. Lab. report BNL 50611, 1976, p. 241).

Workshop on Producing High Luminosity High Energy Proton-Antirpoton Collisions, Berkeley, 1978. (Copies of transparencies and reports presented, published by LBL, Berkeley, 1978).

Workshop on Cooling of High Energy Beams, Madison, 1978. (Proc. edited by D. Cline, published by the University of Wisconsin, Madison, 1979).

Workshop on $\bar{p}p$ in the SPS, Geneva, 1980. (Summary report on High Energy Beam Cooling edited by F. Mills, Proc. published as report CERN/SPS-$\bar{p}p$-1, Geneva, 1980).

Beam Cooling Workshop, Stoughton, 1982. (Proc. edited by D. Cline and F. Mills, published by University of Wisconsin, Madison, 1982).

Workshop on $\bar{p}p$ options for the supercollider, Chicago 1984 (Proc. edited by J.E. Pilcher and R. White A.N.L. report, Feb. 1984).

Symposium on cooler rings and their applications, Tokyo, 1990 (Proc. edited by T. Katayama and A. Noda, World Scientific, 1991).

Workshop on beam cooling and related topics, Montreux 1993 (Proc. edited by J. Bosser (CERN-report 94–03).

Design reports of antiprotons facilities using stochastic cooling:
(Numbers refer to the Bibliography)

AA design report, 1978:15.
LEAR design report, 1980:10.
TEVATRON design study, 1983:27.
ACOL design study, 1983:28.
FERMILAB upgrade 1989:8.

Summary reports and reviews:

1977:1
1978: 3 and 13
1980:1
1981: 1, 3 and 12
1982: 1, 7
1983: 7, 7
1984: 1, 2, 3 and 4
1986: 5
1987:3
1988: 7, 13
1990:1
1993: 2,5

Special problems of cooling of heavy-ion beams are discussed in

1981: 3
1982: 16
1986: 4
1989: 5
1990: 5, 6
1991: 6
1992: 1, 3
1993: 4

Bunched beam stochastic cooling

1978: 14
1981: 2, 6 and 8
1982: 4
1983: 6, 15 and 16
1984: 6
1988: 6
1990: 4, 9
1992: 4, 5
1993: 3
1994: 2,6

Stochastic cooling at optical frequencies:

1993: 1
1994: 2,6

APPENDIX 1

HISTORY

For a long time the principle of stochastic cooling was regarded as too far-fetched to be practical. A first experimental demonstration was tried only seven years after its invention and three years after the first publication of the idea (Table A1). The inventor, S. van der Meer, and the early workers had (mainly) emittance cooling of high-intensity beams in mind with a view to improving the luminosity in the CERN ISR. The proof of principle in the ISR was an important achievement of this early period.

A new era began in 1975 when P. Strolin, coming back from a visit to Novosibirsk, and L. Thorndahl realized the interest of stochastic cooling — for both emittance and momentum — of low-intensity $\bar{p}$ beams for the purpose of stacking. Stochastic cooling at low intensity is different from the original van der Meer cooling. The extension of the theory first done by H. Hereward and L. Thorndahl, and the design of the momentum cooling hardware (L. Thorndahl, G. Carron et al.), are perhaps as fundamental as the original invention and the earlier feasibility studies by S. van der Meer and W. Schnell.

Following upon this broadening of the field of application, in 1975 P. Strolin et al. worked out $\bar{p}$ collection schemes for the ISR using stacking in momentum space, and C. Rubbia et al. made their first proposals of the $\bar{p}p$ scheme for the SPS using similar techniques of stochastic cooling and accumulation. This work gave new life to the idea at a time when the ISR was routinely stacking such high currents that proton beam-cooling became unnecessary — or even impossible. Further milestones between 1975 and 1980 were the invention of the filter method of momentum cooling, the refinement of the stochastic stacking schemes, the results of the initial cooling experiment (ICE), and last, but not least, the success of the AA. The ICE ring was used to make a careful comparison of cooling theory with reality, including bunched beam cooling. By the middle of 1978 all systems worked so well that beam lifetimes at 2 GeV/c of the order of a week were reached, compared with lifetimes of a few hours without cooling. This permitted a measurement of the stability of the antiproton, and this experiment improved the lower limit in one go from 120 μs to 80 h.

One essential ingredient in this experiment was the technique developed at ICE to observe as few as 50 circulating particles in a non-destructive manner. This was made possible by stochastic cooling which reduced the momentum spread to 10^{-5} so that a resonant Schottky noise pick-up with the corresponding quality factor could be used.

Running-in of the AA started in the summer of 1980, and since 1981/82, stacks of several 10^{11} $\bar{p}$ were routinely accumulated from batches of a few 10^6 $\bar{p}$ per second. The AA used a total of seven cooling systems for longitudinal cooling of different 'regions' of the beam and for horizontal and vertical emittance cooling. Time constants were of the order of a second at up to 5×10^6 $\bar{p}$ or 30 minutes for $5 \cdot 10^{11}$ $\bar{p}$, thus nearing design specifications. The AA was at the heart of CERN's antiproton programme, which culminated with the observation in the Sp$\bar{p}$S collider of the Intermediate Vector Bosons predicted by the unifying electroweak theory.

In the ISR stochastic 'post-cooling' of antiprotons from the AA was used (amongst other applications) to improve the beam lifetime and the resolution in conjunction with

an internal hydrogen jet target. In this way charmonium states formed in proton antiproton interactions could be observed with high precision. This was another important achievement of the ISR just prior to its final shutdown. In fact the very last ISR beam was such an antiproton beam circulating in ring 2. It was finally dumped at 6.00 h on 25 June 1984 thus bringing to a definite end the glamorous career of a unique machine. The use of an internal target in conjunction with phase-space cooling of the circulating beam — as proposed by the Novosibirsk group many years before — had thus been put to work for the first time. This arrangement has stimulated much interest as an option for LEAR and as one of the basic techniques for the new ion cooling rings.

The low-energy antiproton ring LEAR, which after ISR and SPS became the third customer of the AA, has given high quality $\bar{p}$ beams to more than 20 user groups on an operational basis since 1983. One particularity of LEAR is an 'adjustable' system which allows stochastic cooling at many different energies. In 1985 alone, beam was extracted at momenta between 60 MeV/c (corresponding to 2 MeV kinetic energy) and 2 GeV/c ($\sim$ 1.3 GeV). Relatively fast cooling with time constants of 1 to 3 minutes for typically 5×10^9 $\bar{p}$, works at different strategic energies; slow cooling with time constants of 10–20 minutes can be used at practically all momenta to keep the beam in shape during the one to five hour extraction. Since the stop of the SPS antiproton programme, end of 1991, LEAR is the only customer of the antiprotons. Since 1990 an internal gas target together with an advanced detector ('JETSET') is installed in straight section 2 of LEAR. JETSET, which runs typically with fills of 4×10^{10} $\bar{p}$, relies heavily on stochastic cooling to obtain a high luminosity (10^{30} cm^{-2} sec^{-1}) and long luminosity lifetime ($\approx$ 15 h).

From about 1978 onwards, other laboratories, especially the Novosibirsk group, who had pioneered electron cooling before, an ANL-LBL-Fermilab collaboration and, more recently, a group at the INS Tokyo, have done work (both experimental and theoretical) on stochastic cooling. This work has placed emphasis on various important aspects such as low-noise cryogenic amplifiers, very high frequency systems, cooling of heavy ions, or cooling of bunches.

In 1983 the Tevatron-1 project at Fermilab and the $\bar{p}$ collector AC to be added to the AA at CERN were approved. Both systems use stochastic cooling and stacking of antiprotons at a rate of several 10^7 $\bar{p}$ per second. New features are cryogenically cooled components on the low-level side (amplifiers, terminating resistors and to some extent even cables and the pick-up plates themselves) to improve the signal-to-noise ratio. In addition bandwidths well in the Gigaherz range are used. The Fermilab antiproton source consists (presently) of two rings, a debunching ring and accumulator ring. Like the CERN collector ring, the 8 GeV debuncher uses fast emittance cooling ($\tau \approx 2$ sec at 7×10^7 particles) of the $\bar{p}$-pulse from the production target. The accumulator ring combines stochastic cooling for stacking in momentum space with emittance cooling to improve the transverse density of the stack. In the accumulator alone there is a total of six specific cooling systems working in the 1 to 2, 2 to 4 and partially in the 4–8 GHz range, each of them combining a large number of pick-up and kicker loops. The system for momentum cooling of the stack tail — which is the largest system — uses more than 100 pick-up and kicker units.

The AC ring at CERN came into operation 1987. Together with the AA, which was modified to work in cascade with the AC, it improved the $\bar{p}$ flux by an order of magnitude

compared to that available with the AA alone. The number and the complexity of the cooling systems of AA-AC is as impressive as in the Fermilab case.

There has thus been a rapid development of stochastic cooling during the decade from 1975-1985 and roughly one order of magnitude has been gained every four years in the cooling power, i.e. in the number of particles which can be cooled with a time constant of 1 s. This has become possible by making larger and larger bandwidths available. Probably an 'absolute' limit in the range of 10^9 to 10^{10} particles per second is reached with the 'conventional' technology which cannot use frequencies much higher than 10 GHz. At these frequencies (most) vacuum chambers transmit waveguide modes and the beam size becomes comparable to the RF-wavelength. Recent proposals discuss stochastic cooling at optical frequencies and bandwidths (Δ f $\simeq 10^{14}$ Hz!). This requires a completely new pick-up, kicker, waveguide and amplifier technology and is also hurt by serious problems of 'mixing' between observation and correction.

Since about 1980, interest in cooling of heavy ion beams developed rapidly, and the combination of stochastic 'pre-cooling' with post-cooling by electrons looks attractive for some applications. A number of ion cooling rings with some resemblance to LEAR have come into operation in the USA, Japan and Europe. All of them foresee electron cooling and many plan to use both electron- and stochastic damping. Three ion coolers: TARN II at the INS Tokyo; the IUCF-cooler at Bloomington Indiana and CELSIUS at Uppsala came into operation in 1988. Five others, TSR at the MPI Heidelberg, ESR at GSI Darmstadt, ASTRID at Aarhus, CRYRING at Stockholm, and COSY at KfK Jülich, followed. Thus in the recent years phase space cooling is used to a large extent at low and medium energy.

For the very highest energies, ideas on bunched beam cooling are being followed up. Studies have been carried out on stochastic cooling of bunches in the SPS collider and more recently for the TEVATRON at Fermilab as well as for the HERA proton ring at DESY and for the relativistic heavy ion collider (RHIC) at BNL. Experiments at the TEVATRON, the Fermilab accumulator ring and at the AAC- and LEAR rings at CERN all seem to indicate that 'RF activity' at frequencies much higher than the bunch frequency is present which poses a problem for the cooling of bunched beams. This is now being actively investigated. Early in 1995 the observation of the top quark at the Fermilab Tevatron was announced which is then another important discovery made possible by stochastic cooling.

Table A1
History

		Prehistory
Liouville	1838	Invariance of phase space area
Schottky	1918	Noise in d.c. electron beams
MURA group (Lichtenberg, Mills, Sessler, Stahle, Symon)	1956–58	Liouville's theorem applied to particle storage rings
		History
van der Meer	1968	Idea of stochastic cooling
ISR staff (Borer, Bramham, Hereward, Hübner, Schnell Thorndahl)	1972	Observation of proton beam Schottky noise in the ISR
van der Meer	1972	Theory of emittance cooling
Schnell	1972	Engineering studies
Hereward	1972–74	Refined theory, low-intensity cooling
Bramham, Carron, Hereward, Hübner, Schnell Thorndahl	1975	First experimental demonstration of emittance cooling
Palmer (BNL), Thorndahl	1975	Idea of low-intensity momentum cooling
Strolin, Thorndahl, Möhl	1975	$\bar{p}$ accumulation, schemes for the ISR using stochastic cooling
Rubbia	1975	$\bar{p}$ accumulation, schemes for the SPS
Thorndahl, Carron	1976	Experimental demonstration of momentum cooling
Thorndahl, Carron	1977	Filter method of momentum cooling
Sacherer, Thorndahl, van der Meer	1977–78	Refinement of theory; imperfect mixing, Fokker-Planck equations
ICE team	1978	Detailed experimental verification
Herr, Möhl	1978	Demonstration of bunched beam cooling
AA team	1981–82	Accumulation of several 10^{11} $\bar{p}$ from batches of several 10^6, cooling times close to design specifications
LEAR team	1982	Stochastic cooling to permit loss-free deceleration of $\bar{p}$. Interest in combining stochastic and electron cooling.
Kilian, Möhl	1982	Stochastic cooling of heavy ions (proposal)
Novosibirsk group	1980–82	Stochastic cooling experiments, work on cooling theory
ANL-LBL-Fermilab group	1979–83	Theoretical and hardware studies, stochastic cooling experiments at FNAL
TARN group at INS Tokoyo	1983–84	Stochastic cooling experiments in the TARN ring
Fermilab group	1983	Design report of the Tevatron-1 project using a debuncher ring and an accumulator ring for fast stochastic cooling and stacking of 4×10^7 $\bar{p}$ per second.
AA team	1983	Design report of the $\bar{p}$ collector ACOL using fast momentum and emittance cooling and stacking of 4×10^7 $\bar{p}$ per second.
Berkeley group	1983	$\bar{p}$ collection scheme for 20 TeV colliders. Collection of 5×10^8 $\bar{p}$ per second
SPS $\bar{p}$ team	1983	Feasibility study of bunched beam cooling in the SPS collider
Fermilab group	1985–86	Construction and running in of $\bar{p}$ source
AA-AC team	1986–87	Construction of the AC
Groups in several different labs	1986–93	Commissioning of 8 different (heavy) ion cooling rings
AA-AC team	1990	Record stacking rate of 6.13×10^{10} $\bar{p}$/h
Fermilab group	1990–94	Upgrading of the $\bar{p}$-source. Record stacking rate of 6.8×10^{10} $\bar{p}$/h (1994).
Several groups	1993	Bunched beam stochastic cooling in large colliders (studies) (Tevatron, RHIC, HERA...)
Mikhailichenko, Zohlents, Zolotorev	1993–94	Stochastic cooling at optical frequencies (proposal!)
Fermilab	1995	Observation of the top quark

APPENDIX 2

DISPERSION INTEGRALS

In this appendix we wish to have a brief look at the dispersion integral Eq. (3.60) as required for the 'signal shielding' calculations. A more general discussion is given by H.G: Hereward [15] in the context of Landau damping.

To deal with the singularity of the integrand we assume that the eigenfrequency ωj of the test particle has a small imaginary part, i.e. we take $\omega j \rightarrow \omega + i\alpha$ such that the free oscillation $a \cdot e^{i\omega_j t}$ corresponding to Eq. (3.55) is damped. Later we go for the limit $\alpha \rightarrow 0$.

With a complex eigenfrequency Eq. (3.60) becomes

$$\tilde{S}(\omega) = -\frac{1}{2\omega} \int \frac{\tilde{G}(\omega_j) n(\omega_j)[(\omega_j - \omega) - i\alpha]}{(\omega_j - \omega)^2 + \alpha^2} \, d\omega_j \, .$$

We are especially interested in the contribution due to the damping term α:

$$\tilde{S}_i(\omega) = -\frac{i}{2\omega} \int \frac{\tilde{G}(\omega_j) n(\omega_j) i\alpha}{(\omega_j - \omega)^2 + \alpha^2} \, d\omega_j \, .$$

The main contribution to this integral comes from the range $\omega_j = \omega \pm \alpha$, near the pole. For small α we can usually assume that $G(\omega_j)$ and $n(\omega_j)$ are constant in this range and thus take the weighting function $G(\omega_j)n(\omega_j) \simeq G(\omega)n(\omega)$ out of the integral. Integrating the rest from a minimum $\check{\omega}_j < \omega$ to a maximum eigenfrequency $\hat{\omega}_j > \omega$

$$\tilde{S}_i \approx -\frac{i}{2\omega} \tilde{G}(\omega) n(\omega) \left[\arctan\left(\frac{\hat{\omega} - \omega}{\alpha}\right) - \arctan\left(\frac{\check{\omega} - \omega}{\alpha}\right) \right]$$

and in the limit $\alpha \rightarrow 0$

$$\tilde{S}_i \approx -\frac{i}{2\omega} \tilde{G}(\omega) n(\omega) \cdot \pi \, .$$

Clearly this is the residuum due to the pole of the integrand. Due to the physics of the problem the value -iπ has to be retained.

The remaining part of the integral is the principal value. It can be expressed in terms of the Hilpert transform (see Erdélyi et al., Tables of integral transforms Vol. 2, MacGraw Hill N.Y. 1954) defined by

$$H[f(x)] = g(y) = \frac{1}{\pi} \int \frac{f(x)}{x - y} \, dx \, .$$

This transform has been tabulated for a large collection of functions. In terms of the Hilpert transform the principal value of Eq. (3.60) may be written as

$$\tilde{S}_p(\omega) = \frac{\pi}{2\omega} H[n(\omega_j)\tilde{G}(\omega_j)] \, .$$

Further details depend on $n(\omega_j)$ and $G(\omega_j)$. For betatron cooling $\tilde{G}(\omega_j)$ is ideally constant (and purely imaginary and negative) whereas $n(\omega_j)$ is 'bell shaped' around the average

betatron frequency ω_β (the centre of the Schottky band). It is therefore convenient to work in terms of the deviation from ω_β denoting $x = \omega_j - \omega_\beta$, $y = \omega - \omega_\beta$. Then

$$\tilde{S}(y) = \frac{\pi}{2\omega}[-i\tilde{G}(y) \cdot n(y) + H(\tilde{G}(x) \cdot n(x)] \ .$$

Two distributions can serve as models and permit the construction of approximations like Eq. (3.61) above:

1) The semi-circular distribution (which models a distribution with a sharp cutt-off)

$$n(x) = \begin{cases} \frac{2}{\pi\Delta^2}\sqrt{\Delta^2 - x^2} & |x| \leq \Delta \\ 0 & \text{elsewhere} \end{cases}$$

For constant $\tilde{G}$ one obtains

$$\tilde{S}(y) = \frac{-\tilde{G}}{\omega_\beta \Delta^2} \ (y + i\sqrt{\Delta^2 - y^2}) \ .$$

2) The Lorentzian distribution (which models a distribution with important tails):

$$n(x) = \frac{\delta/\pi}{\delta^2 + x^2}$$

which for constant $\tilde{G}$ yields

$$\tilde{S}(y) = \frac{-\tilde{G}}{2\omega_\beta}\left[\frac{y + i\delta}{\delta^2 + y^2}\right]$$

APPENDIX 3

THE FACTORS OF TWO

There may be some confusion about a factor of 2 in the simple equation (2.31) etc. For momentum spread cooling the assumptions underlying this equation are fulfilled, thus

$$\frac{1}{\tau_{\Delta p}} = \frac{W}{N}\left[2g(1-\widetilde{M}^{-2}) - g^2(M+u)\right]$$

For betatron oscillation cooling a factor of $\frac{1}{2}$ enters due to the phase relationship as sketched in Figs. 4 and 5. Therefore the amplitudes (and the r.m.s. beam radius) are cooled according to

$$\frac{1}{\tau_a} = \frac{1}{\tau_\sigma} = \frac{1}{2}\,\frac{W}{N}\,[\ldots]$$

It is customary to quote the cooling rates for Δp and σ_x and σ_y.

If however one uses the cooling rate for the transverse emittance ε, then (as $\varepsilon\alpha\sigma^2$) all three rates are described by the same relation

$$\frac{1}{\tau_{\Delta p}} = \frac{1}{\tau_\varepsilon} = \frac{W}{N}[\ldots]\,.$$

All rates are 'instantaneous cooling decrements' defined by $\frac{1}{\varepsilon}\,\frac{d\varepsilon}{dt}(t)$ etc. Only in special cases is this rate independent of time and only then the cooling is purely exponential.

ELECTRON COOLING

J. Bosser

CERN, Geneva, Switzerland

ABSTRACT

The aim of any cooling process is to reduce the emittances of an ion beam as quickly as possible. Amongst the few processes in use, the one using *cold* electron beams and called *Electron Cooling* is efficient when applied to low-energy ion beams. After a short reminder of the principles and of some essential definitions, a more descriptive than mathematical explanation of the cooling forces will be given. A full section will be devoted to the diagnostics while another will aim to emphasise the technological aspects. Finally, the secondary effects on the ion beam and a short review of the present and future coolers are presented.

1. INTRODUCTION, PRINCIPLE OF ELECTRON COOLING

Electron cooling was first proposed by Prof. G. Budker in the late sixties as a method of improving the properties of stored ion beams. Its experimental confirmation was performed at Novosibirsk and later at CERN and Fermilab. Due to the high energy of the production maximum for antiprotons, stochastic cooling, proposed by S. van der Meer, was the most appropriate choice of the cooling technique for their accumulation. In spite of that, the use of *Electron Cooling* in low energy ion rings, in order to improve the beam lifetime and properties, has had spectacular results and will open the way to better resolutions in nuclear and atomic physics experiments with stored beams.

Let us consider an ion beam circulating in a storage ring at average velocity $\vec{v}_0 = \vec{\beta}_0 c$ along the theoretical trajectory (Fig. 1.1). In an orthogonal frame $\Re_0$ moving at velocity $\vec{v}_0$ most of the ions will not be at rest. Their relative velocity can be pictured as in Fig. 1.1b).

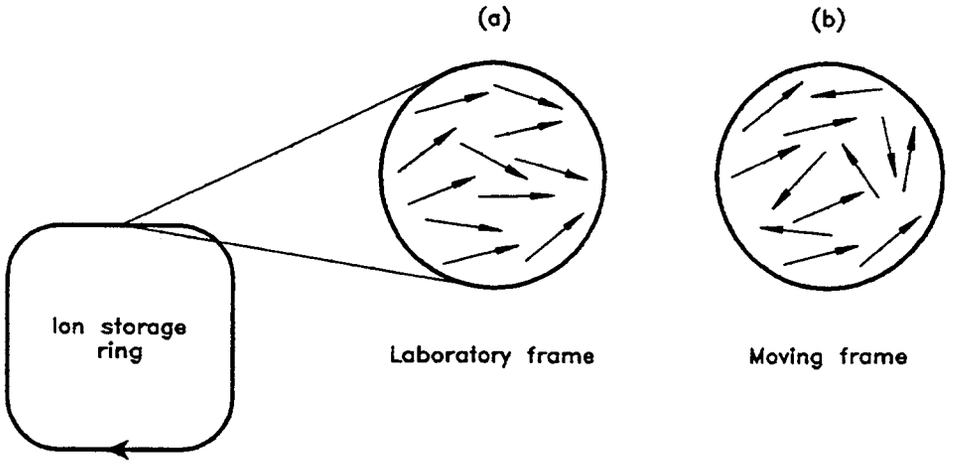


Fig. 1.1 Illustration of electron cooling of an ion beam:

a) Schematic of the ion velocity in the lab frame
b) The ion beam velocity in the moving frame where the electrons are at rest

More precisely the relative velocity of each individual ion can be expressed as the sum of a longitudinal and a transverse component (Fig. 1.2); such that:

$$\vec{v}_i = v_{\parallel} \cdot \vec{u}_{\parallel} + v_{\perp} \cdot \vec{u}_{\perp}$$

$$\vec{v}_i = v_x \cdot \vec{i} + v_z \cdot \vec{j} + v_s \cdot \vec{k};\ (v_s = v_{\parallel}).$$

where the symbol $\parallel$ means relative to the axis colinear or parallel to $\vec{v}_0$, $\perp$ to the plane orthogonal to $\vec{v}_0$, and the vectors $(\vec{u}_{\parallel}, \vec{u}_{\perp})$ and $(\vec{i}, \vec{j}, \vec{k})$ form an orthonormal basis of the $\Re_0$ space.

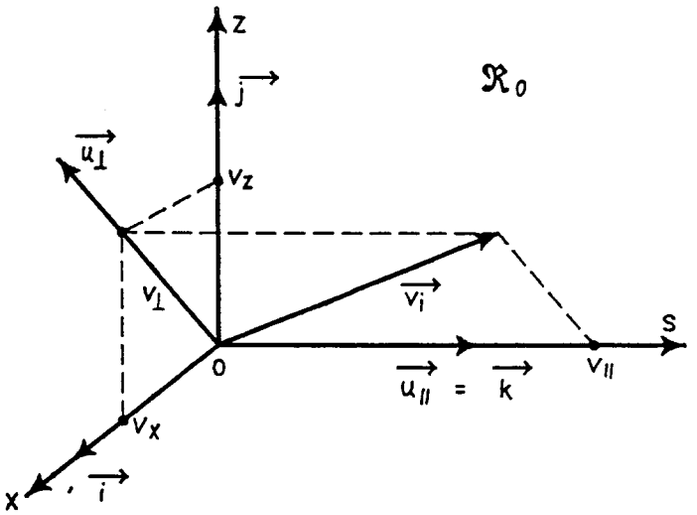


Fig. 1.2 Description of the velocity components in the moving frame $\Re_0$

The rms ($\sqrt{\langle v^2 \rangle}$) velocities are closely related to the emittances. We will have to be more precise on that subject but at the level of this introduction we can say that the emittance ε (in each plane) is proportional to $\langle v^2 \rangle$. The cloud or gas of ions has therefore, in $\Re_0$, a kinetic energy which can be compared to the definition of the temperature T of a dilute gas in thermodynamics using the classical relation

$$\frac{3}{2} k \cdot T = \frac{1}{2} m_i \cdot \langle v_i^2 \rangle = \frac{1}{2} m_i \cdot \left[\langle v_x^2 \rangle + \langle v_z^2 \rangle + \langle v_s^2 \rangle \right]$$

where k is the Boltzmann constant, and m_i the ion mass.

We can thus associate for each individual rms velocity component a temperature

$$k \cdot T_x = m_i \cdot \langle v_x^2 \rangle,\ k \cdot T_z = m_i \cdot \langle v_z^2 \rangle,\ k \cdot T_s = m_i \cdot \langle v_s^2 \rangle$$

or:

$$k \cdot T_{\parallel} = m_i \cdot \langle v_{\parallel}^2 \rangle,\ k \cdot T_{\perp} = m_i \cdot \langle v_{\perp}^2 \rangle .$$

In fact many processes will contribute to an increase of the ion rms velocities. Amongst the most important are the collisions with the residual gas molecules within the beam pipe and the intrabeam scattering. These effects become significant for low-momentum ions. Their

emittance will continuously increase until they reach the machine acceptance and be lost. The ion-beam lifetime is then relatively short.

The goal of any cooling process is to reduce each *temperature component* and thus the corresponding emittances or at least to counteract the diffusion effects mentioned above. As a result the emittances will be reduced and the ion-beam lifetime increased. Figure 1.3 represents the effect of cooling on the size (1.3a) and on the momentum spread (1.3b) of a stored ion beam.

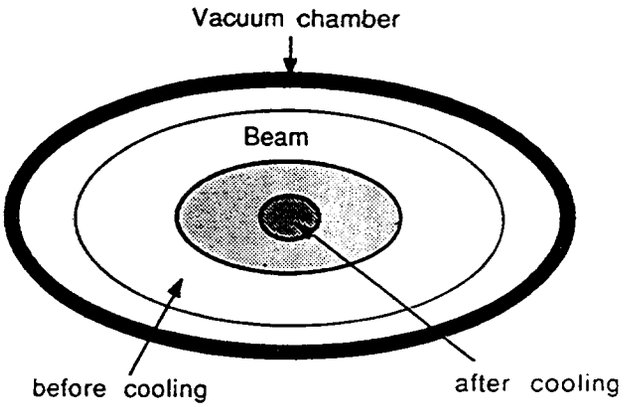


Fig. 1.3a) Schematic effect of ion beam cooling on beam size

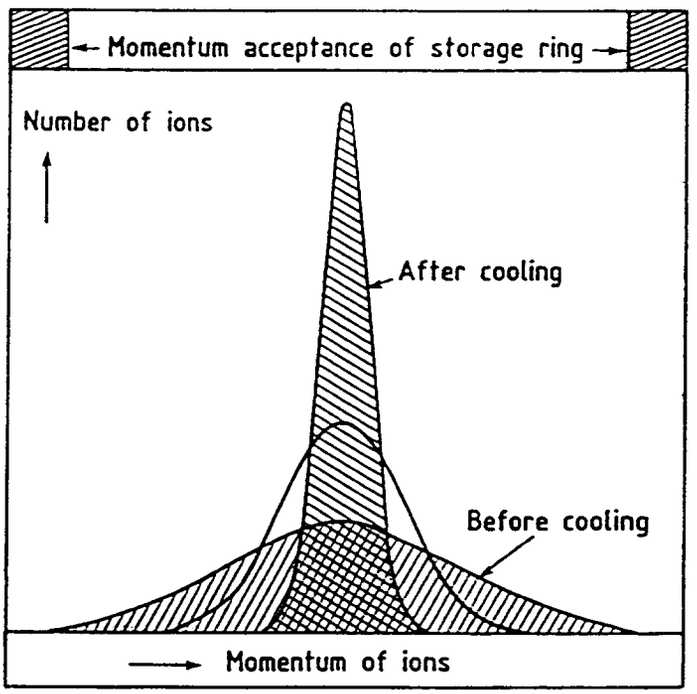


Fig. 1.3b) Schematic effect of cooling on momentum spread

Amongst many cooling processes the one using cold electrons and named *Electron Cooling* or e-cool offers the possibility of obtaining relatively small emittances in short cooling times. This is mainly efficient for non-relativistic ion beams which will be the case in this lecture.

The principle consists of immersing the ion beam in a very cold (in the moving frame $\Re_0$) electron beam over a given length. If we suppose, at first, that the electron beam has no velocity and therefore no energy in $\Re_0$, due to Coulomb interaction the ions will undergo "collisions" with electrons (binary collision model). As a result the ions will give up some of their energy to the electrons which will therefore be heated. As a consequence the electrons must be renewed in order to obtain a very cold (in each plane) ion beam.

In practice the electrons, produced by a cathode and accelerated under a voltage U_0 such that $e \cdot U_0 \cong 1/2\,(m_e \cdot v_0^2)$, are not motionless in $\Re_0$. We will see that they have a flattened velocity distribution such that $\langle v_{e\parallel}^2 \rangle << \langle v_{e\perp}^2 \rangle$. The goal of an efficient electron gun is to obtain transverse electron velocities (or $\langle v_{e\perp}^2 \rangle$) as small as possible.

According to our didactic explanations we can understand that electron cooling will proceed until the ion and the electron relative energies (always in $\Re_0$) become equal. That means that:

$$\mathrm{k} \cdot T_e = \mathrm{k} \cdot T_i \qquad or \qquad m_i \cdot v_i^2 = m_e \cdot v_e^2$$

$$v_i = v_e \left[\frac{m_e}{m_i}\right]^{1/2}$$

and since the ion mass m_i is much greater than the electron mass m_e, the ion velocity will become very small. The cooling time is usually of the order of 10 ms to 10 s after which the ions are at very low temperature and therefore the ion beam has very small emittances.

We can sketch the principle as in Fig. 1.4. The electrons are produced continuously by a cathode heated to T_C. They are accelerated to velocity $\vec{v}_0 = \vec{\beta}_0 \cdot c$ by a gun and directed into the drift or cooling region where they overlap the ion beam over a length ℓ_c. At the end of this section the electrons are steered away from the ions and recuperated by a collector. On their way from the gun to the collector the electrons are usually submitted to a longitudinal magnetic field which has two purposes:

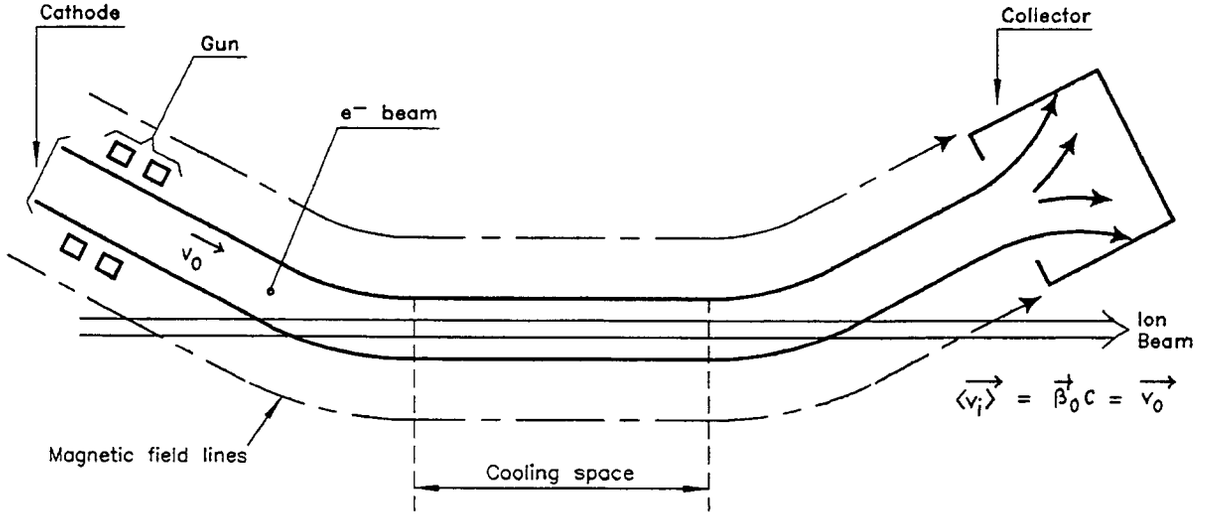


Fig. 1.4 Principle of electron cooling

- First, the potential due to the electron space charge inside the beam is not constant. This will be explained later but, as a consequence, a radial electric field exists which will tend to make the e-beam diverge. The magnetic force will counteract this effect and keep the e-beam cylindrical.

- Second, we will see (section 5) that the fact that electrons are "magnetised" improves drastically, at relative low ion speed, the cooling effect. The cooling time is appreciably reduced.

We can leave electron cooling for the moment and look at two didactic examples:

a) One may take some analogy from thermodynamics where the entropy $S = K\ell n[(\Omega)]$ (Ω the number of microstates) is a measure of the disorder. For a quasi-static transformation $dS = dQ/T$ where dQ represents the quantity of heat received at absolute temperature T. If a small body (representing the ions) of mass m and specific heat c_p at initial temperature T_1 is immersed in a large quantity of water (representing the electrons) at constant temperature $T_2 < T_1$, it will cool down to T_2 and its change in entropy will be:

$$\Delta S = mc_p \ell n(T_2 / T_1) < 0 .$$

A negative change of entropy results in a reduction of the disorder (and by analogy more orderly ions).

b) Another example could be that of a ball traversing a "sand wall" or, more explicit for our purpose, that of a moving foil (Fig. 1.5). One may then consider the electrons as a foil moving with velocity $\vec{v}_0 = \vec{\beta}_0 c$. Ions moving faster than the foil (electrons) penetrate it and lose energy (dE / dx) along the direction of their momentum during each passage, until all transverse component are diminished and their longitudinal velocity is equal to the foil velocity. Slower ions traverses the foil from the opposite side. Ideally at the end all ions will have the same longitudinal velocity as the foil and no transverse component.

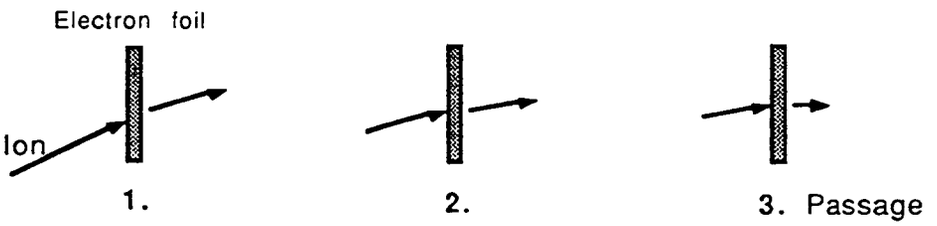


Fig. 1.5 Illustration of cooling as an energy loss in a moving foil

The energy loss can be calculated from the Bohr equation:

$$-\frac{dE}{dx} = \frac{4\pi Z^2 n_e e^4}{(4\pi\varepsilon_0)^2 m_e v_i^2} \alpha_1 \qquad (1.1)$$

where Z is the charge on the ion, n_e the electron density and α_1 is a material constant.

Due to multiple scattering, the beam will blow up as it passes a foil of thickness x_0. This will also cause a smearing of the longitudinal energy, known as energy straggle. The rms width θ_{ms} of the angular distribution is given by:

$$\theta_{ms}^2 = \frac{2\pi Z^2 e^4 n_e}{(4\pi\varepsilon_0)^2 E_i^2} x_0 \alpha_2 \tag{1.2}$$

where E_i is the kinetic energy of the ion and α_2 is another material constant. We therefore distinguish two effects:

i) the energy loss (friction);
ii) the multiple scattering (diffusion).

The first effect results in cooling, the second one corresponds to heating. The diffusion is characterised by the coefficient D, defined as:

$$D = \frac{d}{dt}(p_i \cdot \theta_{ms}^2)$$

where p_i is the momentum of the ion). The energy loss and the diffusion coefficient are related through:

$$\frac{dE}{dx} = \frac{1}{2m_e} \cdot \frac{\partial D}{\partial v_i} \quad (\text{for } \alpha_1 = \alpha_2).$$

Coming back to the electron-cooling process itself, one sees that the production of a cold electron beam by the gun is of importance. On the other hand the efficiency of the e-beam recuperation by the collector will influence the vacuum pressure and indirectly the cooling process.

2. LIST OF CONSTANTS AND TYPICAL PARAMETERS OF AN ACCELERATOR

Most of the constants which will be used in this chapter are listed below. The accelerator parameters listed will be mainly used for numerical examples to get a feeling of the electron cooling process. At the end are given some fundamental relations between the laboratory frame and the moving frame components.

2.1 Usual constants

Boltzmann constant k	$1.38066 \cdot 10^{-23}\ \mathrm{J \cdot K^{-1}} = 8.618 \cdot 10^{-5}\ \mathrm{eV \cdot K^{-1}}$
Speed of light c	$3 \cdot 10^8\ \mathrm{ms^{-1}}$
Elementary charge e	$1.602 \cdot 10^{-19}$ C (or s · A)
Permeability of free space μ_0	$4\pi \cdot 10^{-7}\ \mathrm{H \cdot m^{-1}}$ (or $\mathrm{m \cdot kg \cdot s^{-2} \cdot A^{-2}}$)
Permitivity constant of free space ε_0	$8.854 \cdot 10^{-12}\ \mathrm{F \cdot m^{-1}}$ (or $\mathrm{m^{-3} \cdot kg^{-1} \cdot s^4 \cdot A^2}$) $= \frac{1}{\mu_0 \cdot c^2}; \frac{1}{4\pi\varepsilon_0} \cong 9 \cdot 10^9$
A, Z ion atomic mass, charge	
Proton mass m_p	$938.28\ \mathrm{MeV}/c^2 = 1.672 \cdot 10^{-27}$ kg $(e/m_p) = 9.579 \cdot 10^7$ C / kg

Electron mass m_e $\quad 0.511\,MeV/c^2 = 9.109\cdot10^{-31}\,kg$

$(e/m_e) = 1.758\cdot10^{11}\ C/kg$

Plasma frequency $\quad \omega_{pl}^2 = \dfrac{n_e e^2}{m_e \varepsilon_0} = 4\pi\cdot n_e \cdot r_e \cdot c^2; \quad \omega_{pl} = 56.5(n_e)^{1/2} = 2\pi f_{pl}\ ,\ s^{-1}$

(n_e is the number of electrons in m^{-3}).

Classical radius $\quad r = \dfrac{e^2}{4\pi\varepsilon_0 mc^2} = \dfrac{\ell^2}{mc^2} \quad = \begin{cases} r_e = 2.818\cdot10^{-15}\ m \text{ for electron} \\ r_p = 1.547\cdot10^{-18}\ m \text{ for proton} \end{cases}$

$$\ell^2 = \frac{e^2}{4\pi\varepsilon_0} = 2.3\cdot10^{-28}\ m^3\cdot kg\cdot s^{-2}.$$

The subscript "*i*" will refer to ions while "*e*" will refer to electrons. Cylindrical coordinates are defined in Fig. 2.1 where (0,*s*) shows the beam direction. An element of volume *dv* is given by: $dv = r\,dr\,d\theta\,ds$.

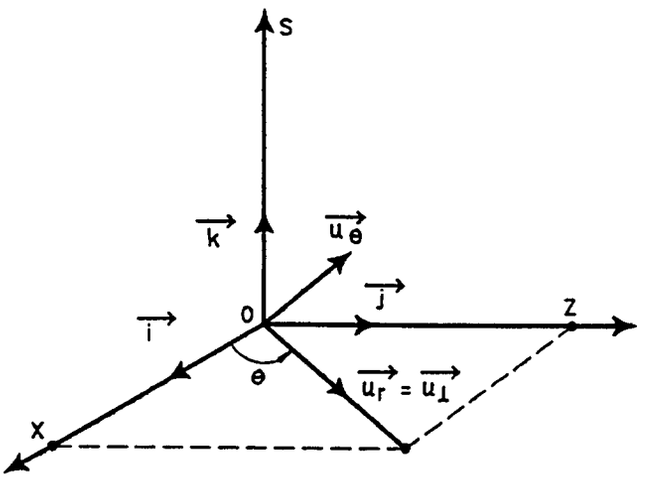


Fig. 2.1 Cylindrical coordinates

rms: If the symbol <X> represents the average or expectation of a random variable X then the rms value of X is expressed by $\sqrt{<X^2>}$.

2.2 Typical accelerator and electron cooler (LEAR) parameters

The ions will be protons ($Z = 1$)	
Dispersion in cooler D	3.58 m
Vertical betatron function in cooler β_v	5.3 m
Horizontal betatron function in cooler β_h	1.9 m
Horizontal tune Q_h	2.305
Vertical tune Q_v	2.730
Accelerator circumference	78.59 m
Off-momentum function $\eta = \dfrac{df/fo}{dp/po} = \dfrac{1}{\gamma_0^2} - \dfrac{1}{\gamma_{tr}^2}$	0.938
Proton momentum p_0	308.6 MeV/c
Proton energy T_0	44.446 MeV

Proton velocity $\vec{\beta}_0 = \vec{v}_0 / c$	0.3124 (γ_0= 1.0527 ; $\beta_0 \gamma_0$= 0.3288)
Nominal revolution frequency f_0	1.192 MHz
Current for 10^9 p circulating in the machine	0.191 mA
Vacuum pressure	10^{-11} Torr [N_2 equivalent for scattering]
Electron energy U_0	26929 eV ≅ (27 keV)
Electron beam intensity I_b	2.3 A
Perveance p_g	$0.52 \cdot 10^{-6}$ A.V$^{-3/2}$
Electron beam radius r_0	2.5 cm
Electron beam density (in the lab. frame) n_e^*	$8.27 \cdot 10^{13}$ m^{-3}
Current density J_c^*	0.13 A/cm^2
Longitudinal magnetic field B_0	455 Gauss
Toroid angle and radius	$\Phi_0 = 36^\circ$ R_t = 1.05 m
Length of cooling section ℓ_c	1.5 m
η_c = length of cooling section/circumference	$1.9 \cdot 10^{-2}$
Number of protons N	$5 \cdot 10^9$

For numerical examples

case 1, $\varepsilon_h = \varepsilon_v$ = 10 π·mm·mrad, $\Delta p/p_0 = 10^{-3}$, case 2, $\varepsilon_h = \varepsilon_v$ = 1 π·mm·mrad, $\Delta p/p_0 = 10^{-4}$.

2.3 Change of frame (Fig. 2.2)

The symbol * denotes the lab. frame. The moving frame moves at speed $\vec{v}_0 = \vec{\beta}_0 \cdot c$ along the (0,s) axis, "||" is related to a component parallel to the (0,s) axis, for example $\vec{v}_{||}$ or $\vec{v}_s$ are colinear with $\vec{v}_0$, "⊥" is related to a component in the plane (0,x,z) orthogonal to the (0,s) axis.

Force $\begin{cases} f_{||}^* = f_{||} \\ f_\perp^* = f_\perp / \gamma_0 \end{cases}$ (Newtons)

Time $t^* = \gamma_0 \cdot t$ (s)

Electron density $n^* = \gamma_0 \cdot n$ (m^{-3})

Velocities $\vec{v}_{||} = \dfrac{\vec{v}_{||}^* - \vec{v}_0}{1 - v_{||}^* \cdot \frac{v_0}{c^2}}$, $\vec{v}_\perp = \dfrac{\vec{v}_\perp^*}{\gamma_0 \left(1 - v_{||}^* \cdot \frac{v_0}{c^2}\right)}$ (m·s^{-1})

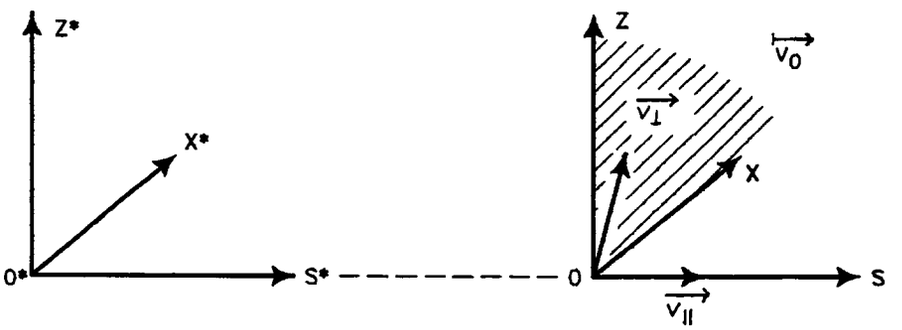


Fig. 2.2 Fixed and moving coordinates

3. EMITTANCES, TEMPERATURES

Emittances are important parameters of any accelerator. Since electron cooling makes use of temperatures, it is worth establishing the relations which exist between emittances and temperatures. When an electron is accelerated under a voltage U_0 (U_0 of the order of a few kV) its longitudinal initial temperature is drastically reduced. We will calculate the scale of this

effect. Finally, some approximate formula for the processes which usually increase the emittance of low energy storage rings are given.

3.1 Emittances

In Fig. 3.1, we consider the observer frame (0*, x*, z*, s*) or rest frame and a moving frame (0, x, z, s) moving at velocity $\vec{v}_0 = \beta_0 \cdot c$ along the (0, s) axis $\left(\gamma_0 = 1/(1-\beta_0^2)^{1/2}\right)$. We will also consider almost non-relativistic particles; this is particularly the case in the moving frame. When necessary, we will use the symbol y to define either x (horizontal plane) or z (vertical plane).

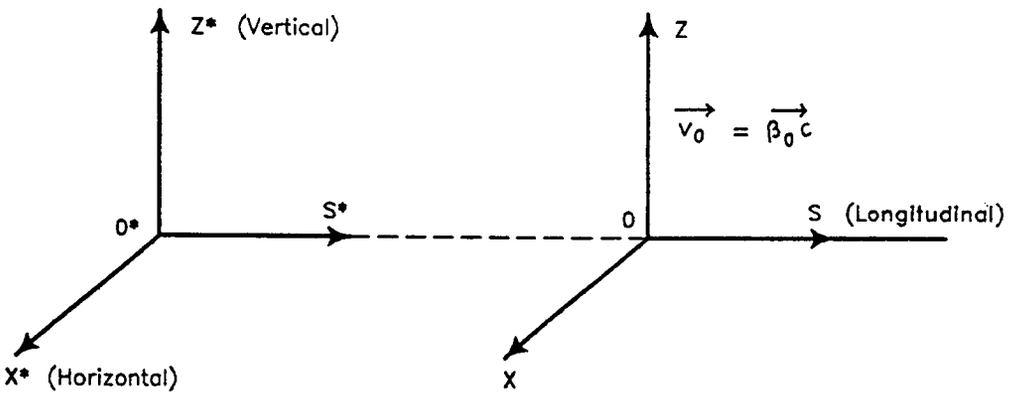


Fig. 3.1 Fixed and moving frame

An important parameter of any particle beam is the rms emittance ε_y related to the beam width $\sigma_y = \sqrt{\langle y^2 \rangle}$ or to its divergence $\theta_y = \sqrt{\langle (dy/ds)^2 \rangle}$ such that (we consider $d\beta_{h,v}(s)/ds = 0$ in the cooler section)

$$\sigma_h = \sigma_x = \left(\frac{\varepsilon_h \beta_h(s)}{\pi}\right)^{1/2} \quad , \quad \theta_x = \left(\frac{\varepsilon_h}{\pi \beta_h(s)}\right)^{1/2} = \theta_h$$

$$\sigma_v = \sigma_z = \left(\frac{\varepsilon_v \beta_v(s)}{\pi}\right)^{1/2} \quad , \quad \theta_z = \left(\frac{\varepsilon_v}{\pi \beta_v(s)}\right)^{1/2} = \theta_v$$

$\beta_{h,v}(s)$ being the betatron amplitude function (in meters), while ε_y, σ_y and θ_y are usually given in π·mm·mrad, mm and mrad respectively. The normalized emittance ε_n, when expressing the emittance as $[\pi\varepsilon] = \pi \cdot \text{mm} \cdot \text{mrad}$,

$$\varepsilon_n = \sigma_y \cdot \theta_y \cdot \beta_0 \cdot \gamma_0 = \varepsilon_y \cdot \beta_0 \cdot \gamma_0$$

remains constant in the absence of cooling and heating.

For circular accelerators the mean value of the betatron function is approximately given by:

$$\langle \beta_h \rangle = \frac{R}{Q_h} \quad ; \quad \langle \beta_v \rangle = \frac{R}{Q_v}$$

where Q_h, Q_v are the tunes in the two planes and R the accelerator radius.

3.2 Temperatures

It is well known, in thermodynamics, that the internal energy U of a gas is related to its absolute temperature T by:

$$U = \frac{3}{2} \mathrm{k} \cdot T .$$

For a dilute gas its internal energy is related to the rms velocity v by:

$$U = \frac{3}{2} \mathrm{k} \cdot T = \frac{1}{2} m \cdot v^2 = \frac{1}{2} m \left[< v_x^2 > + < v_z^2 > + < v_s^2 > \right]$$

or, by definition: $\mathrm{k} \cdot T_p = m \cdot v_p^2 = 2 \cdot Ec_p$ (Ec_p is the kinetic energy, and the subscript p refers to each plane).

Note: v_p is the rms velocity spread in each plane, $\sqrt{3/2} \cdot v_p$ is the total rms velocity spread when one assumes an isotopic distribution of the velocities. $1/2\, m\, v_p^2$ is the kinetic energy spread per plane. It is sometimes convenient to express T in eV, (the reader must remember that 1 eV $\simeq$ 11000 K) and later omit k.

3.2.1 Ion transverse temperature

We can apply these considerations to a beam by working in the particle rest frame $\Re_0$ which moves around with the nominal particle velocity. The beam temperatures are then given by the velocity spreads in this frame. Under these conditions it is easy to show that the temperature is, in the moving frame, given by:

$\mathrm{k} \cdot T_h = m_i \cdot c^2 \cdot \beta_0^2 \cdot \gamma_0^2 \cdot \theta_h^2$ horizontally and $\mathrm{k} \cdot T_v = m_i \cdot c^2 \cdot \beta_0^2 \cdot \gamma_0^2 \cdot \theta_v^2$ vertically.

The transverse temperature $\mathrm{k} \cdot T_\perp = \mathrm{k}(T_h + T_v)$ will be, according to our simplifications:

$$\mathrm{k} \cdot T_\perp = m_i \cdot c^2 \cdot \beta_0^2 \cdot \gamma_0^2 \cdot \left[Q_h + Q_v \right] \frac{\varepsilon}{R} \quad \text{when } \varepsilon_h = \varepsilon_v = \varepsilon.$$

Thus the reduction of $T_\perp$ corresponds to a reduction of the transverse emittance and vice versa.

3.2.2 Ion longitudinal temperature

The particle may have a small velocity spread difference dv^* around $v_0 = \beta_0 \cdot \mathrm{c}$. We know that $p_0 = m_i \cdot \gamma_0 \cdot v_i^*$ so that $\Delta p = m_i \left(\gamma_0 \Delta v^* + v_0 \Delta \gamma \right)$. In the moving frame then:

$$\mathrm{k} \cdot T_\parallel = m_i \cdot c^2 \cdot \beta_0^2 \left(\frac{\Delta p}{p_0} \right)^2 .$$

A reduction of the momentum spread Δp results in a reduction of the longitudinal temperature $T_\parallel$ and vice versa.

3.2.3 Electron temperatures

The temperature T_c of the electrons at the output of a cathode is defined as:

$$k \cdot T_c = k\left(T_{c\parallel} + T_{c\perp}\right)$$

where $T_{c\parallel}$ and $T_{c\perp}$ are respectively the longitudinal and transverse rms electron temperature spreads and $k \cdot T_c$ is about 0.1 eV. Let us apply to the electron beam emitted by the cathode, an accelerating voltage U_0 such that $e \cdot U_0 \cong (1/2) m_e v_0^2$, v_0 being the velocity of the moving frame.

In the fixed, or laboratory frame, the transverse electron temperature remains unchanged after acceleration: $k \cdot T_{e\perp}^* = k \cdot T_c$ (with $k \cdot T_c$ of the order of 0.1 to 0.5 eV). Therefore in the moving frame:

$$k \cdot T_{e\perp} = \gamma_0^2 (k \cdot T_{e\perp}^*) \cong k \cdot T_c .$$

The longitudinal temperature is, however, drastically reduced after acceleration to the nominal kinetic energy $E_c = (1/2) m_e v_0^2$. Let us look carefully, in the non relativistic case, at this phenomenon in order to get an estimate of the longitudinal temperature rms spread $T_{e\parallel}$ in the moving frame. We can write, in the laboratory frame, that the kinetic energy E_c after the electron beam has been accelerated under a voltage U_0 is:

$$E_c = \frac{(m_e v_{e\parallel}^*)^2}{2} = e \cdot U_0 + k \cdot T_{c\parallel} = W_0 + k \cdot T_{c\parallel} .$$

Let us write: $v_{e\parallel}^* = v_0 + \Delta v_e^*$ (with $\Delta v_e^* << v_0$) and $\frac{m_e v_0^2}{2} = W_0$;

Δv_e^* is then the rms electron beam longitudinal velocity spread we are looking for. Thus:

$$\frac{m_e}{2}\left(v_0^2 + 2 v_0 \cdot \Delta v_e^* + \Delta v_e^{*2}\right) = \frac{m_e}{2} v_0^2 + k \cdot T_{c\parallel} ,$$

$$m_e \cdot v_0 \cdot \Delta v_e^* \cong k \cdot T_{c\parallel} ,$$

$$\Delta v_e^* = \frac{k \cdot T_{c\parallel}}{m_e v_0} = \Delta v_{e\parallel} .$$

If we use the equality $k \cdot T_{e\parallel} = m_e (\Delta v_{c\parallel})^2$ then:

$$k \cdot T_{e\parallel} = \frac{\left(k \cdot T_{c\parallel}\right)^2}{2\left(e \cdot U_0\right)} .$$

If for example we take $k \cdot T_{c\parallel} = 0.1\,\text{eV}$ and since according to our typical accelerator table $U_0 = 27$ kV:

$$k \cdot T_{e\parallel} = 1.7 \cdot 10^{-7}\ \text{eV} << k \cdot T_c \cong 2\, k \cdot T_{e\perp} .$$

This is called the kinematic contraction or flattened distribution effect of the electron beam. It shows that in the next sections of this lecture, we can use the inequality $k \cdot T_{e\parallel} << k \cdot T_{e\perp}$.

On the other hand, if the accelerating voltage U_0 has a ripple of about 1 V, this can be considered as an effective cathode longitudinal temperature spread $k \cdot T_{c\parallel} = 1$ eV such that:

$$k \cdot T_{e\parallel} = 1.7 \cdot 10^{-5} \text{ eV}$$

and again, in the moving frame, the equivalent electron temperature spread is much smaller in the longitudinal plane than in the transverse one.

Numerical example

It is left as an exercise to compute the ion and electron transverse and longitudinal velocities in $\Re_0$. For that purpose cases 1 and 2 of section 2.2 have to be used and for the electron one should take $k \cdot T_c = 0.5$ eV and 1 eV. Then compare $v_{e\parallel}$, v_i, $v_{e\perp}$.

Exercise
Show that in the relativistic case: $k \cdot T_{e\parallel} = (k \cdot T_{c\parallel})^2 / \beta_0^2 \gamma_0^2 m_e c^2$.

3.3 Natural heating of the ion beam

Secondary effects will run counter to cooling and, without cooling, will lead to an emittance growth so that the circulating ion beam will be lost after several minutes. Among all these heating phenomena, we will only mention the effect of the residual gas and the intrabeam scattering.

3.3.1 Residual gas disturbance

Repeated small-angle scattering of stored ions by residual gas molecules leads to:

a) Emittance growth. The angular spread can be calculated with Eq. (1.2):

$$\theta_{ms}^2 = \frac{8\pi Z_{gas}^2 \cdot Z^2 \cdot n_{gas} \cdot r_i^2 \cdot c \cdot L_{res}}{\beta_0^3 \gamma_0^2} t ,$$

where we replaced α_2 by the Coulomb logarithm responsible for this scattering ($L_{res} \cong 10$), x_0 by $\beta_0 c \cdot t$ and $E_i = (\gamma / 2) m_i \cdot v_i^2$.

The transverse emittance growth is then:

$$\frac{\dot{\varepsilon}_{ms}}{\pi} = \frac{8\pi Z_{gas}^2 \cdot Z^2 \cdot n_{gas} \cdot r_i^2 \cdot c < \beta_y > \cdot L_{res}}{\beta_0^3 \gamma_0^2} \cdot \frac{1}{2} ,$$

$$n_{gas} = 3.5 \cdot 10^{16} \text{ Torr}^{-1} \text{ cm}^{-3} \cdot P_{gas} .$$

where the factor 1/2 comes from averaging over the betatron phase.

When the emittance equals the machine acceptance A_x or A_z, the ions begin to be lost.

b) Energy loss which can be calculated from Eq. (1.1) after including the effects of the residual gas molecules:

$$-\frac{dE}{dx} = \frac{4\pi Z_{gas} \cdot Z^2 \cdot n_{gas} \cdot r_e^2 \cdot m_e \cdot c^2}{\beta_0^2} L_{res}$$

where we put $\alpha_1 = \alpha_2 = L_{res}$. Again, when the particles have been decelerated down to the inner side of the machine longitudinal acceptance the beam is lost.

We can compare the emittance growth rate with the energy loss:

$$-\frac{dE/dt}{E_i} = \frac{\dot{\varepsilon}_{ms}}{<\beta>}\frac{m_i}{m_e}.$$

More detailed explanations are given in the lecture on "Beam Interactions with Residual Gas" in this course [4].

3.3.2 Intrabeam scattering

An elaborate description of this process is given in the lecture on "Intrabeam scattering" in these proceedings [5], from where the reader can find out the emittance growth time τ_{ibs} and how it evolves when the ion emittance or temperature decreases. Other phenomena have to be taken into account. Some of them will be explained later but, of course, physics experiments using internal targets [6] will contribute to a fast degradation of the emittances. This motivates the use of cooling techniques (stochastic, electronic or other) which aim to counteract these diffusion processes and consequently improve the ion beam quality and lifetime.

4. DEFINITIONS OF THE COOLING FORCES AND TIMES

The cooling process can be explained in terms of plasma physics (Refs. [1 to 3]). Explanations, however, become more difficult in the case of magnetised electrons. We will therefore use a classical explanation, using electromagnetic laws, after a brief reminder of a few plasma definitions. Except for an elementary, and in consequence, approximate demonstration, electron cooling requires tedious and difficult computations which are out of the scope of this lecture. We give here just a few parameters and definitions which will be of some use in the section on cooling forces.

4.1 Reminder of plasma physics

Imagine a positive ion embedded in the electron cloud. The neighbouring electrons will rearrange themselves to shield the ion Coulomb field. The electric field decays exponentially and the shielding radius is called the Debye length λ_d, (see Fig. 4.1).

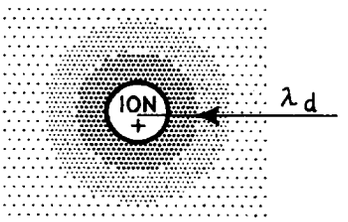


Fig. 4.1 Debye radius

$$\frac{1}{\lambda_d^2} = \frac{e^2}{\varepsilon_0 \cdot \mathrm{k}}\left[\frac{n_e}{T_e} + \frac{n_i \cdot Z^2}{T_i}\right]$$

$$\lambda_d \cong \left[\frac{\varepsilon_0 \cdot \mathrm{k} \cdot T_e}{n_e \cdot e^2}\right]^{1/2}$$

when $n_e >> n_i$ and where T_e is the electron beam temperature in K.

The Debye length can be compared to the average radius between the electrons:

$$\frac{4}{3}\pi r_{av}^3 = \frac{1}{n_e} \quad or \quad r_{av} = \left[\frac{3}{4 \cdot \pi \cdot n_e}\right]^{1/3} .$$

The time it takes for the perturbed electrons to rearrange themselves around the ion is about one period of the plasma frequency:

$$f_{pl}^2 = \frac{n_e \cdot r_e \cdot c^2}{\pi} = \frac{1}{(t_{pl})^2}$$

If the electron gas has a very low temperature and if the ion moves at relative velocity v_i the screening length ℓ_s is about (Fig. 4.2):

$$\ell_s = v_i \cdot t_{pl} .$$

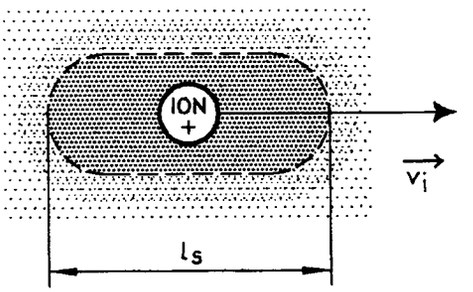


Fig. 4.2 Screening length

One can thus, empirically, understand that the ion while moving will create an electric field in its wake. The force due to this electric field will try to slow down the ion itself. The energy consumed in the electron disturbance will be lost by the ion.

4.2 Relation between cooling force and time

We have seen that the objective of electron cooling is to decrease the ion velocity spread v_i, in the moving frame. This implies a reduction of the ion energy and therefore the existence of a friction or cooling force. The cooling time will depend on the cooling force intensity.

Let us consider a single particle at velocity v_i. For an exponential decrease of the ion velocity we can write:

$$v_i(t) = v_i(t=0) \cdot \exp(-t / \tau_{v_i}) .$$

Then the cooling time variable τ_{vi} (in s) is given by:

$$\frac{1}{\tau_{v_i}} = -\frac{1}{v_i}\frac{dv_i}{dt} = -\frac{1}{m_i \cdot v_i}\left(m_i \frac{dv_i}{dt}\right) = -\frac{F(v_i)}{p_i}$$

where F is the cooling force and p_i the ion momentum.

The friction rate is therefore given by:

$$\frac{1}{\tau_{v_i}} = -\frac{F}{p_i}.$$

It shows that an understanding of the cooling (or friction) force is essential for determining the cooling time. Section 5 will be devoted to this subject where it will be shown that F depends on v_i and v_e and so $\tau_{v_i} = (v_i, v_e)$. The computation of the cooling $\Delta t_i = t[v_i = v_i \text{ final}] - t[v_i = v_i \text{ initial}]$ is thus not straightforward.

In the laboratory frame, the cooling time variable will then be (see Section 2.3):

$$\tau^*_{v_i} = \gamma_0 \frac{1}{\eta_c} \tau_{v_i}$$

where η_c is the ratio of the cooling section length to the accelerator circumference.

What is observed by an experimentalist is the average cooling time $\Delta t = <\Delta t_i>$.

5. ESTIMATION OF COOLING FORCES

Although most electron coolers make use of a longitudinal magnetic field, it is worth considering the case of non-magnetic cooling forces. Firstly, it is an elegant didactic approach to the cooling process and secondly, for large relative ion velocities $\vec{v}_i$, the magnetised forces are relatively weak with respect to the non-magnetised forces. The exact explanations are out of the scope of this lecture (consult Bibliography). All the computations take place in the moving frame.

5.1 Without magnetic field

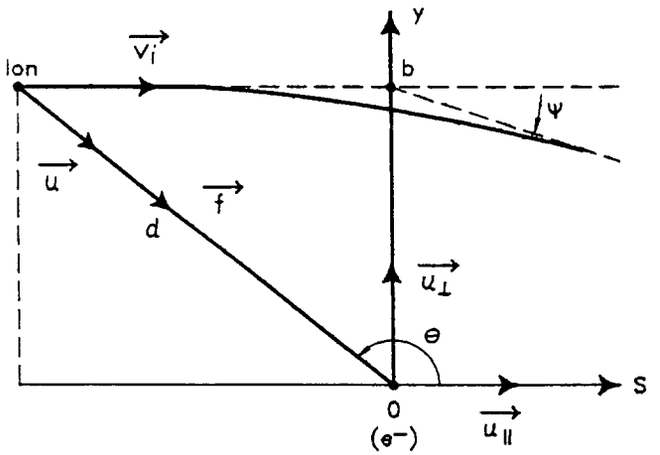


Fig. 5.1 Geometry of the collision

As a first, but illustrative approach, we use the single electron-ion interaction represented by Fig. 5.1. The ion initially at position (∞, b) with a velocity $\vec{v}_i = v_i \cdot \vec{u}_{\parallel}$ will interact with an electron at rest at (0, 0). Due to the Coulomb force exerted between the two particles the ion will be deviated by an angle which is supposed to be small. The deviation ψ is given by:

$$\cot g\left(\frac{\psi}{2}\right) = -\frac{b\, m_e (v_i)^2\, 4\pi\varepsilon_0}{Z\, e^2}\,.$$

Let us estimate the exchange in energy, due to this interaction, when the time runs from $-\infty$ to $+\infty$. The Coulomb force $\vec{f}$ will induce a change in momentum $\vec{dp}\,/\,dt$ given by:

$$\vec{f} = \frac{Z e^2}{4\pi\varepsilon_0}\cdot\frac{1}{d^2}\vec{u} = Z\,\mathbf{e}^2\frac{1}{d^2}\vec{u} = \frac{d\vec{p}}{dt} \quad ; \quad \mathbf{e}^2 = \frac{e^2}{4\pi\varepsilon_0}\,.$$

Approximately: $$\Delta\vec{p} = \int_{-\infty}^{\infty}\vec{f}\,dt \equiv Z\,\mathbf{e}^2\int_{-\infty}^{\infty}\frac{\vec{u}}{s^2+b^2}dt = \Delta p_{\parallel}\,\vec{u}_{\parallel} + \Delta p_{\perp}\vec{u}_{\perp}$$

where $$\vec{u} = \frac{\vec{f}}{|\vec{f}|} = -\cos\theta\cdot\vec{u}_{\parallel} - \sin\theta\cdot\vec{u}_{\perp} = \frac{-s}{\sqrt{s^2+b^2}}\vec{u}_{\parallel} + \frac{-b}{\sqrt{s^2+b^2}}\vec{u}_{\perp}\,.$$

For small-angle scattering the longitudinal component of the force, being the integral of an odd function, $\Delta p_{\parallel}$ is zero, while for the transverse component we get:

$$\Delta p_{\perp} = -Z\,\mathbf{e}^2 b\int_{-\infty}^{\infty}\frac{dt}{\left(s^2+b^2\right)^{3/2}}$$

and since $s = v_i\cdot t$, $$\Delta p_{\perp} = \frac{-Z\,\mathbf{e}^2 b}{b^2 v_i}\lim_{A\to\infty}\left[\frac{s}{\sqrt{s^2+b^2}}\Bigg|_{-A}^{A}\right] = \frac{-2Z\,\mathbf{e}^2}{v_i\, b}$$

where, as usual, b is called the <u>impact parameter</u>.

Due to the electrical field compression, the effective time of interaction or collision time is rather short and is of the order of:

$$\Delta t = \frac{b}{\gamma' v_i} \quad \text{(Fig. 5.2)}$$

($\gamma' \cong 1$ since we work in $\Re_0$).

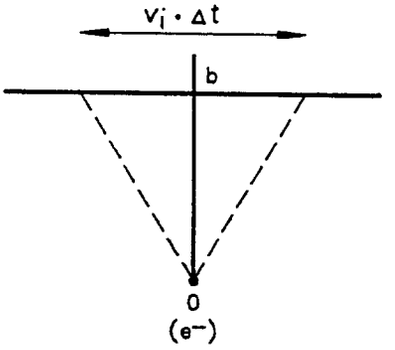


Fig. 5.2 Approximate collision length

The energy lost by the ion, which is that gained by the electron is then:

$$\Delta E(b) = \frac{(\Delta p)^2}{2m_e} = \frac{2Z^2 e^4}{m_e v_i^2 b^2}. \tag{5.1}$$

It is useful to notice the dependence on v_i.

Up to now we have considered a single collision. Of course we must take into account multiple collisions with all possible impact parameters b. If n_e is the electron beam density then the number of electrons in the volume $\pi\ b^2\ ds$ will be $n = \pi\ b^2\ n_e\ ds$ while (Fig. 5.3) $dn = 2\ \pi\ b\ n_e\ ds\ db$ is the number of electrons between b and $b + db$ over the length ds. The energy lost by the ion per unit of length is then:

$$\frac{dE}{ds} = 2\pi \int_{b\min}^{b\max} b n_e\, \Delta E(b)\, db$$

$$\frac{dE}{ds} = \frac{4\pi Z^2 e^4}{m_e v_i^2} n_e\ \ell n\left(\frac{b_{\max}}{b_{\min}}\right).$$

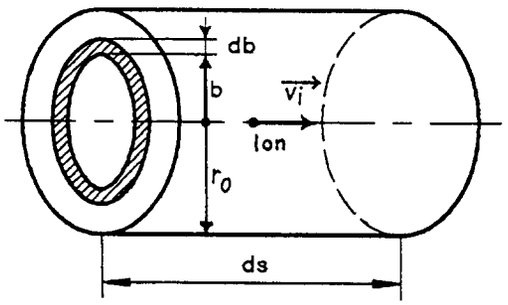


Fig. 5.3 Ion moving in the electron cloud with different impact parameters

An estimation of the Coulomb Logarithm $L_c(v_i) = \ell n\,(b_{\max} / b_{\min})$ has to be found. Concerning $b_{\min}$, it can be estimated by the maximum momentum transfer to the electron (classical head-on collision):

$$\frac{2Z\ell^2}{v_i\, b_{\min}} = \Delta p_{\max} = 2m_e v_i \rightarrow b_{\min} = \frac{Z\ell^2}{m_e v_i^2}.$$

For b_{max} we take $b_{max} = \min(\lambda_d, r_0)$. Of course $b_{max} >> b_{min}$. Usually L_c, of the order of 10, is logarithmically dependent on v_i and is taken as constant.

In fact, the electrons are not quite mono-energetic and therefore the friction force must be weighted by the electron speed distribution $f(v_e)$ which can be expressed in a Gaussian form:

$$f(v_e) = \frac{e^{-\left[\frac{v_{e\parallel}^2}{2\Delta_{e\parallel}^2} + \frac{v_{e\perp}^2}{2\Delta_{e\perp}^2}\right]}}{(2\pi)^{3/2}\,\Delta_{e\perp}^2\,\Delta_{e\parallel}} \qquad \text{and } 1 = \int f(v_e)\, d^3 v_e$$

$$\Delta_{e\parallel}^2 = \frac{k \cdot T_{e\parallel}}{m_e} \quad ; \quad \Delta_{e\perp}^2 = \frac{k \cdot T_{e\perp}}{m_e}$$

We have shown in section 3.2 that $T_{e\parallel} << T_{e\perp}$. The friction force is then:

$$\vec{F} = -\frac{4\pi Z^2 \ell^4 n_e}{m_e} L_c \int \frac{\vec{v}_i - \vec{v}_e}{|\vec{v}_i - \vec{v}_e|^3} f(v_e)\, d^3 v_e. \tag{5.2}$$

where (see Fig. 5.4):

$$\vec{v}_i = v_{i\parallel} \cdot \vec{u}_\parallel + v_{i\perp} \cdot \vec{u}_\perp$$
$$\vec{v}_e = v_{e\parallel} \cdot \vec{u}_\parallel + v_{e\perp} \cdot \vec{u}_\perp .$$

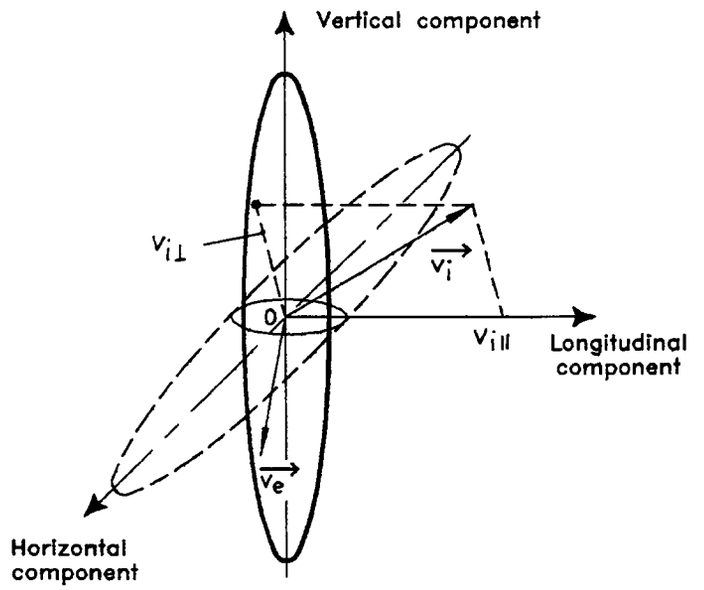


Fig. 5.4 Electron and ion relative velocities in $\Re_0$

L_c is considered to be independent of the velocity. This has to be numerically integrated.

At this level it is important to notice the analogy between the expression of the force $\vec{F}$ and that of an electrical field $\vec{E}$, at a point P due to a distribution of charge ρ (Fig. 5.5).

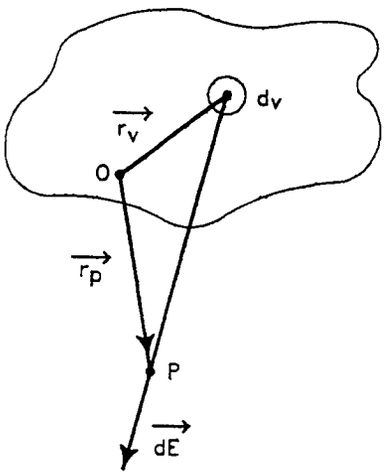


Fig. 5.5 Analogy with an electrical field

$$\vec{E} = \int \frac{\vec{r}}{|\vec{r}|^3} \rho \, d^3 r_v \,, \qquad \left(d^3 r_v \equiv d^3 v\right)$$

where $\vec{r} = \vec{r}_p - \vec{r}_v$

$$\rho = -\frac{4\pi \, n_e \cdot Z^2 \cdot \ell^4}{m_e} L_c \cdot f(\vec{r}_v)$$

$\vec{F}$ is similar to $\vec{E}$ in the velocity space.

In the real case of a Maxwell distribution of the electron speed for which $\Delta_{e\parallel} << \Delta_{e\perp}$ computations have been made [1]. The following gives the asymptotic expressions of the forces.

In the longitudinal direction ($v_{i\perp} = 0$) (Fig. 5.6)

$$F_{\parallel}(v_{i\parallel}) = -\frac{4\pi \, Z^2 \cdot \ell^4}{m_e} n_e \cdot L_c \begin{cases} \dfrac{1}{v_{i\parallel}^2} & ; \quad |v_{i\parallel}| >> \Delta_{e\perp} \qquad (5.3a) \\ \dfrac{1}{\Delta_{e\perp}^2} & ; \quad \Delta_{e\perp} >> |v_{i\parallel}| >> \Delta_{e\parallel} \qquad (5.3b) \\ \dfrac{v_{i\parallel}}{(2\pi)^{3/2} \, \Delta_{e\perp}^2 \, \Delta_{e\parallel}} & ; \quad |v_{i\parallel}| << \Delta_{e\parallel} \qquad (5.3c) \end{cases}$$

where:

$$L_c = \ell n \, (b_{\max} / b_{\min}) \quad \text{and} \quad b_{\max} = \min\left\{ \frac{< |\vec{v}_i - \vec{v}_e| >}{\omega_{p\ell}} \,, \; \tau < |\vec{v}_i - \vec{v}_e| > , r_0 \right\}$$

$$b_{\min} = \frac{Z \, \ell^2}{m_e < |\vec{v}_i - \vec{v}_e| >}$$

$\omega_{p\ell}$ the plasma pulsation, τ the ion time in the electron beam.

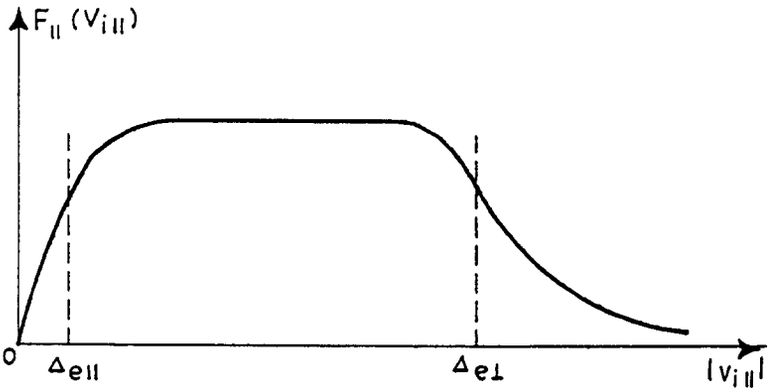


Fig. 5.6 Shape of the "non-magnetised" longitudinal cooling force

In the transverse direction ($v_{i\parallel} = 0$) (Fig. 5.7)

$$F(v_{i\perp}) = -\frac{4\pi Z^2 . e^4}{m_e} n_e \cdot L_c \begin{cases} \dfrac{1}{v_{i\perp}^2} & ; \quad |v_{i\perp}| >> \Delta_{e\perp} \qquad (5.4a) \\ \dfrac{\sqrt{\pi}}{8} \dfrac{v_{i\perp}}{\Delta_{e\perp}^3} & ; \quad |v_{i\perp}| << \Delta_{e\perp} \qquad (5.4b) \end{cases}$$

The transverse force is maximum when the ion speed is the same as the electron transversal rms speed.

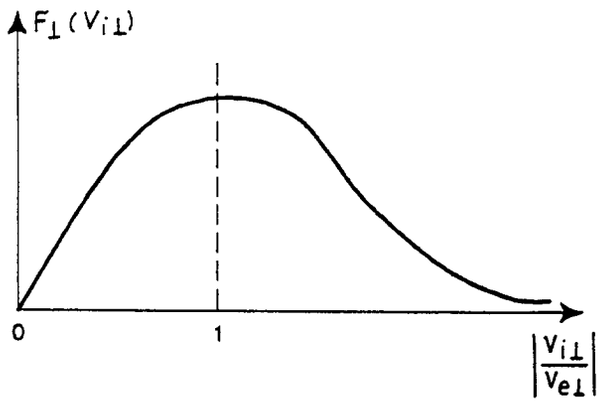


Fig. 5.7 Shape of the "non-magnetised" transverse cooling force

We observe that:

- the forces are not independent of the ion relative velocities
- for large ion velocities the forces scale as $1/(v_i{}^2)$, suggesting that a beam with a relatively large emittance will have a large cooling time
- for small velocities the forces are proportional to v_i.

Numerical example

Let us simplify and take a mono-energetic electron beam

$$f(v_e) = \delta\,(v_{e\parallel}, v_{e\perp})\,.$$

Then, using Eq. (5.2) and considering one single particle at speed v_i:

$$\vec{F} = -F_1 \frac{\vec{v}_i}{|v_i|^3} \quad ; \quad F_\parallel = \frac{-F_1\, v_{i\parallel}}{|v_i|^3} \; ; \quad F_\perp = \frac{-F_1\, v_{i\perp}}{|v_i|^3}$$

with:

$$F_1 = \frac{4\pi\, Z^2\, e^4\, n_e\, L_c}{m_e} = \frac{4\pi\, Z^2\, n_e^*\, L_c \cdot m_p \cdot r_e \cdot r_p}{\gamma_0}\, c^4 = \frac{m_p\, Z^2}{\gamma_0} \quad F_0 \quad ; \quad \mathrm{m^3 \cdot kg \cdot s^{-4}}$$

since:

$$n_e = \frac{n_e^*}{\gamma_0}\,.$$

With $v_{i\parallel} = v_{i\perp}$, the cooling time variable for that particular velocities will be (see section 4.2)

$$\tau_{v_i} = \frac{A \cdot m_p \cdot v_i\, |v_i|^3}{F_1\, v_{i\perp}} = \frac{\gamma_0\, A}{Z^2} \frac{(v_i)^4}{F_0\, v_{i\perp}}\,.$$

In the laboratory frame ($Z = 1 = A$) and taking into account the cooling length

$$\tau_{v_i}^* = \frac{\gamma_o^2}{\eta_c} \frac{(v_i)^4}{F_1\, v_{i\perp}} = \frac{\gamma_0^2}{\eta_c} \frac{|v_i|^3}{\sqrt{2}\, F_1}\,.$$

Using the numbers of our typical accelerator (section 2.2) with

$$\frac{v_i}{c} = \beta_0^* = 10^{-3} = \sqrt{2}\, \frac{v_{i\perp}}{c}$$

gives $\tau_{v_i}^* = 2.726$ s.

When expressed in practical units $\tau^* \propto \beta_0^4\, \gamma^5\, \theta_y^3 / I_b = \beta_0^4\, \gamma_0^5 \left(\varepsilon_y / \beta_{h,v}\right)^{3/2} / I_b\,.$

We see that as $\tau^* \propto \beta_0^4\, \gamma_0^5$, electron cooling for relativistic particles appears to be less efficient.

Exercise
Starting from Eq. (5.2), derive Eqs. (5.3a) to (5.3c). Use the Gauss theorem and Eq. (5.4a)

5.2 Friction force with magnetised electrons

The magnetic field limits the transverse motion of the electrons which appear to be frozen in this plane. The electrons rotate around their axis at the cyclotron frequency f_ℓ with a radius equal to r_ℓ. If the impact parameter b is much larger than r_ℓ and if the collision time $\Delta t = b / v_i$ is larger than $1 / f_\ell$ then, since the transverse motion is "frozen", the electron cooling efficiency is determined by the electron longitudinal temperature which is a few orders of magnitude

lower than the transverse one. Cooling is highly improved. This type of collision is often called adiabatic collision ($r_\ell < b$).

The case where $b_{min} < b < r_\ell$, is called a fast collision and looks like the case previously studied (without magnetisation of electrons). We can explain it in the following way (see Fig. 5.8). If the ion is motionless and the electron rotates at angular frequency ω_ℓ on a circle of radius r_ℓ the force exerted on the electron is

$$\vec{f} = \text{Cst}\ \vec{\ell} / \ell^3 \qquad (\text{Cst is a constant}).$$

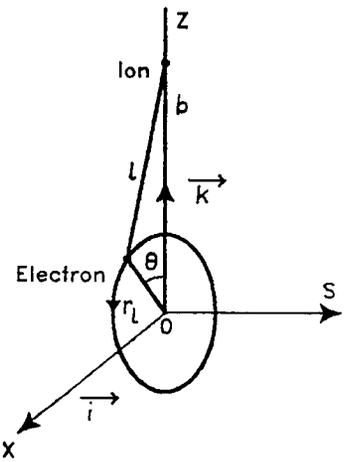


Fig. 5.8 Illustration of the Coulomb force exerted between a fixed ion and an electron moving on a circle of cyclotron radius over one full cyclotron period

$$\vec{\ell} = -r_\ell \sin\theta\, \vec{i} + (b - r_\ell \cos\theta)\vec{k}$$
$$\ell^2 = b^2 + r_\ell^2 - 2 \cdot r_\ell\, b \cos\theta$$
$$d\vec{\ell} = \left[-r_\ell \cos\theta\, \vec{i} + r_\ell \sin\theta\, \vec{k}\right] d\theta.$$

The work of the force $dW = \vec{f} \cdot d\vec{\ell}$ over one turn, which is equivalent to the exchange in energy between the ion and the electron, is zero:

$$W = \int_{\theta=0}^{\theta=2\pi} \vec{f} \cdot d\vec{\ell} = 0.$$

There is no exchange of energy in the transverse plane (in the adiabatic case).

If now the ion moves very quickly with respect to the electron then the interaction time is shorter than the cyclotron period. The electron will move on a part of the circle during this time and (Fig. 5.9):

$$W = \int_{\theta=0}^{\theta_0} \vec{f} \cdot d\vec{\ell} \neq 0.$$

We are now in the previous situation (no magnetic field) where the electron is considered as "free".

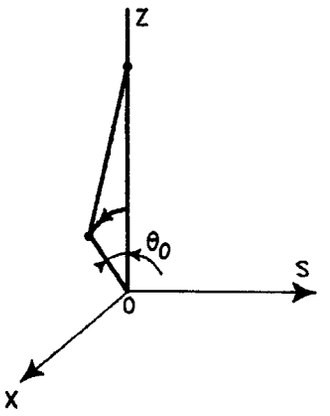


Fig. 5.9 Illustration of the Coulomb force exerted between a fixed ion and an electron moving on a circle of cyclotron radius over part of one period

We can also make a comparison with the energy transfer of an ion to a harmonically bounded electron (see Classical Electro dynamics: J.D. Jackson; chapter on collision between charged particles [Bibliography]). In this case the electron is considered to be bounded in an atom and executes an orbital trajectory at angular frequency ω_t. The collision time is again given by:

$$\Delta t = \frac{b}{\gamma' v_i}.$$

If the collision is shorter than $1/\omega_t$ the electron is assumed to be free and the previous results apply. If on the other hand the collision time is very long compared to the orbital period, the electron will make many cycles of motion as the incident particle passes slowly by and will be influenced adiabatically by the field. We can define a threshold impact parameter b_t given, since:

$$\Delta t = 1/\omega_t \,,\; b_t = \frac{\gamma' v_i}{\omega_t} = \gamma' v_i \Delta t \,.$$

The exact energy transfer to harmonically bound charge is given by:

$$\Delta E(b) = \frac{2Z^2 e^4}{m_e v_i^2} \quad \frac{1}{b^2}\left[\xi^2 K_1^2(\xi) + \frac{1}{\gamma'^2}\xi^2 K_0^2(\xi)\right] \tag{5.5}$$

with

$$\xi = \frac{\omega_t b}{\gamma' v_i}$$

and $K_{1,0}$ the modified Bessel functions. The term between square brackets behaves asymptotically as:

$$[\;] = \begin{cases} 1 & \text{for} \quad \xi = \dfrac{b}{b_t} << 1 \\ \left(1 + \dfrac{1}{\gamma'^2}\right)\dfrac{\pi}{2}\xi\, e^{-2\xi} & \text{for} \quad \xi = \dfrac{b}{b_t} >> 1 \end{cases}$$

We see that for $b << b_t$ the energy transfer is essentially the result given for a free electron, since then (5.5) equals (5.1), while for $b >> b_t$ (and therefore small ion velocity spread) it falls exponentially to zero thus showing that there is no transfer of energy.

The same type of argument can be applied to the cooling process with magnetised electrons where, for low velocity ions, there is no energy transfer with the electrons' transverse movement whereas the exchange in energy occurs mainly with the electrons' longitudinal degree of liberty. Since the electron longitudinal velocity spread is very small the magnetisation will significantly enhance the cooling force in the case of low velocity ions.

A third explanation (I. Meshkov, private communication) can be given with the help of Fig. 5.10. Here again we have to take into account only the longitudinal displacement of an electron along the magnetic field lines $\vec{B}_0$ which are crossed by the ion at angle θ.

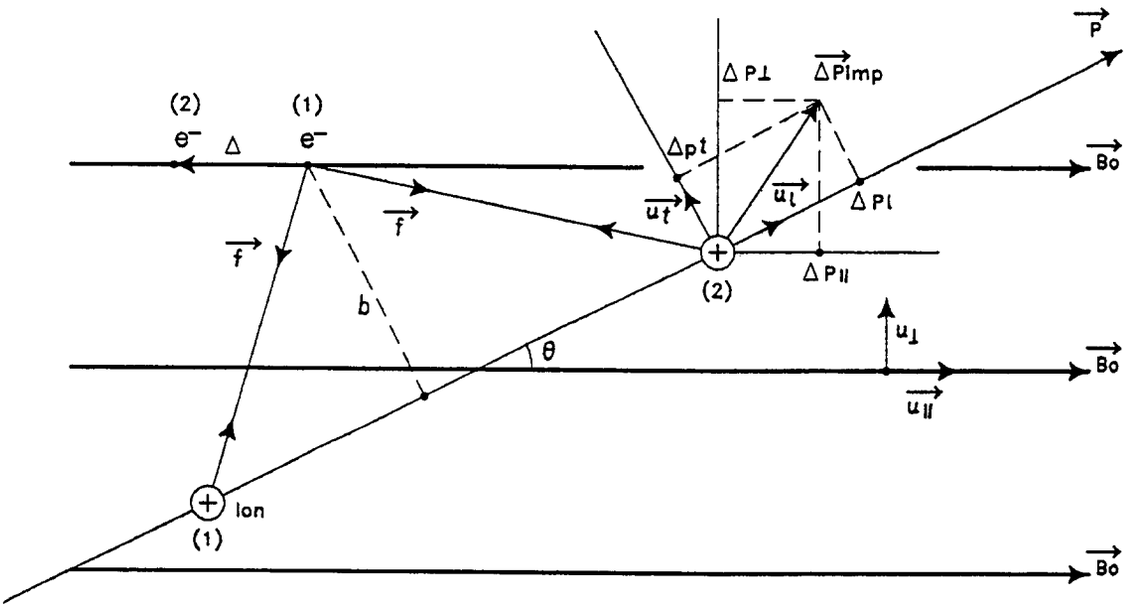


Fig. 5.10 Description of the third explanation for magnetised cooling

Starting from position (1), the ion and electron will interact. After a time t, when in position (2), the electron will have executed a small displacement $\Delta << b$ such that:

$$\Delta = \frac{\text{acceleration} \cdot t^2}{2} = \frac{f}{m_e} \sin\theta \cdot \frac{t^2}{2} \cong \frac{Ze^2}{2m_e b^2} \sin\theta \cdot \left(\frac{b}{v_i}\right)^2 = \frac{Ze^2}{2m_e v_i^2} \sin\theta .$$

This gives a small change in the impact parameter and it is easily shown that $\Delta b \cong \Delta$.

This small difference will result in a change in the drag force depending on whether the ion is on the first (1) or second (2) half of its trajectory. The momentum transfer induced by the collision is:

$$\Delta\vec{p}_{imp} = \Delta p_{\parallel}\vec{u}_{\parallel} + \Delta p_{\perp}\vec{u}_{\perp} = \Delta p_{\ell}\vec{u}_{\ell} + \Delta p_{t}\vec{u}_{t}$$

with $\Delta p_{\parallel} << \Delta p_{\perp}$ since the phenomenon is the same as for a collision of a gas molecule with a hard wall.

The result is:

$$\Delta p_{\ell} = \Delta p_{\perp} \sin\theta + \Delta p_{\parallel} \cos\theta \cong \Delta p_{\perp} \sin\theta$$

$$\Delta p_{t} = \Delta p_{\perp} \cos\theta - \Delta p_{\parallel} \sin\theta \cong \Delta p_{\perp} \cos\theta .$$

To get an estimate of $\Delta p_{\perp}$ we can write (see section 5.1):

$$\Delta p_{\perp} \cong \int_{-\infty}^{\infty} 2\left(\frac{Z\ell^2}{b^2} - \frac{Z\ell^2}{(b+\Delta b)^2}\right)\frac{ds}{v_i^2} \cong \frac{4Z\ell^2}{v_i}\frac{\Delta b}{b^2} \cong \frac{2Z^2\ell^4}{m_e v_i^3}\frac{\sin\theta}{b^2}$$

since the integration is effective over the distance b only.

We then come to the simplified form:

$$\Delta p_{\ell} = \frac{2Z^2\ell^4}{m_e v_i^3}\sin^2\theta \; ; \; \Delta p_{t} \cong \frac{2Z^2\ell^4}{m_e v_i^3}\frac{\sin\theta}{b^2}\cos\theta .$$

The friction force resulting from the electron cloud is:

$$F_{\ell,t} = -\int_{b_{min}}^{b_{max}} \Delta p_{\ell,t}\, n_e\, v_i\, 2\pi b\, db$$

such that:

$$F_{\ell} \cong -\frac{4\pi Z^2\ell^4}{m_e} n_e\, L_c\, \frac{\sin^2\theta}{v_i^2} \; ; \; F_{t} \cong -\frac{4\pi Z^2\ell^4}{m_e} n_e\, L_c\, \frac{\sin\theta\cdot\cos\theta}{v_i^2} .$$

When comparing $F_{\ell,t}$ to (5.3) or (5.4), it is important to notice that there is no influence of the electron transverse velocity spread $\Delta_{e\perp}$, keeping in mind that our explanation is valid for an ion velocity which scales to about $v_i \geq 100\ \Delta_{e\parallel}$.

In consequence, to the non-magnetic case determined before, a magnetic force F^{ad} has to be added to the previous force. Again numerical computation must be used [1]. We give the asymptotic expressions of the forces and define:

a) A relative velocity:

$$\vec{u}_{ad} = \vec{v}_i - \vec{v}_{e\parallel} = \begin{cases} v_{i\parallel} - v_{e\parallel} \\ v_{i\perp} \end{cases} .$$

b) A threshold impact parameter $b_t = v_i / \omega_{\ell}$ to distinguish between the two impact parameter regions:

- inner impact, or fast collision region $b_{min} < b < b_t$ where the formula given in (5.1) applies,

- outer impact, or adiabatic region $b_t < b < b_{max}$ where the following formula apply and for which the Coulomb logarithm:

$$L_c^{ad} = \ell n(\lambda_d / r_e) .$$

For $|v_i| >> \Delta_{e\|}$, we first set:

$$F_1 = \frac{4\pi Z^2 e^4}{m_e} n_e \cdot L_c = 4\pi Z^2 \frac{n_e^*}{\gamma_0} L_c \cdot m_p \cdot r_e \cdot r_p \cdot c^4 ; \quad \mathrm{m^3 \cdot kg \cdot s^{-4}} .$$

Then:

$$\vec{F}^{ad}(\vec{v}_i) = -\frac{1}{2} F_1 \, L_C \, v_{i\perp} \int \frac{u_{ad\perp} \cdot u_{ad\equiv}}{u_{ad}^3} \cdot \frac{\partial f(v_e)}{\partial v_{e\|}} d^3 v_e$$

where $\equiv$ means $\|$ *or* $\perp$. If $f(v_e) = \delta(v_{e\|})$ and $|v_i| >> \Delta_{e\|}$, the evaluation of the integral leads to:

$$F_{\|}^{ad}(v_i) = -\frac{1}{2} F_1 \cdot L_c^{ad}(u_{ad}) \frac{v_{i\perp}^2}{v_i^2} \cdot \frac{v_{i\|}}{v_i^3}$$

$$F_{\perp}^{ad}(v_i) = -\frac{1}{2} F_1 \cdot L_c^{ad}(u_{ad}) \frac{v_{i\perp}^2 - 2v_{i\|}^2}{v_i^2} \cdot \frac{v_{i\perp}}{v_i^3} .$$

Using a similar procedure when $|v_i| << \Delta_{e\|}$ one obtains:

$$F_{\perp}^{ad}(v_i) = -\frac{1}{\sqrt{2\pi}} F_1 \, \ell n\left(\frac{\Delta_{e\|}}{v_{i\perp}}\right) L_c^{ad}(\Delta_{e\|}) \frac{v_{i\perp}}{\Delta_{e\|}^3}$$

$$F_{\|}^{ad}(v_i) = -\frac{1}{\sqrt{2\pi}} F_1 \, L_c^{ad}(v_{i\perp}) \frac{v_{i\|}}{\Delta_{e\|}^3} .$$

The longitudinal force component for both regions ($v_i << \Delta_{e\|}$ *and* $v_i >> \Delta_{e\|}$) can be approximately expressed by:

$$F_{\|}^{ad} \cong -9 \cdot F_1 \cdot L_c^{ad}(u_{ad}) \frac{v_{i\perp}^2 \, v_{i\|}}{6|v_i|^5 + \Delta_{e\|}^3 \, v_{i\perp}^2} .$$

As mentioned above the magnetic effect appears mainly when the relative ion velocity v_i is small. This is of course doubly true, since, the transverse rms speed of the magnetised electron, with respect to the ion, is rather small. On the other hand, once the cooling process has started, and therefore the relative ion speed is lowered, the cooling effect is enhanced more and more resulting in much shorter cooling times.

If we analyse the practical case where $v_i >> \Delta_{e\|}$ we can deduce from the formula that:

- The individual forces $F_{\perp}^{ad}$ *or* $F_{\|}^{ad}$ depend both on the longitudinal and transverse components of the ion velocity. This can provoke a coupling of the motion in the different planes,

- The transverse force cancels when $v_{i\perp} = \sqrt{2} \cdot v_{i\|}$ corresponding to an angle $\phi = 54.7^0$. For larger angles the ion beam is transversally heated,

- Even when $v_i < 10\,\Delta_{e\parallel}$, the forces go approximately as $1/v_i^2$ which is quite an improvement when compared to the non-magnetic case.

Figure 5.11 illustrates the shape of the two transverse cooling forces (remember that $\Delta_{e\parallel} << \Delta_{e\perp}$) where F^0 refers to the non-magnetic case.

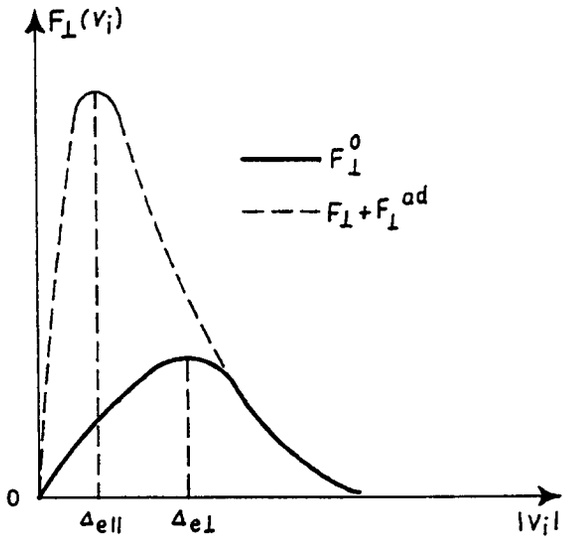


Fig. 5.11 Shape of the transverse cooling force with and without magnetisation

Equilibrium distribution of the ion beam

Theoretical equilibrium is reached when the ion temperature equals that of the electrons:

$$T_i = T_e \quad or \quad v_i = v_e\left(\frac{m_e}{m_i}\right)^{1/2}$$

which is quite favourable for the ion due to the small mass ratio (m_e / m_i). However, several other diffusion processes mentioned before such as interaction with residual gas molecules, intra-beam scattering etc., lead to much larger values. Typically one can say that transverse emittances of the order of 1π mm · mrad and $(\Delta p / p_0) \cong 10^{-5}$ are currently obtained in cooling times of the order of a second. However, the longitudinal cooling time is about five times shorter than the transverse cooling time.

It is not very useful to analyse how these limit values are distributed in the longitudinal and transverse planes. For that purpose, it is left as an exercise to compute, for a very cooled proton beam, the longitudinal proton temperature when $\Delta p / p_0 = 10^{-5}$ and the transverse proton temperature when $\varepsilon_y = 1\pi$mm · mrad; and then to compare these values with that of the electron beam when $kT_{e\parallel} = 10^{-6}$ eV and $kT_{e\perp} = 0.5$ eV.

It is easily understood that the electron beam must be well aligned with the axis defined by the theoretical closed ion orbit. Any inclination with respect to this axis will induce a "transverse electron velocity" as seen by the ions and therefore come to larger final emittances.

6. DIAGNOSTICS

Many parameters have to be measured amongst which are:

- the longitudinal and transverse emittances. The rate at which the emittances are reduced gives the cooling time. Emittances can be determined in an indirect way by the observation of the so-called Schottky signals [7] and in a more straightforward way with scrapers and ionisation beam profile monitors.

- the beam position. Since the ion and electron beams must be perfectly aligned it is important to have an accurate measurement of their position in the drift space.

- the cooling times and forces.

6.1 Longitudinal and transverse emittance measurement

6.1.1.Schottky signals

One of the easiest ways to measure the cooling effects on the emittances, in a relative manner at least, is to make use of the so-called Schottky pickups. These are mainly wide-band electromagnetic detectors which measure the statistical properties of the ion beam. We will give a brief explanation of the principle applied to unbunched beams.

a) Longitudinal Schottky signal

For a single ion circulating in a circular machine (charge $Z \cdot e$, revolution period $T_p = 1/f_p$; f_p is near f_0 but not necessarily equal to f_0) the beam current seen by a sum pickup (Fig. 6.1a) at a given location in the ring is composed of an infinite train of Dirac pulses separated by time T_p (Fig. 6.1b):

$$I(t) = Z \cdot e \cdot f_p \sum_{n=-\infty}^{\infty} \delta\,(2 \cdot \pi \cdot f_p \cdot t + \theta_i - 2 \cdot \pi \cdot n)$$

where θ_i is referred to time $t = 0$.

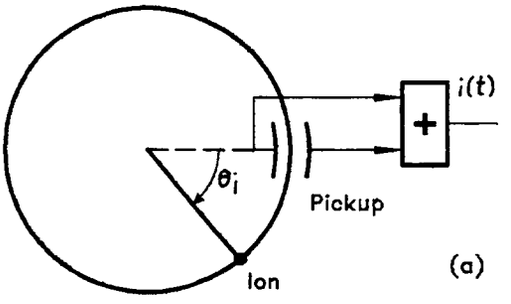


Fig. 6.1a) Longitudinal Schottky signal: explanation of symbols

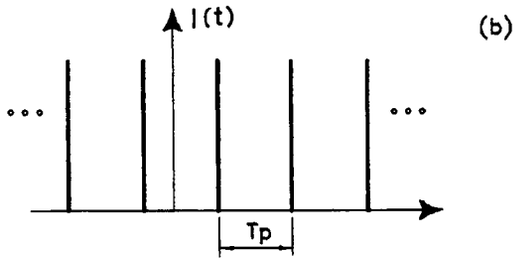


Fig. 6.1b) Time domain Dirac pulses

In the frequency domain:

$$I(t) = Z \cdot e \cdot f_p \sum_{k=-\infty}^{\infty} \exp\ j \cdot k \cdot (\theta_i + \omega_p \cdot t).$$

Looking at the positive frequencies only (Fig. 6.1c)

$$I(t) = Z \cdot e \cdot f_p \left[1 + 2 \sum_{n=1}^{\infty} \cos\left\{ n \cdot (2 \cdot \pi f_p \cdot t + \theta_i) \right\} \right].$$

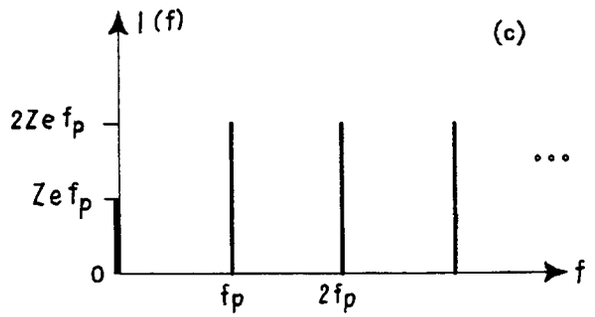


Fig. 6.1c) Positive frequency domain

For N_p particles moving with the same frequency f_p, when looking at the nth harmonic, the average will be 0 except for a DC term if the θ_i ($i = 1,...N_p$) are randomly distributed. The rms current however

$$< I^2 > = < \left[2 \cdot Z \cdot e \cdot f_p (\cos n\theta_1 + \cos n\theta_2 + ... + \cos n\theta_{N_p}) \right]^2 >$$

does not vanish because $< \cos\theta^2 > = 1/2$. We thus obtain:

$$I\,\mathrm{rms} = \sqrt{< I^2 >} = 2Z \cdot e \cdot f_p \sqrt{\frac{N_p}{2}}.$$

In fact particles have slightly different frequencies resulting from the relative momentum spread:

$$\Delta f = n \Delta f_p = n \cdot f_0\, \eta\, \frac{\Delta p}{p_0};\ \eta \text{ the off-momentum function}$$

$$\eta = (\frac{1}{\gamma^2} - \frac{1}{\gamma_{tr}^2}).$$

Displaying the power spectrum density $<I^2>/\Delta f$ on a spectrum analyser, at harmonic n, gives a measurement of $\Delta p/p_0$. During longitudinal cooling, the decrease in Δf can be observed to deduce the ratio $\Delta p/p_0$ and therefore the longitudinal emittance. This is represented in Fig. 6.2a) and in practice by Fig. 6.2b).

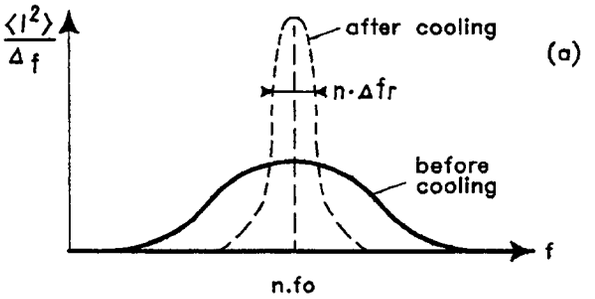


Fig. 6.2a) Evolution of the power spectrum during cooling

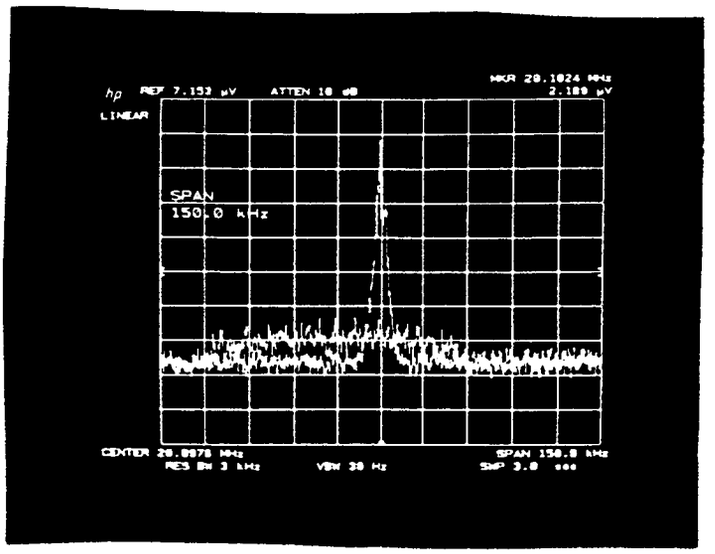


Fig. 6.2b) Display of the power spectrum for 10^9 0^{8+}

In the case of strong cooling the phases θ_i are no longer random, the *Schottky* signals are strongly enhanced and their shape is modified (see Refs. [8, 11]). Then a different interpretation is necessary to deduce $\Delta p/p_0$ and the emittances. The same applies to the next paragraph b) on transverse Schottky signals.

b) Transverse Schottky signal

For a single particle the beam current I(t) must be replaced by the dipole moment (Fig. 6.3a):

$$d(t) = a(t) \cdot \mathrm{I}(t)$$

where $a\,(t\,)$ is the transverse displacement. This displacement can be measured by taking the difference of the signals measured on the two plates of the pickup. The *pth* particle executes a sinusoidal betatron oscillation of amplitude a_p which can be written:

$$a(t) = a_p \cos(q_p \cdot \omega_p \cdot t + \varphi_i)\,.$$

Here, $q_p \cdot f_p$ is the observed frequency at a fixed location in the ring, q_p being the non-integer part of the betatron tune.

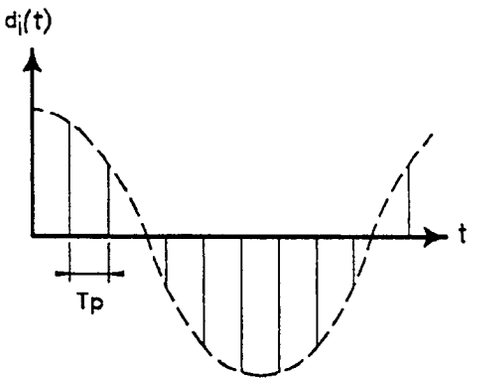


Fig. 6.3a) Transverse oscillation versus time of a single particle

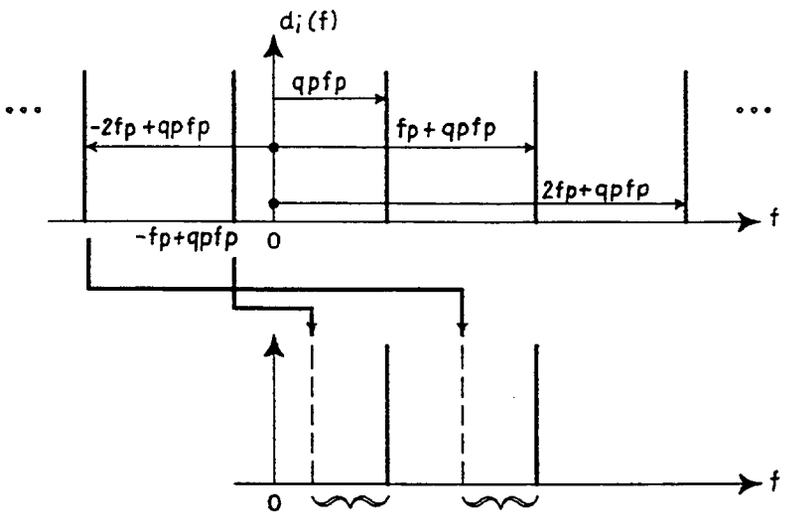


Fig. 6.3b) Frequency domain representation of the single-particle transverse oscillation

In the frequency domain (Fig. 6.3b)

$$d(t) = \underbrace{a_p \cos(q_p \omega_p t + \varphi_i)}_{a(t)}\ \underbrace{Ze\, f_p \sum_{n=-\infty}^{\infty} \exp(j n \omega_p t)}_{I(t)} = a_p Ze\, f_p \Re e\left\{ \sum_{n=-\infty}^{\infty} \exp j\left[(n+q_p)\omega_p t + \varphi_i\right]\right\}.$$

The spectrum is again a series of lines spaced by the revolution frequency of the particle but shifted by $q_p \cdot f_p$. Looking at positive frequencies only (Fig. 6.3b) one obtains two betatron lines per revolution frequency band.

For N_p particles in the beam, at the same f_p and q_p, but randomly distributed in azimuth and in betatron phase, averaging gives $<d> = 0$ but:

$$< d^2 >=< a_p^2 > (Z \cdot e)^2 \cdot f_p^2 \frac{N_p}{2}$$

$$d_{rms} = Z \cdot e \cdot f_p \cdot a_{rms} \sqrt{\frac{N_p}{2}}$$

where a_{rms} is the rms oscillation amplitude, which is closely related to the emittance.

Each Schottky band has a finite width which results from the spread of the revolution frequency:

$$\frac{\Delta f}{f_0} = \eta \frac{\Delta p}{p_0}$$

but now, to first order, we must also consider the spread of the betatron frequency related to the chromaticity ξ since $\Delta q = Q_0 \, \xi \, \Delta p / p_0$. Consequently

$$\Delta f = f_0 \frac{\Delta p}{p_0} \left[(n \pm q) \eta \pm Q_0 \, \xi \right].$$

A difference Schottky pickup (Fig. 6.4a) will display the two side bands (Fig. 6.4b). As the beam is cooled, the amplitude observed on the spectrum analyser firstly increases (curve 2) since the longitudinal cooling is faster (Δp decreases) and secondly the amplitude is reduced (curve 3) owing to the transverse cooling. The real case is illustrated by Fig. 6.4c).

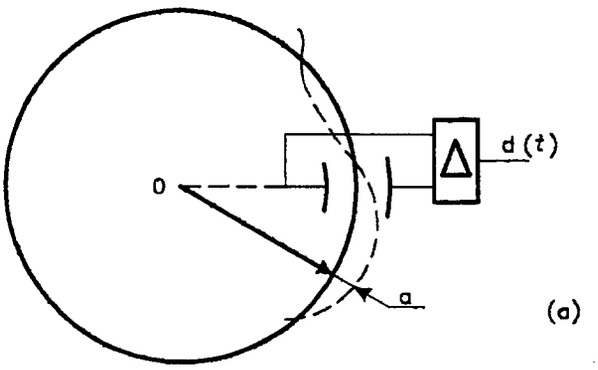


Fig. 6.4a) Implementation of a transverse Schottky pickup

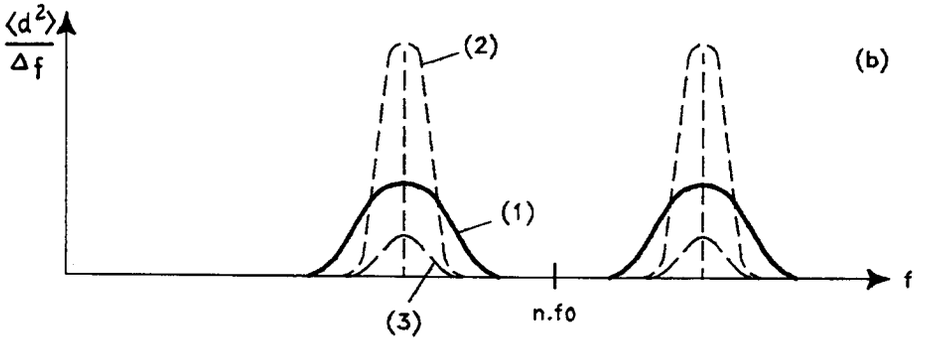


Fig. 6.4b) Evolution in time of the transverse Schottky power spectrum

Center frequenccy: 119.2 MHz; Span: 1.2 MHz

Fig. 6.4.c) Practical measurement on a spectrum analyser with $5 \cdot 10^9$ $\bar{p}$

6.1.2. The scraper (Fig. 6.5)

Suppose that the transverse ion-beam distribution is Gaussian (though the procedure applies to any distribution) of variance σ and that we have a means to measure the beam intensity. We introduce a metallic block progressively into the vacuum chamber until the beam intensity is reduced by 2.5%. Then a second block opposite to the first one is introduced until another 2.5% of the beam intensity is removed. The distance d between the two blocks is equal to 4σ or more precisely 95% of the initial particles are inside 4σ. Since $\sigma_{h,v} = \sqrt{\varepsilon_{h,v}\, \beta_{h,v}(s) / \pi}$, knowledge of $\beta_{h,v}$ allows us to determine the horizontal ε_h or vertical ε_v emittances depending on the movement of the blocks.

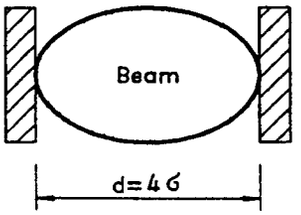


Fig. 6.5 Principle of the scraper method

6.1.3 Ionisation beam profile monitor (Fig. 6.6)

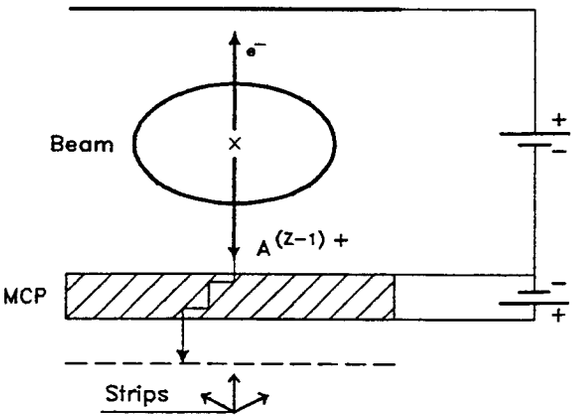


Fig. 6.6 Principle of the ionisation beam profile monitor

The ion beam when passing through the residual molecular gas, or an artificial molecular gas curtain, will ionise the molecules. The positive ions are then accelerated toward a micro-channel plate (MCP). Each ion will produce 10^3 to 10^5 electrons at the MCP output. These secondary electrons are collected on metallic strips or on a phosphor screen. The beam profile (H or V) is thus obtained. If the profile can be measured in a few ms the cooling rate can also be estimated.

6.2 Beam position monitors

Beam position monitors are usually of the electrostatic type. Since this type of detector cannot measure unbunched beams:

- the ions must be bunched by the RF cavities. The bunch duration is of the order of a micro second.

- the electron beam is modulated in density. Its current intensity is related to the acceleration voltage by $I = p_g U_0^{3/2}$ (see section 7.1); one can therefore modulate the high voltage power supply at frequencies of the order of a few kHz.

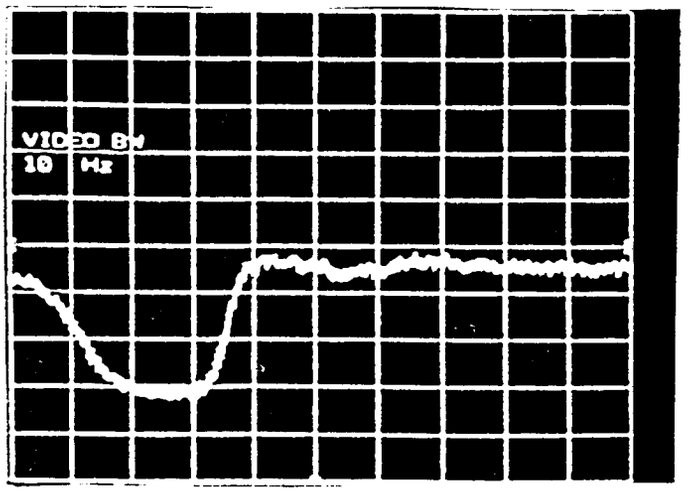
In consequence the PU electronic systems must have a large bandwidth (1 kHz to several MHz). Synchronous demodulation is sometimes used and the absolute accuracy is about 0.1 mm.

6.3 Measurement of the longitudinal cooling time and force

6.3.1 Longitudinal cooling time

Two methods have been used though both are quite difficult and marred by significant errors.

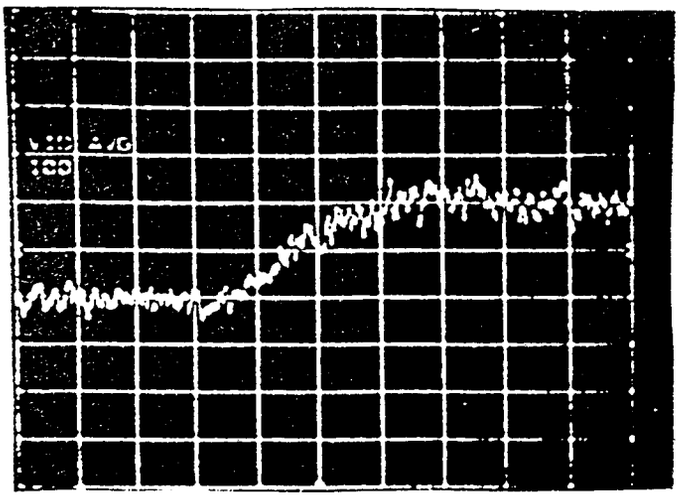
The first method consists in stepping away the electron beam energy via the high voltage power supply and then resetting the supply to the operational value. By observing on a spectrum analyser which is triggered at the same time as the voltage step, the longitudinal spectral Schottky density in a narrow bandwidth around a harmonic of the revolution frequency, the cooling time can be determined. Figure 6.7 illustrates this method.



Center frequency: 24.236 MHz; Span: 0 Hz; Sweep time: 10 s

Fig. 6.7 Longitudinal cooling time measurement using first method. The curve gives the noise density (which is proportional to the square root of the particle density) at the nominal momentum. Horizontal scale is 1 s/div. At $t = 0$ the cooling is *stepped away* and the beam density decreases due to the various diffusion mechanisms . At $t \cong 3$ s, the cooling is reset and the beam recooled to equilibrium in about 0.7 s.

The second method uses the spectrum analyser in the same way, but instead of stepping away from the operational voltage, radio-frequency noise at a different harmonic of the revolution frequency is put on a longitudinal gap with bandwidth and power adjusted to blow up the beam's momentum spread by about 10^{-3}. When this noise is switched off the spectrum analyser is triggered and the spectral density evolution is observed. An example of the signal observed is shown in Fig. 6.8.



Center frequency: 24.236 MHz; Span: 0 Hz; Sweep time: 1 s

Fig. 6.8 Longitudinal cooling time measurement using second method. The curve shows the increase in noise density as the heating noise is switched off. Horizontal scale is 0.1 s/div. At $t \equiv 0.3$ s, the heating is switched off and the beam reaches a new equilibrium density after about 0.4 s.

6.3.2 Longitudinal friction force

Two methods can be used to measure this force:

a) First method

The frictional force at low relative velocities is determined by analysing the distribution in equilibrium between a constant stochastic heating power, as mentioned in section 6.3.1 and the cooling force itself (Fig. 6.9). In order to obtain the velocity dependence of the frictional force $F(v)$ from the equilibrium distribution $\rho\,(v)$, one has to solve the one-dimensional Fokker-Planck equation for a diffusion constant $D(v)$:

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial v}\left(-F(v)\rho\,(v) + D(v)\frac{\partial \rho}{\partial v} \right).$$

In the equilibrium case $\partial\rho/\partial t = 0$, the shape of the frictional force is determined by the normalised slope of the distribution function:

$$F(v) = D(v)\frac{\partial \rho/\partial v}{\rho(v)}.$$

The diffusion constant D is derived experimentally from an independent measurement. Figure 6.10 shows the results of two measurements made with 50 MeV $\bar{p}$ (on our standard machine), one with aligned beams and the other with misaligned beams (angle of about 1 mrad).

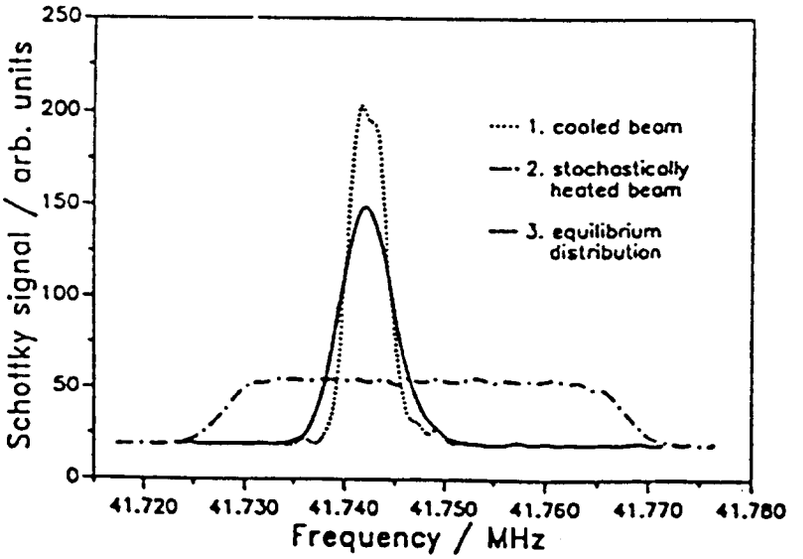


Fig. 6.9 Principle of the longitudinal frictional force measurement

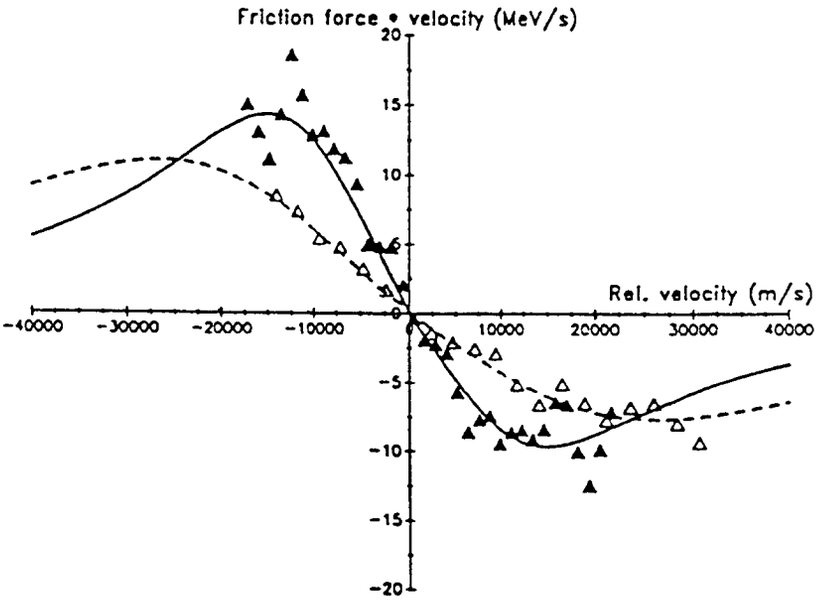


Fig. 6.10 The longitudinal frictional force $F(v)$ as a function of v for 50 MeV $\bar{p}$. The filled triangles correspond to the measurement made with aligned beams, the unfilled triangles to misaligned beams

b) Second method [9]

This uses the principle of the Betatron. A large coil is powered by a variable current and since

$$r\vec{o}t\,\vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

a longitudinal electric field is induced which results in a force on the ions. Equilibrium occurs when the electric force is equal to that of the electron cooling drag force.

7. COMPONENTS OF THE ELECTRON COOLER

All the components of an e-cooler must be designed with care. The cathode together with its gun has to provide a quasi-monokinetic electron beam whereas the collector efficiency must be almost equal to unity. On the other hand, the guiding solenoidal magnetic field must be free of any error since it may influence the electron rms transverse temperature. Firstly, we will consider the electrical arrangement of Fig. 7.1 which concerns only the e-beam.

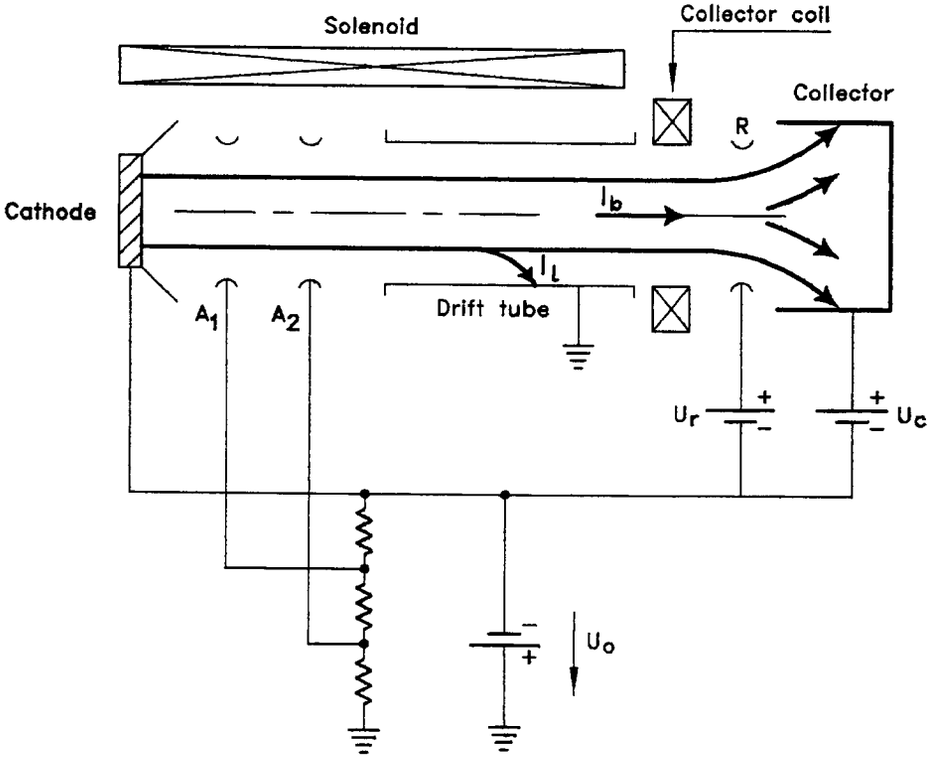


Fig. 7.1 Electron beam electrical arrangement

The electron beam, having a diameter of a few cm, emitted from a cathode is accelerated by a set of anodes (A1, A2, ...) forming the gun. When the beam enters the drift tube it has an energy of about $e.U_o$ eV. At the output of the drift tube the e-beam I_b is decelerated by an electrode at a potential U_r with respect to the cathode. U_r is of the order of 1.5 kV, relative to the cathode so that, in a very simplified approach, the e-beam energy is about $e.U_r$. The beam enters the collector at potential $U_c > U_r$ (U_c is also relative to the cathode) and therefore it reaches the collector wall at energy $e.U_c$. The dissipated power on the collector is thus $P_c = I_b.\ U_c$ (5 to 10 kW) necessitating a cooling of the collector walls (usually by water).

Except for the collector, the cooler is surrounded by a solenoid giving a longitudinal magnetic field. This magnetic field ends near the collector entrance since one wants the e-beam to diverge in order to be spread over a large area of the collector. When optimising the trajectories, the presence of at least one coil at the collector entrance is of some help. Some of the electrons may be lost. They are represented on Fig. 7.1 by I_ℓ .

7.1 Electron gun

7.1.1 Cathode

The current density J emitted by a cathode at temperature T_c is given by the Richardson-Dushmann equation:

$$J = \text{Constant} \cdot T_c^2 \cdot e^{-\frac{W_\phi}{k \cdot T_c}}$$

where W_ϕ is the extraction work. More interesting in our case is the energetic distribution of the emerging electrons. It follows a Maxwell-Boltzmann distribution. The mean energy of the electrons is equal to 2 kT_c. The distribution is, however, not isotropic. The energy carried by the longitudinal component (orthogonal to the cathode surface) is kT_c the velocity component in the two other orthogonal directions has a mean energy 0.5 k·T_c. Therefore T_C = 1000 K and kT_C = 0.0862 eV, which is equal to the longitudinal temperature $T_{\parallel}$ and transverse temperature $T_{\perp} = T_x + T_z$.

One sees the importance of having a thermoemissive cathode which operates at low temperature while having a long life time. The use of cold cathodes is of course interesting.

7.1.2 Gun

The gun itself consist of a Pierce electrode surrounding the cathode and several (two or more) accelerating electrodes all operating in the space-charge regime which means that a virtual cathode is formed near the cathode. If $-\Phi m$ is the potential-minimum plane, only those electrons with velocity normal to the cathode greater than $(2q \cdot \Phi m / m_e)^{1/2}$ pass through the potential and proceed to the anodes. Electrons which are emitted with a velocity normal to the cathode less than this value return to the cathode.

The current I_b at the gun output is given by:

$$I_b = p_g \cdot U_0^{3/2}$$

where p_g is the so-called gun perveance, of the order of 10^{-6} A.V$^{-3/2}$, and U_0 the accelerating voltage.

The overall device (cathode + gun) is usually embedded in a longitudinal magnetic field. The goal of the gun is to provide an essentially monokinetic cylindrical electron beam having a diameter of a few cm. Since one has to take into account the space charge effects, these are usually simulated by computer. The accelerating electrodes will induce a transverse electric field and even with careful adjustment of the magnetic field (resonant optics) the transverse energy dispersion may rise to 0.5 eV or indeed 1 eV. That is an important deviation from the ideal case mentioned previously and needed in order to produce efficient cooling. Presently studies are underway which aim to get very "cold" electron guns and cathodes.

Two types of gun are mainly used: the resonant or the adiabatic type.

a) Resonant gun in a simplified form (Fig. 7.2)

Submitted to a longitudinal magnetic field B_0, the electron will describe in the transverse plane a circle C_1 of radius $r_t = p_{e\perp} / |e| B_0$ and move longitudinally by $\lambda_t = 2\pi p_{e\parallel} / |e| B_0$ each cyclotron period. When the electron passes at time t_1 the first anode A_1 at potential V_1 then, due to the radial electrical field E_{r1}, it will be submitted to a transverse kick and follow, transversally, the path C_2. At time t_2 when it passes the anode A_2 at potential V_2 it will receive a second kick due to the transverse electrical field E_{r2}. If an appropriate relation or resonance is founded between $(V_2\text{-}V_1)$, $(t_2\text{-}t_1)$ and B_0 the electron may come, at the output of A_2, to its initial trajectory C_1. In this case the initial transverse energy remains unchanged.

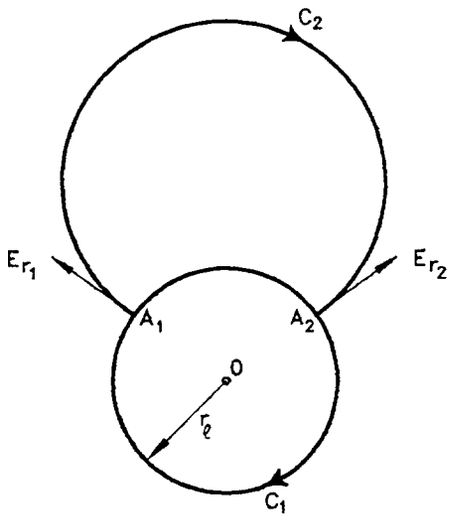


Fig. 7.2 Principle of the resonant gun

b) Adiabatic gun

If the transverse electric field varies slowly so that:

$$\frac{dE_r}{ds} \leq \frac{E_r}{\lambda_\ell}$$

the condition of adiabatic motion is satisfied which means that the perturbations due to transverse fields are negligible. This, usually, needs higher magnetic fields (small λ_ℓ) than for a resonant gun.

7.2 Drift space

7.2.1 Transversal and longitudinal temperature spread

During acceleration the transverse energy spread should theoretically remain constant, of the order of 0.2 to 0.5 eV as mentioned before. The longitudinal velocity spread, with respect to the moving frame is subject to an important reduction as seen in section 3. The electron velocity distribution is thus flattened.

7.2.2 Effect of the space charge on the longitudinal velocity

The electron beam is assumed to be circular with a radius r_0, and to travel along the axis of a circular tube of radius r_c, at ground potential. The electron beam density is then (Fig. 7.3):

$$n_e^* = \frac{p\, U_o^{3/2}}{e \cdot \pi\, r_0^2 \cdot \beta_0 \cdot c} \ \mathrm{m}^{-3}.$$

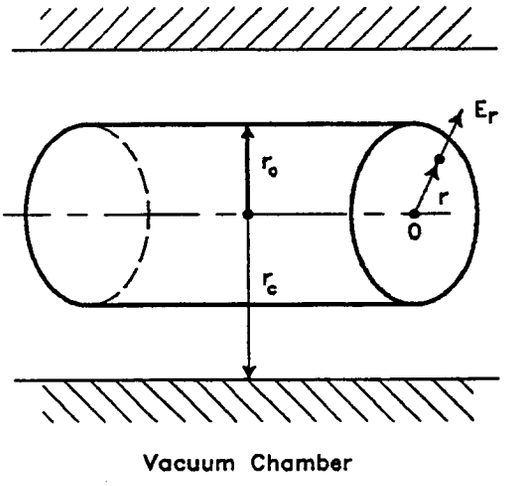


Fig. 7.3 Geometry of the drift tube

Due to the electron space charge a transverse radial electrical field E_r will exist within the e-beam. It can be computed using the Gauss theorem. The expressions of the electric field and potential are:

For $0 \le r \le r_0$

$$U_f = \frac{en_e^*(r^2 - r_0^2)}{4\varepsilon_0} - \frac{en_e^* r_0^2}{2\varepsilon_0} \ell n\left(\frac{r_c}{r_0}\right)$$

$$E_r = -\frac{en_e^*}{2\varepsilon_0} r$$

For $r_0 \le r \le r_c$

$$U_f = -\frac{en_e^*}{2\varepsilon_0} r_0^2 \ell n(r_c / r)$$

$$E_r = -\frac{en_e^*}{2\varepsilon_0} \frac{r_0^2}{r}.$$

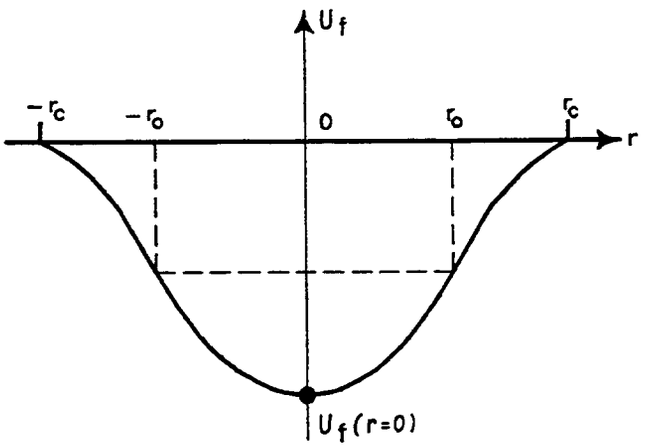


Fig. 7.4 Potential distribution versus radius in the drift tube

This means that even at the centre of the beam (r = 0) there exists a negative potential (Fig. 7.4):

$$U_f = -\frac{e n_e^* r_0^2}{4\varepsilon_0}\left[1+2\ell n\left(\frac{r_c}{r_0}\right)\right]$$

$$U_f = -n_e^* \pi r_0^2 r_e\left(\frac{m_e c^2}{e}\right)\left[1+2\ell n\left(\frac{r_c}{r_0}\right)\right]$$

so that the kinetic energy of the electrons at the centre of the beam is:

$$Ec = e\,U_0 - n_e^* \pi\, r_0^2 r_e\left(\frac{m_e c^2}{e}\right)\left[1+2\ell n(r_c / r_0)\right]$$

$$Ec = e\,U_0\left[1-\frac{p_g U_0^{1/2}}{\beta_0}\cdot 91.849\right].$$

Example:
At 309 MeV/c, U_0 = 27·10 kV, β_0 = 0.312, p_g = 0.52·10^{-6}, r_c = 70 mm gives U_f (r = 0) = - 676 V. Therefore the theoretical accelerating voltage must be corrected by this amount in order to obtain the requested velocity v_0.

More important is the fact that electrons having a radius $0 < r \leq r_0$ will undergo different acceleration voltages and will not move at the same longitudinal velocity as the electron on axis. The electron beam is far from being monokinetic. The longitudinal rms velocity spread remains however unchanged.

7.2.3 Effect of the space charge on the transverse velocity

The Ampere theorem can be used to get the magnetic field due to the electron beam itself at radius $0 < r < r_0$:

$$\vec{B} = \frac{\mu_0}{2}(n_e^*\, e\, \beta_0\, c) r\, \vec{u}_\theta\,.$$

The total force which acts on an electron of the beam with no initial transverse velocity, due to the space charge effect of the beam alone will be:

$$\vec{f} = -e\cdot\left[\vec{E}(r)\cdot\vec{u}_r + \vec{v}\times\vec{B}\right] = \left[\frac{e^2 n_e^* r}{2\varepsilon_0} - \frac{\mu_0}{2} e^2 n_e^* r \beta_0^2 c^2\right]\vec{u}_r = \frac{e^2 n_e^* r}{2\varepsilon_0}(1-\beta_0^2)\vec{u}_r\,.$$

The ratio of the electrical force to the magnetic force is $1/\beta_0^2$. Hence the electron would acquire a radial velocity and therefore increase its transverse temperature.

The effects of the space charge explain the necessity for an additional guiding field, given by the main solenoid and toroids $\vec{B}_0 = B_0\,\vec{k}$ so that at least $\vec{f}$ will cancel (Brillouin flow). The guiding field will exert a force $\vec{f}_b = -e\cdot r\cdot B_0\cdot\dot{\theta}\cdot\vec{u}_r$. The movement will become rather complex but we see that, even with no transverse energy at the cathode output, an angular velocity $v = r\cdot\dot{\theta}$ will exist which introduces an additional transverse temperature (Numerical example: 0.1 eV). In order to reduce these space-charge effects one may trap the ions, resulting from the collisions between the e-beam and the residual gas molecules in the drift region, so as to neutralise the electron space charge and therefore cancel the electric field. In this respect it is clear that the uncontrolled occurrence of such ions will modify the above expression.

7.2.4 Effect of the space charge on the cooling process

We consider,

a) The parabolic potential shape mentioned before which gives a parabolic electron longitudinal velocity

b) The longitudinal cooling force component on ions

c) The influence of the dispersion function $D(s)$ which gives the ion horizontal deviation r_i from the nominal orbit:

$$r_i = D(s) \cdot \frac{\Delta p_i}{p_0} = D(s) \cdot \gamma^2 \cdot \frac{\Delta \beta}{\beta_0} = \frac{D(s)}{\eta} \cdot \frac{\Delta f}{f_0} .$$

Therefore the ion horizontal position is:

$$x_i(s) = d_i(s) + r_i(s) .$$

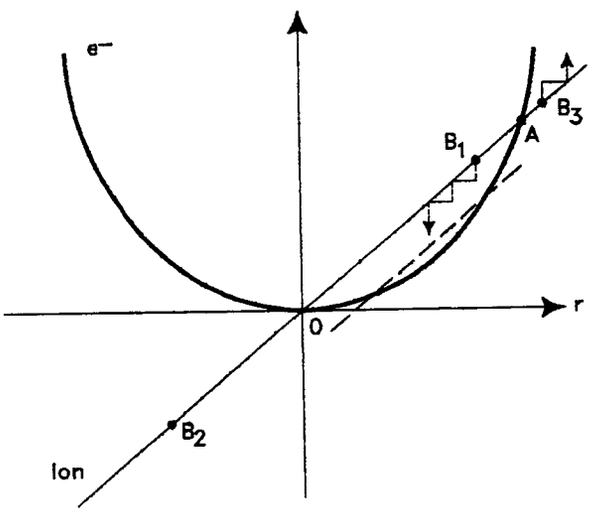


Fig. 7.5 Ion and electron velocity distribution, versus radius, in the drift space

We have plotted on Fig. 7.5 the velocity distributions as a function of the radius. If we neglect the betatron oscillation d_i we see that:

- an ion at B_1 will be submitted to a strong cooling force and will therefore rapidly converge to the centre 0,
- an ion initially at B_2 will also converge to 0, but more slowly since the cooling force (proportional to $1/|\vec{v}_i - \vec{v}_e|^2$) is rather weak,
- an ion at B_3 will diverge and be lost.

At this point, it is important to see that the two beams have to be well aligned and the average velocities of the two beams well equalised. In the case of a misalignment (dashed curve of Fig. 7.5) a large part of the ion beam may be lost.

In the vertical plane such a process does not occur.

7.3 The collector

The task of this device is to collect the incoming electrons with a maximum efficiency. Here again many types of collector have been proposed. The simplest one consists of a Faraday cup. The electrons impinging on the collector surface will:

- dissipate power on the collector walls which have to be cooled
- create secondary electrons.

The primary electrons which do not enter the collector, or the secondary electrons which escape from the collector volume, are reaccelerated toward the drift tube and the cathode. Nearby the cathode they are reflected at the virtual cathode level. They bounce back and forth in the cooler, are heated transversally, and eventually hit the vacuum wall. This current loss I_ℓ imposes a load for the high voltage power supply and upon impact with the vacuum chamber surface the electrons liberate gas and deteriorate the vacuum. The ratio:

$$\frac{I_\ell}{I_{e-beam}} = \frac{I_\ell}{I_b}$$

determines the collector inefficiency. Good collectors have an inefficiency of the order of 10^{-4}.

In order to reduce the dissipated power and the maximum energy of the secondary electrons it is essential to reduce U_c. According to the negative space-charge potential due to the e-beam, U_c must be of the order of a few kV. An important characteristic of the collector is its perveance $p_c = I_c/U_c^{3/2}$. It must be such that:

$$I_b = p_g \cdot U_0^{3/2} < p_c \cdot U_c^{3/2}$$

p_g being the gun perveance. Therefore in order to have small dissipation (U_c small) p_c must be large.

Computer codes have to be used to calculate the correct dimensions, potentials and magnetic field. However the influence of the secondary electrons and of the ionised ions is very difficult to simulate.

7.4 Magnetic field

We have seen that the presence of a longitudinal field enhances the cooling forces and therefore reduces the cooling time (the field is of the order of $3 \cdot 10^{-2}$ to $1.5 \cdot 10^{-1}$ Tesla). We will see (section 8) that the vertical component of the magnetic field in the toroids creates a horizontal kick where the ion beam enters and leaves the toroids. On the other hand, for resonant type electron guns, the magnetic field must be set at well defined values to get very small transverse velocities.

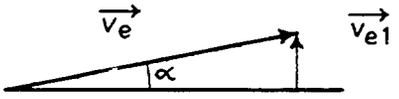


Fig. 7.6 Geometry of an imperfect magnetic field

Since the electron and ion beams must be well aligned it is very important that the magnetic field lines are strictly colinear with the theoretical electron trajectory. If the solenoid field makes an angle (Fig. 7.6) with respect to the theoretical trajectory over a length ℓ which is larger than

$$\beta_0 c \frac{1}{\omega_t},$$

the electron trajectory will follow the magnetic lines and will therefore acquire a transverse velocity

$$v_\perp \cong \alpha \beta_0 c .$$

Its transverse temperature will in consequence be increased. Therefore all the magnetic system must be built within well specified tolerances. Additional coils are used to correct the eventual imperfections. We must also mention that some small horizontal and vertical dipoles are needed to perform the e-beam steering.

7.5 Vacuum system

Storage rings in general operate under ultra-high vacuum conditions ($<10^{-10}$ Torr). This is of course valid for the cooler itself which must be bakeable in situ at 300 ^{0}C and make use of high vacuum materials and technology. A carefully estimated vacuum budget has to be established in order to evaluate the necessary pumping speeds. Pressure bumps mainly occur at the level of the hot cathode, in the drift tube where some high energy electrons are lost, and at the collector.

Suitable pumps are non-evaporable getters (NEG). Titanium sublimation pumps are sometimes used for special applications and for short periods. The NEG-pumps are activated during the system bake-out.

7.6 Controls

Operation of a cooler requires quite an elaborate control system with some of the electronics placed at the terminal potential. Figure 7.7 gives a synoptic diagram.

Communication with the system at ground potential is made via a fibre-optic link. For data acquisition, standard electronic components (ADC, DAC, etc...) are used in conjunction with systems such as CAMAC or VME. With the advent of workstations (WS) many processes may run in parallel for the control of the apparatus. Programs to control the cooler can be run from different WS simultaneously accessing parameters that may be on different CAMAC loops via an Ethernet link. The J11 microprocessor situated in the system crate of each CAMAC loop acts as the CAMAC/ETHERNET interface.

If energy ramping of the cooler is needed, the use of function generators is imperative in order to perturb the machine as little as possible.

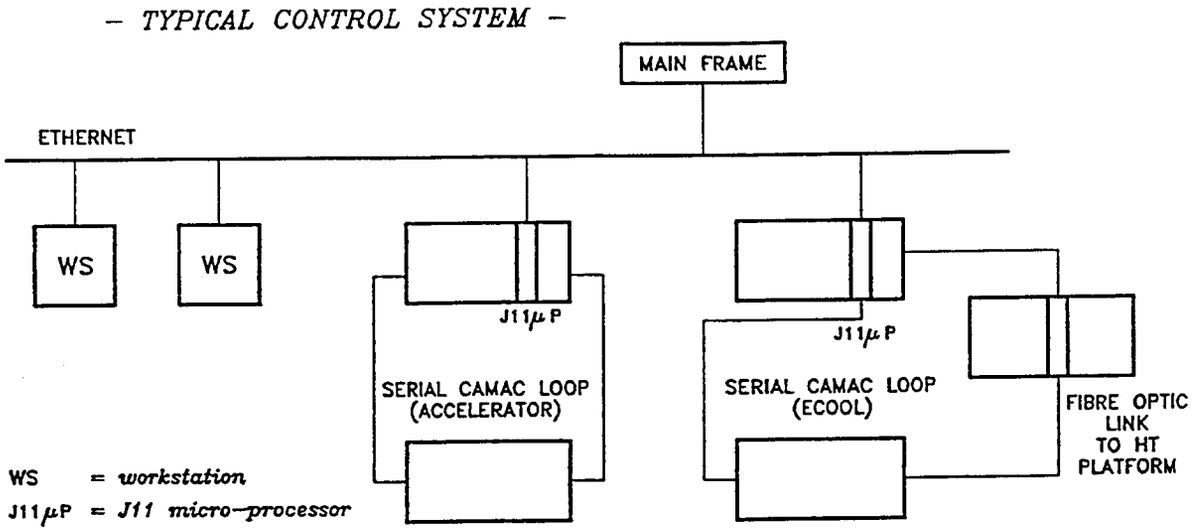


Fig. 7.7 Control synoptic diagram

8. EFFECTS ON THE ION BEAM

The interaction between the electrons and the ions is not without some disturbances. The ions will be submitted to a tune shift due either to their own space charge, which increases when cooling proceeds, or to the electron beam which acts as a lens. The solenoid will also twist the ion-beam trajectory and therefore induce a coupling between the transverse planes. When passing through the toroids, the ions will be influenced by the vertical component of the e-beam guiding field and will therefore receive a horizontal kick. As in many other accelerators, when reaching a given density threshold, the ion beam becomes unstable transversally. These coherent instabilities have to be damped. Lastly when the ion velocity comes close to that of the electron beam (the goal of any e-cooler) one can foresee nuclear interactions, which may deteriorate the ion beam lifetime. Let us now look at some details of all these phenomena.

8.1 Tune shifts

A tune shift is introduced on the ion beam due to the ion-beam space charge and to the electron beam. The ion-space charge effect is given by :

$$\Delta Q_{sp} = \frac{Z \cdot N_i}{2 \cdot \pi \cdot \varepsilon_y \beta_0^2 \gamma^3} r_i \cdot \frac{1}{B_f}$$

where B_f is the bunching factor (B_f = 1 for coasting beam). It will influence the stability at low energies when β_0 is small and when the beam is strongly cooled, ε_y small. The electron beam acts as a lens giving a tune shift:

$$\Delta Q_e = \frac{Z}{2 \cdot \beta_0^2 \gamma^3} n_e^* \cdot r_i \cdot \ell_c \cdot < \beta_y > .$$

For a fixed perveance electron gun, since $n_e^* \propto \beta_0^2$, ΔQ_e should remain constant:

Numerical example $\Delta Q_{sp} = 1.08 \cdot 10^{-3}$ for case 1, $1.08 \cdot 10^{-2}$ for case 2, and $\Delta Q_e = 4.2 \cdot 10^{-3}$.

In experiments $\Delta Q = 0.01$ was achieved.

This effect must be carefully taken into consideration mainly for variable energy storage rings and even at fixed energy due to the change in ion density. This is true since for any triplet (m,n,k) of integer the inequality $n \cdot Q_x + m \cdot Q_z \neq$ k must be fulfilled to keep the ion beam stable.

8.2 Coupling of transverse phase-space plane

The solenoid longitudinal magnetic field causes the ion beam to execute a cyclotron rotation around the longitudinal axis at frequency

$$\omega_i = \frac{Z \cdot e \cdot B_0}{m_i}.$$

Therefore, at each passage, the solenoid magnetic field will twist the ion beam by

$$\delta\theta = \frac{Z \cdot e \cdot B_0}{m_i} \frac{\ell_c}{v_0} = \frac{Z \cdot e \cdot B_0 \ell_c}{p_0} \quad \text{(Numerical example } 6.63 \cdot 10^{-2} \text{ rad)}.$$

This will induce a coupling between the horizontal and vertical planes and eventually depolarise polarised beams. It may also affect the cooling if there are different betatron functions in the two directions. To compensate this drawback, a solenoid producing a field in the opposite sense must be installed in the accelerator (usually near the cooler itself). Skew quadrupoles may also be used.

8.3 Effects on the closed orbit

The transverse component (vertical) of the guiding magnetic field in the toroids of the cooler gives a horizontal kick to the ion beam. The deflection angle is

$$\theta[rad] = \frac{\int B_z \, d\ell}{(p_0 / e)} = \frac{B_0 R_t}{(p_0 / e)} \ell n(\cos\Phi_0),$$

and the displacement of the ion beam is

$$\Delta x = \frac{Z e B_0 R_t^2}{m_i \beta_0 c \gamma_0} \left| \Phi_0 - tg\Phi_0 + \frac{1}{2} tg\Phi_0 \ell n(1 + tg^2\Phi_0) \right|$$

where Φ_0 and R_t are the bending angle and radius of the toroid. Numerically $\Delta x = 2.72 \cdot 10^{-3}$ m. Two vertical dipoles, positioned at each end of the cooler, correct to a first approximation for this displacement. For a fixed magnetic field the displacement increases as the ion energy is decreased.

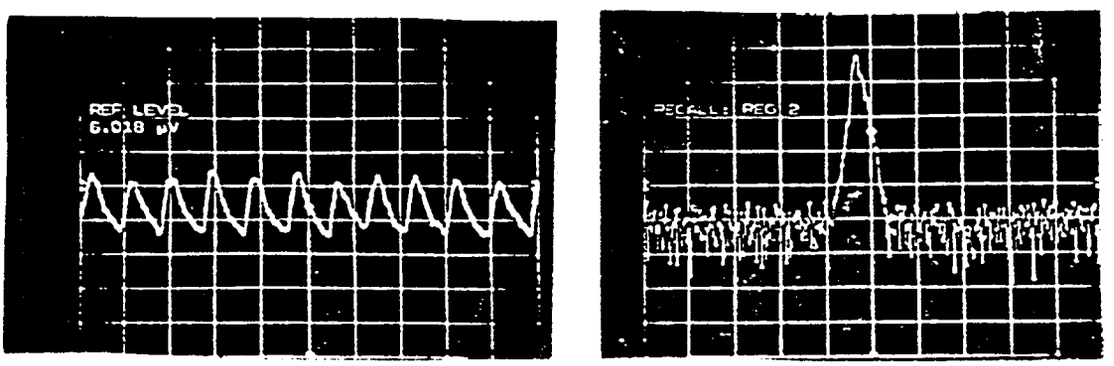
8.4 Transverse instabilities of the ion beam

One may refer to the lecture [10] and [11] on this subject.

8.4.1 Observation of instabilities

If as a result of the cooling, the density $N / \varepsilon_h \varepsilon_v \varepsilon_l$ becomes larger than a given threshold, the ion beam becomes unstable. Roughly speaking, for our standard machine this happens when N, the number of circulating particles, exceeds 10^9. This phenomenon is observed with:

- *Schottky transverse pickups.* Figure 8.1 shows this measurement with a spectrum analyser set at a fixed frequency (frequency span = 0) at one of the betatron sidebands $(n \pm Q) \cdot f_0$. More explicitly, at the time where the threshold occurs one sees an abrupt change of the transverse Schottky signal followed by a cooling period of about 10s until the density threshold is again reached. In Fig. 8.1a) the height of the transverse Schottky signal near 40 MHz is displayed. The horizontal scale is 1 s/div. The instability occurring at this frequency causes a large coherent oscillation which smears out and leads to an emittance growth during 0.2 s. This blow-up is then compensated during 0.8 s by the cooling before the next burst of instabilities occurs.



a) b)

a) Center frequency: 49.82 MHz, Span: 0 Hz, Sweep time: 10 s

b) Center frequency: 159.60 kHz, Span: 20 kHz, Sweep time: 1 s

Fig. 8.1 Observation of transverse instabilities on (a) the Schottky pickup and (b) on the position pickup

- *Transverse position pickups.* It is well known that in the case of instability the beam centre is represented by a travelling wave pattern described by:

$$y(s,t) = y_0 e^{j\left[(n \pm Q)\cdot \omega_0 \cdot t \pm \frac{n \cdot s}{R}\right]} \tag{8.1}$$

where n, an integer, is named the mode of oscillation. At a fixed azimuth s the pickup will see the mode frequency:

$$\omega_n = (n \pm Q)\omega_0 \,. \tag{8.2}$$

In most cases the instabilities are observed for lower order ($Q_h = 2.3, Q_v = 2.7$, modes are $n = 3, 4, 5$) slow waves corresponding to the minus sign in Eq. (8.2). In Fig. 8.1b) the spectral density at low frequency near one of the bands where the instability occurs is

displayed. The spike representing the beam oscillation jumps up with each burst of the instability. Usually, when the instability occurs, part of the beam is lost until $N < 10^9$ ions. These instabilities are also observed in many other situations.

8.4.2 Causes of instabilities

Normal beams tend to resist coherent instabilities by virtue of Landau damping. A small difference between the oscillation frequency prevents the ensemble responding coherently to the driving force exerted by the beam induced fields. On the other hand, cooled beams become susceptible responding coherently for at least two reasons:

- The tune spread due to non-linearities and the momentum dependence decreases.
- Some of the induced fields, such as the direct space-charge field, increase as the beam cools down.

When differentiating formula (8.2) one can define

- a mode frequency shift: $\Delta\omega_n = \pm\Delta Q \cdot \omega_0$
- a mode frequency spread: $\delta\omega_n = (n \pm Q) \cdot \delta\omega_0 \quad - \omega_0 \cdot \delta Q$.

If the mode frequency shift has an imaginary part, see section 8.4.1, we can see that the oscillation pattern can self amplify.

As a rule of thumb stabilisation of transverse instabilities by Landau damping requires that

$$|\Delta\omega_n| \leq \frac{1}{\pi} \left| (n \pm Q)\delta w_0 - w_0 \cdot \delta Q \right|.$$

To stay within the scope of this lecture we can not go deeper into this subject. One can just mention the Keil-Schnell criteria which states that at low energy the space-charge contribution to the longitudinal stability is obtained when (see lecture on impedances and Landau damping these proceedings [6-7]):

$$\left(\frac{\Delta p}{p_0}\right)^2 > \left|\frac{Z_{\|}sc}{n}\right| \frac{e}{m_i c^2} \frac{I_i}{\beta_0^2 \gamma |\eta|}$$

where I_i is the ion-beam intensity, and

$$\frac{Z_{\|}sc}{n} = \frac{377}{2\beta_0 \gamma_0^2} g$$

is the longitudinal impedance in Ω, and

$$g = 1 + 2\ell n \left(\frac{\text{chamber height}}{\text{beam height}}\right).$$

(Numerically g is 5.9 Ω for cooled beams, 3.5 Ω for normal beams). Other contributions to the impedance come from the changes in the wall resistivities of chamber sizes, RF cavities, ferrites and dielectric structures seen by the beam. So, to a first approximation:

$$\frac{\Delta p}{p_0} > 8.783 \cdot 10^{-9} \frac{\left[N' \cdot f_0\right]^{1/2}}{\left[\beta_0 \gamma_0\right]^{3/2} |\eta|^{1/2}}$$

where N' is the number of particles expressed in units of 10^9. For $N' = 1$ one obtains:

Momentum MeV/c	309	200	100	60
$\left(\frac{\Delta p}{p_0}\right)_{\min}$	$5 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$1.6 \cdot 10^{-4}$	$2.63 \cdot 10^{-3}$

On the other hand, the transverse stability criterion is given by the following inequality:

$$\frac{Z_\perp}{|n-Q|} \leq \frac{F}{G} \frac{m_i c^2}{e} \frac{2Qb^2}{I_i R^2} \beta_0 \gamma_0 \frac{\Delta p}{p_0} |(n \pm Q)\eta - \xi Q|$$

where F, G are form factors of the order of 1, b is the vacuum chamber height (35 mm for our standard machine), R is the mean radius of the accelerator (≅12.5 m), Q the tune ($Q_h = 2.305$, $Q_v = 2.73$),

$$\eta = \frac{1}{\gamma_0^2} - \frac{1}{\gamma_{tr}^2} \cong 1 \text{ at our standard machine}$$

and ζ is the chromaticity. $Z_\perp / |n-Q|$ is the transverse impedance converted into longitudinal units via:

$$\frac{Z_\parallel}{n} = \frac{b^2}{R} \frac{Z_\perp}{n-Q}.$$

This is valid for simple structures but not for the space-charge contribution to the transverse impedance, which in the transverse coasting beam case is:

$$\frac{Z_\perp}{n-Q} = \frac{377}{2\beta_0^2 \gamma_0^2} \left(\frac{b^2}{a^2} - 1\right) \Omega$$

a being the beam height. In consequence transverse stability requires that:

$$\frac{\Delta p}{p_0} > \left(\frac{Z_\parallel}{n}\right) \cdot \left(\frac{e}{m_i c^2}\right) \cdot \frac{I_i R^2}{2 \cdot Q \cdot b^2} \cdot \frac{1}{\beta_0 \gamma_0} \cdot \frac{1}{|(n \pm Q)\eta - \xi Q|}.$$

If one takes:

$$(n-Q)\eta - \xi Q = 0.3$$

we come to:

$$\frac{\Delta p}{p_0} > 7.13 \cdot 10^{-12} \frac{N' f_0}{(\beta_0 \gamma_0)^2} .$$

For $N' = 1$ one obtains:

Momentum MeV/c	309	200	100	50
$\left(\frac{\Delta p}{p_0}\right)_{min}$	$7.85 \cdot 10^{-5}$	$1.25 \cdot 10^{-4}$	$2.57 \cdot 10^{-4}$	$4.18 \cdot 10^{-4}$

The above tables show that for a beam of a few 10^9 particles the loss of Landau damping occurs (with our standard machine) approximately when the longitudinal cooling reaches the ratio $\Delta p / p_0 = 10^{-4}$. This is directly linked to the double-peaked spectrum signal observed with strong cooling. Of course, many other phenomena may induce instabilities, such as the interaction of electrons from the residual gas trapped in the proton beam. By the way, the impedances (Z/n) have to be measured rather than estimated.

It is worth mentioning that coherent motion gives signals, on a spectrum analyser, which are proportional to N, the number of particles, while for Schottky (random) signals the amplitudes are proportional to $\sqrt{N}$.

8.4.3 Partial cure of transverse instabilities

In order to suppress the transverse coherent instability a damper is an important electronic device for stabilising dense beams. A simple damper (Fig. 8.2) consists mainly of:

- a horizontal and vertical electrostatic pickup,
- a horizontal and vertical kicker placed at a given distance S_0 from the pickups.

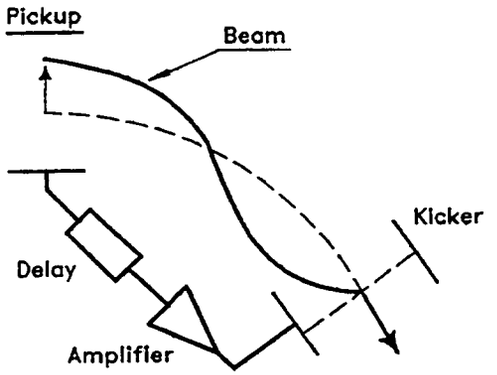


Fig. 8.2 Principle of the damper

The transverse position signal, taken from the position pickup is linearly amplified, appropriately delayed then applied to the kicker plates. If we take the phase from Eq. (8.1):

$$\theta = (n-Q)\cdot\omega_0\cdot t - \frac{ns}{R}.$$

$s = S_0$ is chosen such that the phase of the particle has increased by an odd multiple of $\pi/2$ between the pickup and the kicker (usually named $\lambda/4$ advance). The amplifier bandwidth is directly related to the mode n of oscillations to be covered and to its time response to the growth rate of the instability. With such a damper high intensity beams with large densities can be obtained which are very useful for physics experiments.

8.5 Other electron-ion interactions

At the beginning of this lecture, we mentioned that in order to explain the cooling process the electron beam may be considered as a target. There are some other reactions that should be considered. The most dominant are:

$A^{Z+} + e^- \rightarrow A^{(Z-1)+} + h\upsilon$	spontaneous radiative recombination
$A^{Z+} + e^- \rightarrow \left[A^{(Z-1)+*}\right] \rightarrow A^{(Z-1)+} + h\upsilon$	dielectronic recombination
$A^{Z+} + e^- \rightarrow (A^{Z+})^* + e^-$	ion excitation
$A^{Z+} + e^- \rightarrow A^{(Z+1)*} + e^- + e^-$	ionisation.

These are processes which should be taken into account near thermal equilibrium and therefore may influence the cold ion-beam lifetime.

Spontaneous recombination

When positive ions are cooled by electrons, occasionally cooling electrons are radiatevely captured by beam ions into atomic states with main quantum number n. If N is the number of stored ions, we can define

$$\frac{1}{N}\cdot\frac{dN}{dt} = -\frac{1}{\tau} = -\frac{R_a}{N}.$$

Therefore the number of lost ions per unit of time is

$$R_a = \frac{N\cdot\eta_c\cdot\alpha_r\cdot n_e^*}{\gamma_0^2};\quad \alpha_r = 9.3\cdot10^{-19}Z^2(\mathrm{m^3s^{-1}\,eV^{1/2}})(\mathrm{k}T)^{-1/2}.$$

For kT = 1 eV; $\alpha_r = 9.3\cdot10^{-19}$, $R_a = 2537$.

In the case of protons (ions) the detection of these neutral atoms (recharged ions) is an indirect way to adjust the cooler parameters, until one obtains the maximum H_0 rate, and to measure the ion beam transverse emittance.

Dielectronic recombination

This reaction will come into play if the circulating ions are not fully stripped. An electron is captured in an auto-ionising state, and the energy gained is used to lift one of the core electrons to an excited state. The auto-ionising states decay either back into the original channel or into a lower state.

The phenomenon is of resonant type and so will depend on the ion particular state. When the resonant condition is fulfilled, the cross-section attains very high values of the order of 10^6 mBarn.

We will not comment on the other two phenomena which are more relevant to nuclear physics (see lecture [4]).

9. PRESENT AND FUTURE ELECTRON COOLER RINGS

Table 9.1 shows the design parameters of the present and near future electron-cooling rings. Most of them are dedicated to atomic physics experiments. The pioneer coolers built at INP Novosibirsk, Fermilab and at CERN (ICE) are no longer operational. The CERN-LEAR cooler is not represented in this list since it is already mentioned in section 2 as a typical accelerator. The numerical applications refer to data and numbers which have been physically measured. COSY will be the next electron cooler to go into operation while classical cooler projects are underway in USSR (Dubna, Kiev). A study of a 6 MeV electron energy cooler is underway at Bloomington (U.S.A.) and the results are awaited with interest.

Figure 9.1 shows the layout of five of the machines mentioned in Table 9.1. It is easy to recognise the fundamental components mentioned in this lecture. Also worth mentioning is the fact that, except for COSY and Dubna K4 - K10, they have reached their design parameters.

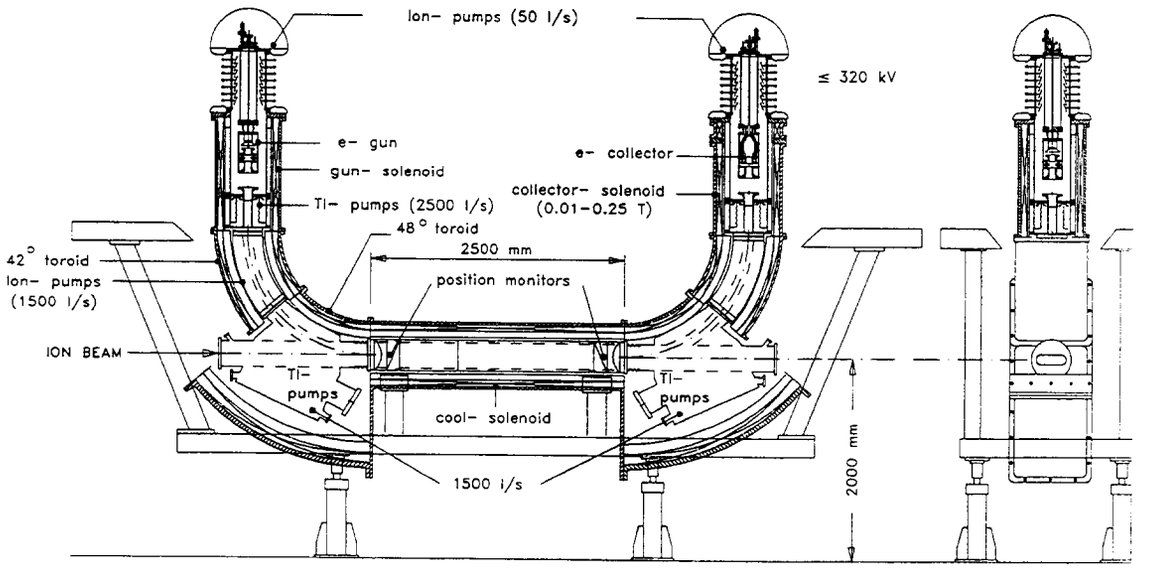


Fig. 9.1a) Cross-sectional view of the ESR electron cooler

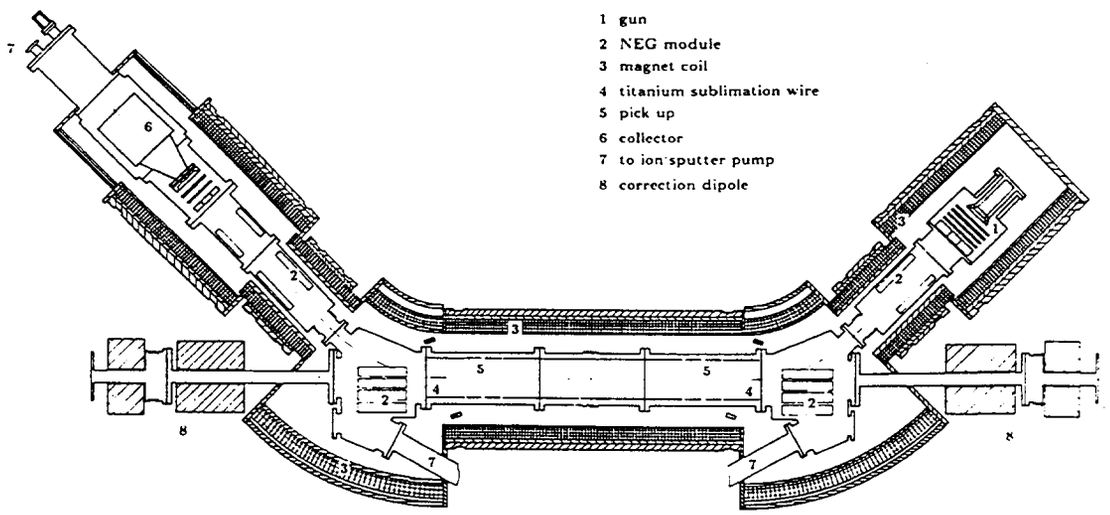


Fig. 9.1b) Layout of the TSR electron cooler

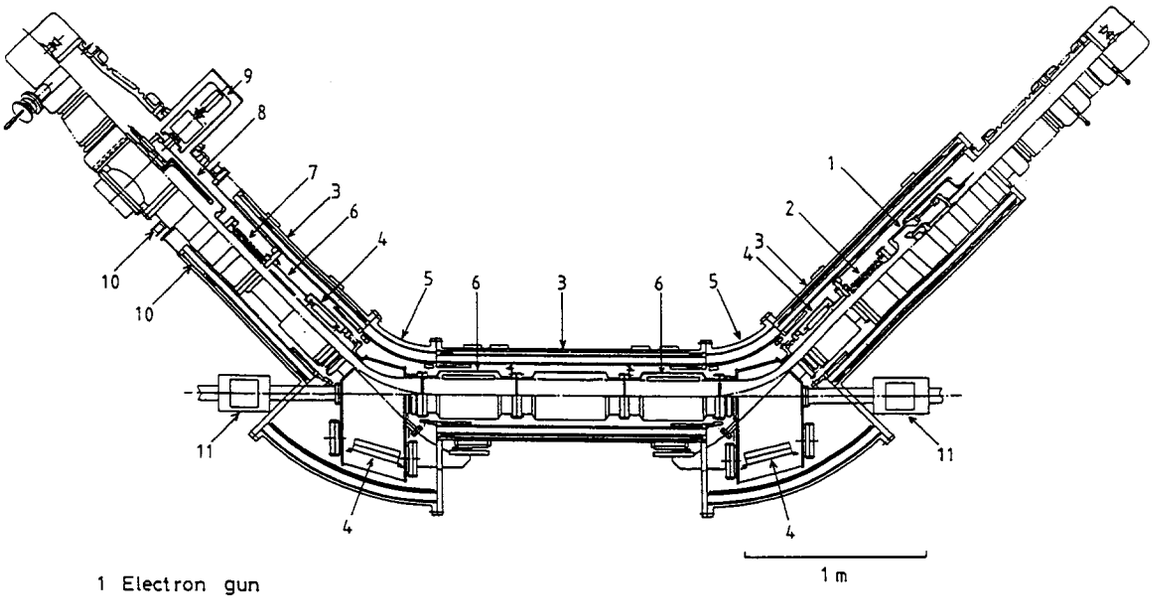


Fig. 9.1c) Layout of the TARN II electron cooler device

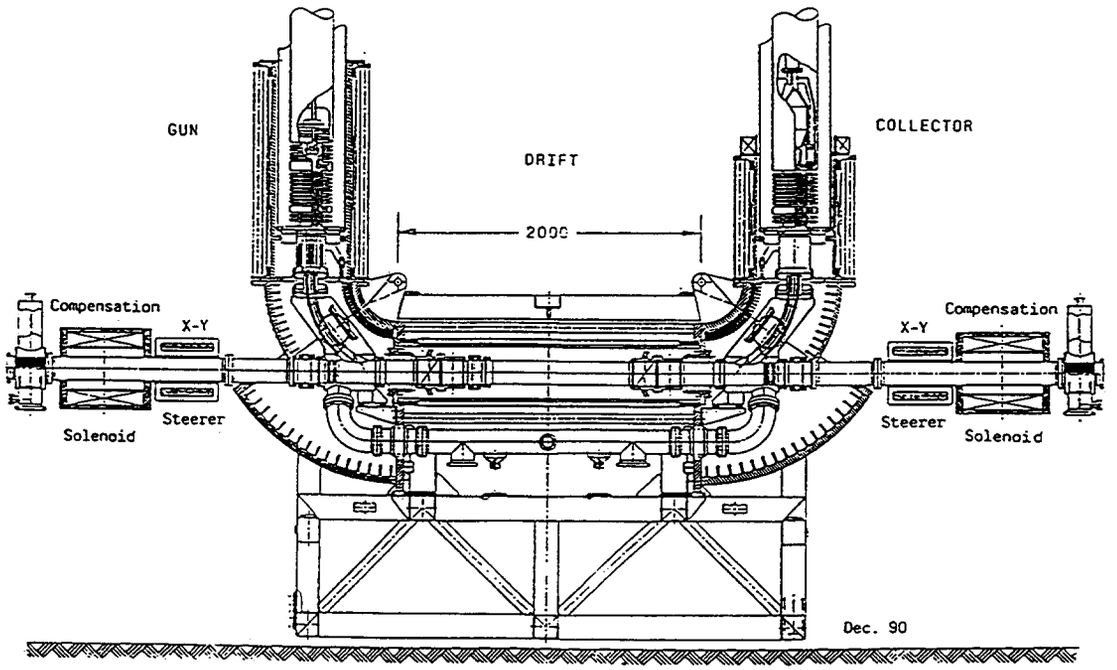


Fig. 9.1d) Longitudinal section of the COSY electron cooler

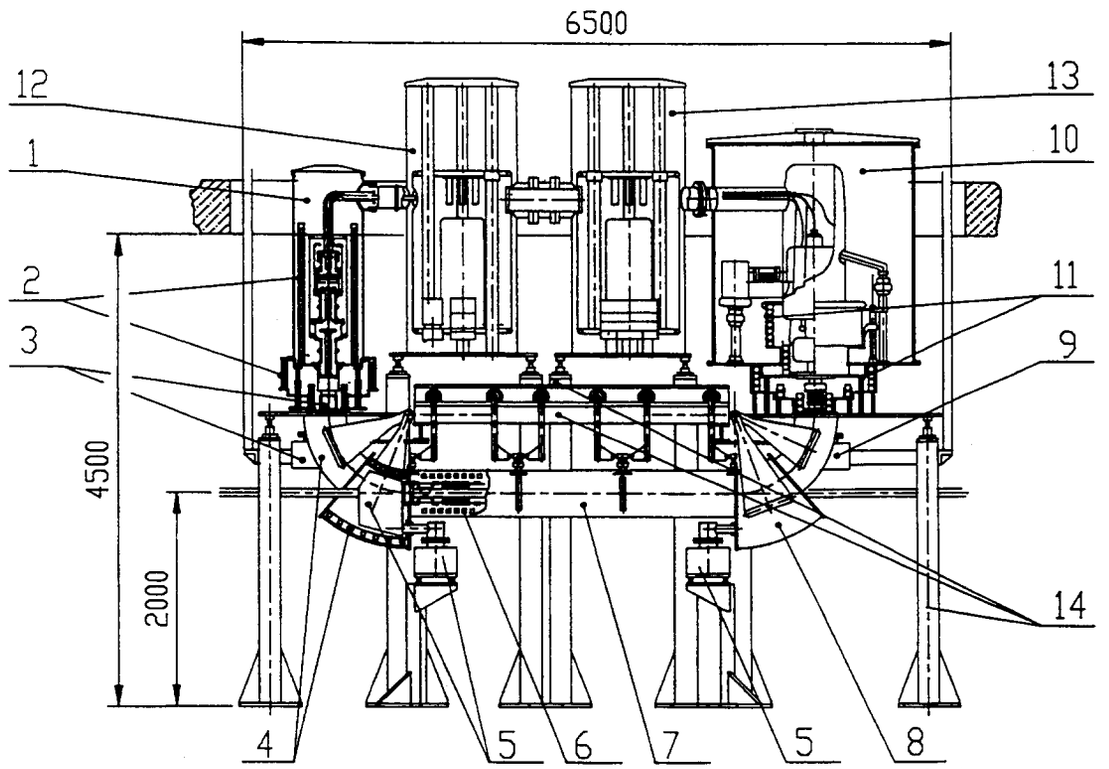


Fig. 9.1e) Electron cooling system of the ring K10: *1 and 10* = electron gun and collector (respectively) in their tanks; *2 and 11* = solenoids of the electron gun and collector (respectively); *3, 5 and 9* = vacuum pumps; *4, 8* = toroid magnets, *6,7* = the winding and the body of the main solenoid, *12, 13* = power supplies; *14* = platform

Table 9.1

Cooling Ring Characteristics

Name of Ring	IUCF COOLER	TSR	CELSIUS	TARN II	ESR	DUBNA		CRYRING	COSY
	Bloomington	Heidelberg	Uppsala	Tokyo	Darmstadt	K4	K10	Stockholm	Jülich
Circumference (m)	86.8	55.4	81.8	77.8	108.4	83.12	146.24	51.6	184
Magnetic rigidity (T·m)	3.6	1.5	7	6.1	10	4.0/0.7	10.0/0.7	1.44	11
Energy range (MeV/A)	500	5-30	1360 (for p)	100, 400 (for p)	30-560	170	830	0.3-24	40-2500
Type of ions:									
- highest charge	3	47	18	10	92	92			1
- highest A	7	127	40	20	238	238			1
- lowest Q/M	0.5	0.5		0.5	0.5	0.2		0.5	1
Expected ion intensity	10 mA	10 mA	10^{11} (p)	10^{8}	up to 10^{11}	3×10^{7} - 8×10^{11}		10^{9}	$\leq 1\times10^{10}$
Average ring vacuum (Torr)	$< 3\times10^{-9}$	$< 10^{-11}$	10^{-11}	10^{-11}	10^{-11}	2×10^{-10}		10^{-11}	$\leq 1\times10^{-10}$
Horizontal acceptance (π mm·mrad)	25	500	120	260	450	50	25	200	180
Vertical acceptance (π mm·mrad)	25	120	120	15	150	50	25	100	36
Longitudinal acceptance for $\varepsilon = 0$ (%)	$\pm$ 0.2	$\pm$ 3	$\pm$ 0.3	$\pm$ 0.1	$\pm$ 2	1.0	2.0	$\pm$ 1.7	$\pm$ 0.5
Length of cooling section (m)	2.8	1.5	2.5	1.5	2.5	3	3	1.1	2
Typical working point:									
- Q_h	4.15	2.75	1.68	1.75	2.1-2.45	2.4	2.8	2.3	3.38
- Q_v	5.15	2.825	1.9	1.8	2.1-2.45	2.8	3.3	2.27	3.38
Average horizontal β function (m)	2.7	3.15	7.75	8	8-10	5.5	8.3	4	4.5
β_h in cooling section (m)	2.3	5.5	8.14	10	8	10.9	5.6	2.3	10
β_v in cooling section (m)	4	5.7	5.04	4	8-10	6.3	5.8	2.9	10
Dispersion in cooling section (m)	0	0-1	1.48	4.7	0-6	0	0	1.6	0
Transition γ_t	4.85	2.96	2.605	1.88	2.7	5.26	2.26	2.23	1-4
Internal target type:	Jet/dust fibre	Storage cell	Cluster/C fibre		Jet	Jet / cluster / "microclusters"			Cluster
- density (atoms/cm^2)	1-100 ng/cm^2	10^{14}			10^{14}	10^{11} - 10^{14}			10^{14}
- nuclear mass		H↑			1-129	1 - 238			
First circulating beam	1988	1989	1989	1989	1990	1996	1997	1992	1992
Cathode/beam diameter (cm)	2.54	5	2	5	5	3	3	4	2.54
Electron energy range (keV)	10-270	3-20	10-300	≤ 120	10-320	15-100	100-250	2-20	20-30 (100)
Electron current (A)	0-4.8	1	0-2.8	≤ 10	0.5-5	0.8-5	5	0.001-3	1.5-2.7
Nominal gun perveance (μP)	0.7	1.7	0.36 (40 kV)	1	1.85	0.44 - 0.16	0.16-0.04	0.1-5	0.1-0.6
Magnetic field (kG):	1-1.5	0.55	0.5-2	1.2	0.1-2.5	1-2	2	2.5	0.8
- at electron energy (keV)				120		15-100	100-200	13	22
Toroidal angle (°)	60	45	90	45	90	90	90	50	53
Gun-collector voltage (kV):	2-7	1.2-2			5	2	2	1.5	3
- at electron-energy (keV)		7			310			13	22
- and current (A)		1			10			1	2
Vacuum in cooler (Torr)	1×10^{-9}	$< 10^{-11}$		10^{-11}	$\leq 10^{-10}$	1×10^{-8}	1×10^{-8}	10^{-11}	1×10^{-10}
Electron losses (collector)	10^{-6}-10^{-3}	$< 10^{-4}$		10^{-4}	$\leq 10^{-3}$	$< 10^{-4}$	10^{-4}	10^{-4}	1×10^{-4}

10. CONCLUSION

Though we have chosen an approach which is more didactic then rigourous, we hope that the usefulness of electron cooling, for low-energy storage rings, has been sufficiently explained.

Amongst all the cooling processes (at least for accelerators operating at momenta lower than about 600 MeV/c) electron cooling is presently the only one which makes it possible to reach very small emittances in very short times. However, it involves a broad knowledge of many physical domains and needs good engineering of almost all the components in order to reach the required performances.

The cooler itself is only part of the process. A careful study must be made, and many precautions must be taken, concerning the insertion of this device into the storage ring. This is particularly true for variable energy accelerators, where the effects on the ion beam (Q shifts, closed orbit, instabilities) must be minimised.

REFERENCES

Almost all the references which refer to electron cooling can be found in the basic report [1] (which has been taken as a reference for this lecture):

[1] H. Poth, Electron Cooling: Theory, Experiments, Application, CERN-EP/90-04 or A. Wolf, thesis "Elektronkühlung für Niederenergische Antiprotonen", Karlsruhe, KfK 4023 (1986).

Written in the frame of a lecture are:

[2] H. Poth, Electron Cooling, CAS, Oxford, CERN/87-03, Vol. 2, September (1985).

[3] B. Seligmann, An Introduction to Electron Cooling, CAS, Jülich, CERN/91-04, September (1990).

More elaborate reports are:

[4] B. Franzke, Beam Interactions with Residual Gas, these proceedings.

[5] A. Piwinski, Intra-Beam Scattering, these proceedings.

[6] C. Ekström, Internal Target, these proceedings.

[7] D. Boussard, Schottky Noise and Beam Transfer Function Diagnostics, Proceedings. CAS, Oxford, September 1985.

[8] S. Baird, Schottky Noise on Very Cold Beams, Proceedings CAS, Uppsala, September 1989, CERN 90-04.

[9] C. Ellert, Aufbau und erste Experimente mit dem Induktionsberchleuniger am Heidelberg Speicherring TSR. MPI-V6, (1991)

[10] A. Hofmann, Vlasov Equation and Landau Damping, these proceedings.

[11] I. Hofmann, Space-charge Impedance, these proceedings.

BIBLIOGRAPHY

J.D. Jackson, Classical Electrodynamics, 2nd Edition, John Wiley & Sons, Inc.

G.I. Budker, Proceedings, International Symposium on Electron and Positron Storage Rings, Saclay (1966), (Presses Universitaires de France, Paris, 1967).

Y. Debernev, I. Meshkov et al., CERN 77-08 yellow report.

L. Spitzer, Physics of Fully Ionised Gases, Interscience Publishers (1962).

Y. Debernev and A.N. Skrinsky, Particle Accelerator 8 (1977) 1.

A.H. Sorensen and Bonderup, Nucl. Instr. and Methods 215 (1983) 27.

N.S. Dikanski et al., INP Novosibirsk preprint 88-61 (1988).

Proceedings of the Workshop on ECOOL-1984, Karlsruhe, Editor: H. Poth, KfK 3846,

Proceedings of the Workshop on ECOOL-1990, Legnaro, Editors: R. Calabrese and L. Tecchio, World Scientific (1991).

LASER COOLING

Ejvind Bonderup

Institute of Physics, Aarhus University, Aarhus, Denmark

Abstract

The velocity distribution of an ionic or atomic beam may be modified through interaction with a laser field with a frequency close to a transition frequency in the ion or atom. Efficient longitudinal cooling may be achieved in an intense field forming a standing wave in a reference frame moving with the average particle velocity. The chapter concentrates on the theoretical description of this effect, and experimental results are presented only briefly towards the end. The origin of the force on an atom from a light wave is first explained in terms of the Lorentz force. In a strong field, quantitative evaluations may conveniently be performed within the dressed-atom picture, based on the eigenstates for the atom-laser system. At low velocities of the particle relative to the standing wave, the force may be interpreted in terms of motion on potential curves with occasional interruptions owing to spontaneous emission. The final section contains a summary of how the force depends on laser detuning and intensity and on relative velocity, and typical numerical values are given for key quantities.

1. INTRODUCTION

This chapter deals with the possibility of decreasing the longitudinal velocity spread of a beam of atoms or ions through the interaction of the particles with a strong laser field with a frequency close to a transition frequency in the atom or ion. The experimental investigation of this method of modifying particle beams has just begun, and at present it is not clear to what extent laser cooling may become a supplement to electron cooling and stochastic cooling in the future. Towards the end of the chapter, the results of two experiments will be mentioned but the emphasis will be on the description of the mechanism behind the effect.

Laser cooling can be applied only to selected atoms and ions since the states involved in the laser induced transitions must form a so-called closed system, i.e. spontaneous transitions to other states should be negligible. To obtain a symmetric compression of the longitudinal velocity distribution of a particle beam, one must apply two laser waves, travelling in the directions parallel and antiparallel to the particles. In the laboratory, the frequencies of the two travelling waves should differ by such an amount that the Doppler-shifted frequencies become equal in a reference frame moving with the average particle velocity. In the moving frame, the superposition of the travelling waves with the same frequency leads to a standing wave, and the basic problem is therefore to understand the influence of a standing laser field on particles moving with velocities of the order of the velocity spread in the beam. The laser affects only particles moving sufficiently slowly with respect to the standing wave. The velocity region accessible to laser cooling increases with the detuning of the radiation frequency away from the atomic transition frequency, provided the increase in detuning is accompanied by a sufficient increase in light intensity. Typically, velocities of 10^3 m/sec may be damped out during an interaction time of 10^{-3} seconds.

We first consider an atom at rest, and the origin and direction of the force from a light wave are explained within a simple classical model with an electron bound in a harmonic oscillator. For weak fields, the transition to quantum mechanics then corresponds to the replacement of the binding frequency by a transition frequency and the introduction of an

oscillator strength, as in the theory of atomic polarizability. The difference between the forces from a travelling wave and a standing wave of the same strength is illustrated. A transparent description of the force on an atom moving through a strong standing laser field may be given in terms of the so-called dressed-atom states which are eigenstates of the interacting atom-laser system, and this picture will be introduced and discussed in some detail. Finally, we summarize how the direction and strength of the force depend on the particle velocity with respect to the standing wave, on the light intensity, and on the detuning of the laser frequency.

2. ORIGIN OF FORCES

2.1 Classical model

Consider an electron with charge $-e$ and mass m bound in a harmonic oscillator with frequency ω_0 and damping constant Γ. In addition to the elastic force, the electron experiences the electric and magnetic fields from a light wave. Assuming a linearly polarized plane wave, we may write the electric field in the form

$$\boldsymbol{E}(z,t) = \boldsymbol{\varepsilon}_x \frac{1}{2}\left\{E(z)e^{-i\omega t} + E^*(z)e^{i\omega t}\right\} = \boldsymbol{\varepsilon}_x \,\mathrm{Re}\left\{E(z)e^{-i\omega t}\right\}, \tag{1}$$

where $\boldsymbol{\varepsilon}_x$ denotes a unit vector in the x direction. The two cases $E(z)=E_0\cos(kz)$ and $E(z) = E_0\exp(ikz)$, with E_0 real, correspond to a standing wave and a travelling wave, respectively. The wavelength of the light is assumed to be long compared to the displacement of the electron from its equilibrium position (0, 0, z), such that the fields may be evaluated at this fixed point in space. For nonrelativistic particle motion, the magnetic force is much weaker than the electric force, and we therefore at first neglect it. To this approximation, the electron displacement x satisfies the equation of motion

$$\ddot{x} + \Gamma\dot{x} + \omega_0^2 x = -\frac{e}{m}\mathrm{Re}\left\{E(z)e^{-i\omega t}\right\}. \tag{2}$$

After the disappearance of transients, we obtain for the dipole moment $\boldsymbol{d}(t) = d(t)\boldsymbol{\varepsilon}_x$,

$$d(t) = -ex(t) = \mathrm{Re}\left\{\alpha(\omega)E(z)e^{-i\omega t}\right\}, \tag{3}$$

where the complex polarizability $\alpha(\omega)$ is given by the expression

$$\alpha(\omega) = \frac{e^2/m}{\omega_0^2 - \omega^2 - i\omega\Gamma}. \tag{4}$$

The time average of the electric force on an electron vanishes. This is in contrast to the magnetic part of the Lorentz force which results from the interaction of the induced oscillating dipole with the magnetic field from the light wave. According to the Maxwell equation curl $\boldsymbol{E} = -c^{-1}\partial\boldsymbol{B}/\partial t$, the electric field in Eq. (1) is accompanied by the magnetic field

$$\boldsymbol{B}(z,t) = -\boldsymbol{\varepsilon}_y \,\mathrm{Re}\left\{\frac{ic}{\omega}\frac{\partial E(z)}{\partial z}e^{-i\omega t}\right\}. \tag{5}$$

Denoting a time average by the symbol < >, we then obtain the force

$$\boldsymbol{F} = -\frac{e}{c} < \dot{x}\boldsymbol{\varepsilon}_x \times \boldsymbol{B}(z,t) > \; = \frac{1}{2}\mathrm{Re}\left\{\alpha(\omega)E(z)\frac{\partial E^*(z)}{\partial z}\right\}\boldsymbol{\varepsilon}_z \; . \tag{6}$$

2.2 Transition to quantum mechanics in weak fields

A quantal treatment of an atom, perturbed by a weak field of the form in Eq. (1), leads to a very similar polarizability [1],

$$\alpha(\omega) = \frac{e^2}{m}\sum_n \frac{f_{n0}}{\omega_{n0}^2 - \omega^2 - i\omega\Gamma_n} \; , \tag{7}$$

where ω_{n0} is the transition frequency from the ground state $| 0 >$ to the excited state $| n >$ of linewidth $\hbar\Gamma_n$, and where the quantity f_{n0}, proportional to the square of a dipole matrix element, is the dipole oscillator strength,

$$f_{n0} = \frac{2m}{\hbar}\omega_{n0}|< n|\sum_{j=1}^{Z} x_j|0 >|^2 \; . \tag{8}$$

Here x_j, $j = 1, 2, \ldots Z$ denotes a coordinate of the j'th electron, and the oscillator strengths fulfil the sum rule

$$\sum_n f_{n0} = Z \; . \tag{9}$$

2.3 Scattering force in travelling wave

In a travelling wave, with $E(z) = E_0 \exp(ikz)$, Eq. (6) leads to the scattering force or radiation pressure

$$\boldsymbol{F}_{sc} = \frac{1}{2}E_0^2 k \;\; \mathrm{Im}\;\alpha(\omega)\boldsymbol{\varepsilon}_z \; . \tag{10}$$

As the names indicate, this force is related to the momentum transfer resulting from the scattering of the light wave by the atom. Energy and momentum are flowing in the direction of motion of the travelling wave, indicated by the unit vector $\boldsymbol{\varepsilon}_z$, and an energy transfer ΔE to the atom is accompanied by a momentum transfer $(\Delta E/c)\,\boldsymbol{\varepsilon}_z$. In a weak field, the energy absorbed by the atom is reemitted in spontaneous decays but since this emission is of equal strength in opposite directions, there is, on the average, no momentum change associated with these processes. The magnitude of the force on the atom therefore equals P/c, where P denotes the power transfer from the light when the electric field drives the oscillating dipole, i.e.,

$$\boldsymbol{F}_{sc} = \frac{1}{c}\;P\;\boldsymbol{\varepsilon}_z = \frac{1}{c}\langle -eE(z,t)\dot{x}\rangle\boldsymbol{\varepsilon}_z \; . \tag{11}$$

From this formula we immediately retrieve Eq. (10).

For a one-electron atom, we now introduce the expressions (7) and (8) into Eq. (10). If the light frequency ω is close to the transition frequency ω_{n0} for a closed two-state system, we find the approximate formula

$$F_{sc} = \frac{1}{4}\frac{e^2}{\hbar c}\frac{\Gamma_n \omega_{n0}|x_{n0}|^2}{\delta^2 + \Gamma_n^2/4} E_0^2 \varepsilon_z \ , \tag{12}$$

where the constant $e^2/\hbar c$ equals the fine-structure constant, and where we have introduced the detuning

$$\delta \equiv \omega - \omega_{n0} \ . \tag{13}$$

On the average, a momentum transfer from a travelling wave to an atom occurs only if the absorption is followed by spontaneous and not by induced emission since the induced radiation is emitted in the direction ε_z. As a result, the scattering force in Eq. (12) must saturate when the absorption rate $F_{sc}/(\hbar\omega_{n0}/c)$ becomes of the order of the spontaneous emission rate Γ_n, i.e., when the *Rabi frequency* $\omega_R \equiv eE_0|x_{n0}|/\hbar$ becomes of the order

$$\omega_R \equiv eE_0|x_{n0}|/\hbar = 2^{1/2}\left(\delta^2 + \Gamma_n^2/4\right)^{1/2} \ . \tag{14}$$

At this electric field strength, we are at the limit of the perturbation treatment since the population of the excited state becomes significant. With the numerical factor $2^{1/2}$ in Eq. (14), a combination of the equations (12) and (14) leads to the correct value of the saturated scattering force,

$$F_{sc}^{sat} = \frac{1}{2}\frac{\hbar\omega}{c}\Gamma_n \ , \tag{15}$$

where the factor 1/2 accounts for the population of the upper level at saturation.

At this point it may be appropriate to give the order of magnitude of important quantities. With a photon energy $\hbar\omega = 2$ eV and a lifetime $\Gamma_n^{-1} = 10^{-8}$ sec we obtain $F_{sc}^{sat} \approx 0.5 \cdot 10^{-19}$ Newton. For a projectile of mass number A, the corresponding acceleration is $\sim 0.3 \cdot A^{-1} \cdot 10^8$ m/sec^2, such that within a microsecond a neon atom experiences a velocity change of ~1.5 m/sec. For a two-state system we have a relation between the matrix element x_{n0} and the line width Γ_n in Eq. (14), $\Gamma_n = (4/3)\alpha\omega^3|x_{n0}|^2/c^2$ where α is the fine-structure constant. In the case considered, we then find that at resonance ($\delta = 0$), saturation is obtained at an electric field strength of ~ 260 V/m. This again corresponds to an energy flux of $0.9 \cdot 10^{-4}$ Watt/mm^2.

2.4 Dipole force in standing wave

We now turn to a standing wave, and with $E(z) = E_0 \cos(kz)$ we obtain from the general formula (6) the so-called *dipole force*

$$F_d = \frac{1}{4}\{\mathrm{Re}\ \alpha(\omega)\}\frac{\partial}{\partial z}\left(E_0 \cos(kz)\right)^2 \varepsilon_z \ . \tag{16}$$

Since this force is proportional to the real part of the polarizability, the interpretation in terms of absorption and emission is quite different from the one given in connection with the scattering force. If the standing wave is considered as the superposition of two waves travelling in opposite directions, i.e., $E(z) = E_0[\exp(ikz)+\exp(-ikz)]/2$, the force in Eq. (6) results from the cross terms whereas the two direct contributions cancel. This indicates that the dipole force is associated with the momentum change in processes where absorption from one of the waves is followed by induced emission into the other one.

When the frequency ω is close to the atomic transition frequency ω_{n0} in a closed two-state system, the dipole force may be expressed as

$$\boldsymbol{F}_d = -\frac{1}{8}\frac{e^2}{m\omega_{n0}} f_{n0} \frac{\delta}{\delta^2 + \Gamma_n^2/4} \frac{\partial}{\partial z} \left(E(z)\right)^2 \boldsymbol{\varepsilon}_z \ . \tag{17}$$

The sign of the dipole force thus depends on the sign of the detuning: For blue detuning, the force is in the direction out of the strong electric field intensity. A comparison of the magnitudes of the dipole force in a standing wave and the scattering force in a travelling wave of equal amplitude and at equal detuning may be obtained from the Eqs. (12) and (17), and replacing the derivative $\partial/\partial z$ by the wave number k of the light wave, we find

$$F_d / F_{sc} \sim |\delta| / \Gamma_n \ . \tag{18}$$

The scattering force from a travelling wave saturates at the value in Eq. (15), and if the detuning is large compared to the line width, this occurs when the Rabi frequency becomes of order $|\delta|$. As we shall see later, from Eq. (51), the dipole force has also reached saturation in the situation $\omega_R > |\delta| > \Gamma_n$, and the large factor $|\delta|/\Gamma_n$ in Eq. (18) then represents the ratio of the saturated dipole force $\sim |\delta| \hbar\omega/(2c)$ and the saturated scattering force (15).

Although the dipole force can become strong, one might not expect it to cool a beam of particles travelling through the standing wave since the expression (17) is periodic as a function of position, with vanishing average value. The reason why cooling does occur is that the forces become velocity dependent. The motion of the particle may only be disregarded as long as the spatial variation of the fields is negligible during the time necessary for the disappearance of transients, i.e., for the attainment of local statistical equilibrium. In strong fields, this time is of order $\gtrsim \Gamma_n^{-1}$, and in Section 4 we shall find an appreciable dipole force on an atom moving through a strong standing laser field already at velocities quite low compared to Γ_n/k, corresponding to motion through a distance k^{-1} during the time Γ_n^{-1}.

3. ATOM IN RADIATION FIELD

3.1 Quantization of radiation field

As an introduction to the dressed-atom picture, we briefly recall the quantization of the radiation field in the absence of charges [2]. In Coulomb gauge, div $\boldsymbol{A}$ = 0, the electric and magnetic fields may be expressed in terms of the vector potential,

$$\boldsymbol{E}(\boldsymbol{r},t) = -\frac{1}{c}\frac{\partial \boldsymbol{A}(\boldsymbol{r},t)}{\partial t} \ , \quad \text{and} \tag{19}$$

$$\boldsymbol{B}(\boldsymbol{r},t) = \ \text{curl}\, \boldsymbol{A}(\boldsymbol{r},t) \ . \tag{20}$$

Introducing a spatial Fourier decomposition of the vector potential within the quantization volume V,

$$\boldsymbol{A}(\boldsymbol{r},t) = \frac{1}{\sqrt{V}} \sum_{\boldsymbol{k}} \boldsymbol{A}_{\boldsymbol{k}}(t) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \ , \quad \boldsymbol{k}\cdot\boldsymbol{A}_{\boldsymbol{k}}(t) = 0 \ , \tag{21}$$

we obtain for the energy in the radiation field,

$$H_r = \frac{1}{8\pi}\int d^3\boldsymbol{r}\left(\boldsymbol{E}^2 + \boldsymbol{B}^2\right) = \frac{1}{8\pi}\sum_{\boldsymbol{k}}\left(k^2\left|\boldsymbol{A}_{\boldsymbol{k}}\right|^2 + \frac{1}{c^2}\left|\dot{\boldsymbol{A}}_{\boldsymbol{k}}\right|^2\right). \tag{22}$$

This expression simplifies if we introduce the function

$$\boldsymbol{f}_{\boldsymbol{k}} = \boldsymbol{A}_{\boldsymbol{k}} + \frac{i}{kc}\dot{\boldsymbol{A}}_{\boldsymbol{k}} \ . \tag{23}$$

Since the vector potential $\boldsymbol{A}(\boldsymbol{r}, t)$ is real, the Fourier coefficients $\boldsymbol{A}_{\boldsymbol{k}}$ fulfil the relation $\boldsymbol{A}^*_{-\boldsymbol{k}} = \boldsymbol{A}_{\boldsymbol{k}}$, and for the energy H_r we therefore have

$$H_r = \frac{1}{8\pi}\sum_{\boldsymbol{k}} k^2\left|\boldsymbol{f}_{\boldsymbol{k}}\right|^2 \ . \tag{24}$$

The vector $\boldsymbol{A}_{\boldsymbol{k}}$ is perpendicular to $\boldsymbol{k}$, and thus it may be written in the form

$$\boldsymbol{f}_{\boldsymbol{k}}(t) = \left(\frac{4\pi c}{k}\right)^{1/2}\sum_{\mu=1}^{2}\left(q_{\boldsymbol{k}\mu}(t) + i\ p_{\boldsymbol{k}\mu}(t)\right)\boldsymbol{\varepsilon}_{\boldsymbol{k}\mu} \tag{25}$$

where $\{\boldsymbol{\varepsilon}_{\boldsymbol{k}1}, \boldsymbol{\varepsilon}_{\boldsymbol{k}2}, \boldsymbol{k}/k\}$ is an orthonormal set, and where $q_{\boldsymbol{k}\mu}$ and $p_{\boldsymbol{k}\mu}$ are real. They fulfil very simple equations of motion which are obtained from the wave equation satisfied by the vector potential,

$$\Delta\boldsymbol{A} - \frac{1}{c^2}\frac{\partial^2\boldsymbol{A}}{\partial t^2} = \boldsymbol{0} \ . \tag{26}$$

The equivalent equation $\ddot{\boldsymbol{A}}_{\boldsymbol{k}} = -k^2c^2\boldsymbol{A}_{\boldsymbol{k}}$ gives for the function $\boldsymbol{f}$ in Eq. (23),

$$\dot{\boldsymbol{f}}_{\boldsymbol{k}} = -ikc\ \ \boldsymbol{f}_{\boldsymbol{k}} \ , \tag{27}$$

and this relation may immediately be rewritten in terms of the quantities $q_{\boldsymbol{k}\mu}$ and $p_{\boldsymbol{k}\mu}$,

$$\begin{aligned} \dot{q}_{\boldsymbol{k}\mu} &= kcp_{\boldsymbol{k}\mu} \ , \\ \dot{p}_{\boldsymbol{k}\mu} &= -kcq_{\boldsymbol{k}\mu} \ . \end{aligned} \tag{28}$$

As is easily verified, these equations of motion are reproduced by the canonical equations with the Hamiltonian equal to the field energy H_r, written as a function of the canonical variables $q_{\boldsymbol{k}\mu}$ and conjugate momenta $p_{\boldsymbol{k}\mu}$,

$$H_r = \sum_{\boldsymbol{k}\mu}\frac{1}{2}kc\left(p^2_{\boldsymbol{k}\mu} + q^2_{\boldsymbol{k}\mu}\right). \tag{29}$$

To make the transition to quantum mechanics, we introduce the canonical commutation relations

$$\left[q_{\boldsymbol{k}\mu}, p_{\boldsymbol{k}'\mu'}\right] = i\hbar\delta_{\boldsymbol{k}\boldsymbol{k}'}\delta_{\mu\mu'} \ . \tag{30}$$

Finally we define photon annihilation operators

$$a_{k\mu} = (2\hbar)^{-1/2}\left(q_{k\mu} + ip_{k\mu}\right) , \tag{31}$$

and in terms of these operators and their adjoints $a^+_{k\mu}$ the Hamiltonian takes the form

$$H_r = \sum_{k\mu} \hbar kc\left(a^+_{k\mu} a_{k\mu} + \frac{1}{2}\right) . \tag{32}$$

To describe the interaction of an atom with the radiation field, we also need the expression for the vector potential A in terms of creation and annihilation operators. From the relation $A^*_{-k} = A_k$ we find that A_k equals $\left(f_k + f^*_{-k}\right)/2$, and the equations (21), (25), and (31) then yield

$$\begin{aligned} A &= \frac{1}{\sqrt{V}} \sum_k \frac{1}{2}\left(f_k \ e^{ik\cdot r} + f^*_k \ e^{-ik\cdot r}\right) \\ &= \sum_{k\mu}\left(\frac{2\pi\hbar c}{Vk}\right)^{1/2}\left[a_{k\mu} \ e^{ik\cdot r} + a^+_{k\mu} \ e^{-ik\cdot r}\right]\varepsilon_{k\mu} \ . \end{aligned} \tag{33}$$

3.2 Dressed atoms

A canonical description of an electron moving in a potential $V(r)$ and interacting with the radiation field is based on the Hamiltonian

$$H = \frac{1}{2m}\left(p + \frac{e}{c}A(r,t)\right)^2 + V(r) + H_r \ , \tag{34}$$

where p denotes the canonical momentum of the particle, and where the Hamiltonian H_r for the radiation field is given in Eq. (29), and in quantized form in Eq. (32). Neglecting a very small term proportional to A^2, we may, in Coulomb gauge, rewrite the operator H for a one-electron system in the form

$$H = \frac{p^2}{2m} + V + H_r + \frac{e}{mc}\sum_{k\mu}\left(\frac{2\pi\hbar c}{Vk}\right)^{1/2}\left[a_{k\mu} \ e^{ik\cdot r} + a^+_{k\mu} \ e^{-ik\cdot r}\right]\varepsilon_{k\mu}\cdot p \ . \tag{35}$$

We now specialize to an atom interacting with a laser field in a single mode. At first, spontaneous emission is neglected such that only a single term need be retained in the summation in Eq. (35), and for brevity we drop the indices $k\mu$. The eigenstates of the Hamiltonian H are then linear combinations of product states of the type $|\, a > |\, n >$, where the first factor is an eigenstate of the atomic Hamiltonian $H_a = p^2/2m + V$, and where $|\, n >$ is the state describing the presence of n laser photons. Let the laser frequency be close to the transition frequency $(E_e - E_g)/\hbar$ in a closed atomic two-state system with upper and lower states $|\, e >$ and $|\, g >$. The energies of the coupled atom-laser system may then be obtained from diagonalization of the Hamiltonian H within the two-dimensional sub-spaces spanned by the nearly degenerate product states $|\, g > |\, n+1 >$ and $|\, e > |\, n >$, i.e., from diagonalization of the 2 x 2 matrices

$$\underline{\underline{H}}_n = \begin{Bmatrix} E_g + \hbar\omega\left(n+\frac{3}{2}\right) & M_n^* \\ M_n & E_e + \hbar\omega\left(n+\frac{1}{2}\right) \end{Bmatrix} . \tag{36}$$

Here the interaction matrix element M_n is given by the expression

$$M_n = \frac{e}{m}\left(\frac{2\pi\hbar}{V\omega}\right)^{1/2} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \; \boldsymbol{\varepsilon}\cdot < e|\boldsymbol{p}|g > \sqrt{n+1} \; , \tag{37}$$

where the vector $\boldsymbol{R}$ indicates the position of the atomic nucleus. In Eq. (35) the variation of the exponential over atomic distances has been neglected (dipole approximation). The commutator relation $[x, H_a] = i\hbar p_x/m$ leads to a simple connection between matrix elements of the momentum operator and of the position operator,

$$< e|\boldsymbol{p}|g >= im\omega_0 < e|\boldsymbol{r}|g > \; , \tag{38}$$

where ω_0 is the transition frequency, $\hbar\omega_0 = E_e - E_g$. Applying this equation in the expression for the off-diagonal element M_n, we may write the eigenvalues of the matrix $\underline{\underline{H}}_n$ in the form

$$E_n = \frac{1}{2}\left[E_g + \hbar\omega\left(n+\frac{3}{2}\right) + E_e + \hbar\omega\left(n+\frac{1}{2}\right)\right] \pm \frac{\hbar}{2}\sqrt{\delta^2 + \omega_R^2} \; . \tag{39}$$

The detuning δ is defined as in Eq. (13), with ω_{n0} replaced by ω_0, and the Rabi frequency ω_R is given by the formula

$$\hbar\omega_R = 2|\boldsymbol{\varepsilon}\cdot\boldsymbol{d}|\left(\frac{2\pi\hbar\omega}{V}\right)^{1/2}\sqrt{n+1} \; , \tag{40}$$

where the dipole matrix element $\boldsymbol{d}$ equals $-e < e \mid \boldsymbol{r} \mid g >$. In Eq. (40) no distinction is made between the frequencies ω and ω_0.

To make the connection to the Rabi frequency defined in Eq. (14), we first show that in the classical limit the single-mode laser field under consideration corresponds to a travelling wave of the type leading to Eq. (14). In fact, from the Eqs. (19), (33), (27), and (25) we find, omitting the fixed values of the indices $\boldsymbol{k}\mu$,

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{r},t) &= -V^{-1/2}k \;\; \mathrm{Im}\left\{f(0)e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-kct)}\right\} \\ &= -\boldsymbol{\varepsilon}\left(\frac{4\pi\omega}{V}\right)^{1/2}\left[p(0)\cos\left(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t\right) + q(0)\sin\left(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t\right)\right] . \end{aligned} \tag{41}$$

For the travelling wave discussed previously, $\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{\varepsilon}_x E_0 \cos(kz-\omega t)$, the energy density, averaged in space and time, equals $E_0^2/8\pi$. If this value is identified with the quantal result $n\hbar\omega/V$, corresponding to the presence of n photons, the Rabi frequencies introduced in the Eqs. (14) and (40) indeed become equal.

As expected, the dressed-atom picture may be developed for any kind of laser mode, and not only for travelling plane waves. The classical fields are then expanded in another orthonormal system than the complex exponentials in Eq. (21) but the transformation of the equations of

motion to canonical form with a Hamiltonian describing a set of uncoupled harmonic oscillators may be carried through with any set of orthonormal basis functions [3]. The dressed-atom energies are obtained through an appropriate replacement of the function $V^{-1/2}\exp(i\boldsymbol{k}\cdot\boldsymbol{R})$ in Eq. (37). In the case of a standing wave, the function is of the form $(2/V)^{1/2}\cos(\boldsymbol{k}\cdot\boldsymbol{R})$, and the right-hand side of Eq. (40) is multiplied by the factor $\sqrt{2}\cos(\boldsymbol{k}\cdot\boldsymbol{R})$. A classical field

$$\boldsymbol{E}(\boldsymbol{r},t)=\boldsymbol{\varepsilon}_x E_0\cos(\boldsymbol{k}\cdot\boldsymbol{r})\cos(\omega t) \tag{42}$$

corresponds to an average energy density $E_0^2/16\pi$, and it therefore leads to the Rabi frequency

$$\omega_R(\boldsymbol{R})=\hbar^{-1}|\boldsymbol{d}\cdot\boldsymbol{\varepsilon}_x|E_0\cos(\boldsymbol{k}\cdot\boldsymbol{R})\ . \tag{43}$$

Strictly speaking, the Rabi frequency depends on the number of photons present but in laser fields the average number $\bar{n}$ is very large, e.g., $\bar{n}\sim 10^{10}$, and the corresponding relative fluctuation $(\bar{n})^{-1/2}$ is therefore so small that this dependence may be neglected. For an atom in the standing wave (42), with wave vector $\boldsymbol{k}$ in the z direction, we then obtain the dressed-atom energies, setting the energy $E_g+\frac{1}{2}\hbar\omega$ equal to zero, [4]

$$\begin{aligned} E_{1n}&=(n+1)\hbar kc-\frac{1}{2}\hbar\delta+\frac{1}{2}\hbar\Omega(Z)\ ,\\ E_{2n}&=(n+1)\hbar kc-\frac{1}{2}\hbar\delta-\frac{1}{2}\hbar\Omega(Z)\ . \end{aligned} \tag{44}$$

Here, Z denotes the z coordinate of the atomic position, and $\Omega(Z)$ the generalized Rabi frequency

$$\Omega(Z)=\left(\delta^2+|\boldsymbol{d}\cdot\boldsymbol{\varepsilon}_x|^2E_0^2\cos^2(kZ)/\hbar^2\right)^{1/2}\ . \tag{45}$$

With an appropriate choice of the relative phase of the atomic states $|e>$ and $|g>$, the off-diagonal elements in the matrix $\underline{\underline{H}}_n$ in Eq. (36) become real with the value $M_n=\hbar\omega_R(Z)/2$, and the eigenstates of $\underline{\underline{H}}_n$ are then given by the expressions

$$\begin{aligned} |1,n;Z>&=C(Z)|e>|n>+S(Z)|g>|n+1>\ ,\\ |2,n;Z\rangle&=-S(Z)|e>|n>+C(Z)|g>|n+1>\ , \end{aligned} \tag{46}$$

with

$$\begin{aligned} C(Z)&=\left\{[1-\delta/\Omega(Z)]/2\right\}^{1/2}\ ,\\ S(Z)&=\left\{[1+\delta/\Omega(Z)]/2\right\}^{1/2}\ . \end{aligned} \tag{47}$$

4. COOLING IN STANDING WAVE

4.1 Application of the dressed-atom picture

The dipole force acting on an atom in a standing laser field has a very simple interpretation within the dressed-atom picture [4]. The energies E_{1n} and E_{2n} in Eq. (44) depend on atomic position, and therefore the atom experiences a force in the z direction

$$F_{in} = (-1)^i \frac{1}{2} \hbar \partial \Omega / \partial Z \ , \quad i = 1,2 \ . \tag{48}$$

The sign, but not the magnitude of the force, depends on whether the system is in the upper state 1 or in the lower state 2. So far, we have neglected spontaneous emission. The transition amplitude for such an event is proportional to a dipole matrix element $< j, n'; Z | \boldsymbol{r} | i, n; Z >$, and since the process always involves an atomic transition $| e > \rightarrow | g >$ and does not affect the number of laser photons present, a dressed state $|i, n; Z >$ can only decay to a state $|j, n\text{-}1; Z >$. The level indices i and j can differ or be equal, and from the expressions (46) we find for the corresponding transition rates Γ_{ji}

$$\begin{aligned} \Gamma_{21} &= \Gamma C^4(Z) \ , \\ \Gamma_{12} &= \Gamma S^4(Z) \ , \end{aligned} \tag{49}$$

where $\hbar\Gamma$ denotes the line width of the upper atomic state. In statistical equilibrium, the populations Π_1 and Π_2 of the upper and lower levels are determined by the relation $\Pi_1\Gamma_{21} = \Pi_2\Gamma_{12}$, and one obtains

$$\Pi_1 = 1 - \Pi_2 = S^4(Z) / \left(C^4(Z) + S^4(Z) \right) \ . \tag{50}$$

The implicit assumption that populations do not couple to off-diagonal elements of the density matrix is valid for generalized Rabi frequencies $\Omega(Z)$ large compared to Γ [5].

For positive detuning, $\delta > 0$, the ratio S/C is larger than unity, according to the expressions (47), and levels of type 1 are more heavily populated than levels of type 2. As a result, the average force obtained from the equation (48) is in the direction away from regions of high electric field intensity, in agreement with Eq. (17). The average force is given by the formulas

$$F = -\frac{1}{2} \hbar \frac{\partial \Omega}{\partial Z} (\Pi_1 - \Pi_2) = -\frac{1}{2} \hbar \delta \frac{\partial \Omega^2 / \partial Z}{\Omega^2 + \delta^2} \ . \tag{51}$$

In the weak field limit, $\omega_R(0) = eE_0|x_{n0}| / \hbar < | \delta |$, Eq. (51) reduces to Eq. (17), with the oscillator strength f_{n0} given by Eq. (8), when we insert the condition $| \delta | > \Gamma_n$. In the opposite limit of strong fields, $|\omega_R(Z)| > | \delta |$, the force saturates at the value $\frac{1}{2} \hbar k \delta \ \sin(2kZ) / \cos^2(kZ)$.

If the atom moves infinitely slowly in the z direction, local statistical equilibrium results at every position. According to the Eqs. (45) and (51), the force F then becomes antisymmetric around the maxima and minima in the electric field energy, and the average value over half a wavelength of the field vanishes. At finite particle velocities v, however, the system does not have time to relax to local equilibrium, and the populations $\Pi_i(Z, v)$ must be determined from the differential equation

$$\partial / \partial(Z / v)\begin{Bmatrix}\Pi_1\\ \Pi_2\end{Bmatrix}=\begin{Bmatrix}-\Gamma_{21} & \Gamma_{12}\\ \Gamma_{21} & -\Gamma_{12}\end{Bmatrix}\begin{Bmatrix}\Pi_1\\ \Pi_2\end{Bmatrix}. \tag{52}$$

The symmetry with respect to the extrema of the electric field energy is now broken, and an average force results. If the wavelength of the field is long compared to the distance travelled by the projectile during the lifetime Γ^{-1} of the upper atomic state, i.e., $v < \Gamma/k$, the populations only differ from the equilibrium values in Eq. (50) by small terms proportional to v. The average force then also becomes proportional to velocity, and it is straightforward to show that it is retarding (accelerating) for blue (red) detuning.

The spontaneous events determine the populations of the dressed states and thereby the average force. The transitions do not in themselves lead to extraction of kinetic energy from the atom since the energy of an emitted photon is balanced by the decrease in dressed-state energy from E_{in} to $E_{j(n-1)}$. At sufficiently high velocities, nonradiative transitions between dressed states, corresponding to the same value of n, also become important. Such a process results in changes in both populations and in kinetic energy. These non-adiabatic transitions will be briefly discussed in Section 4.2.

As regards the direction of the average dipole force on a slowly moving atom, we shall give another argument which is based on the simple pictorial representation of the dressed-atom description [6]. In Fig. 1, the energy levels in Eq. (44) are sketched as functions of position for consecutive values of n. When spontaneous emission is neglected, the system evolves along a single curve. The energy required to climb a hill is taken from the kinetic energy of the atom, but this kinetic energy is fully regained on the way down the other side of the hill. Apart from modulations, the particle therefore maintains its velocity as it travels through the standing wave.

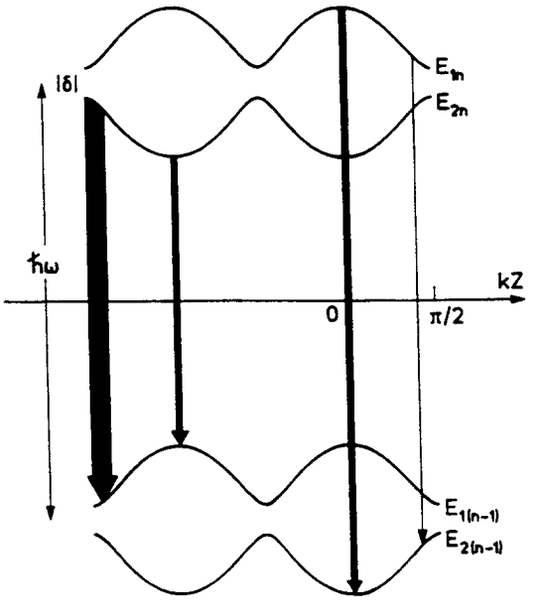


Fig. 1 Two sets of dressed-state energies as functions of position. The relative values indicated for the transition rates Γ_{21} and Γ_{12} at various positions correspond to a positive detuning δ, in the limit $\omega_R(0) >> \delta$.

This situation is changed when spontaneous processes are introduced. A decay from a level E_{in} to a level $E_{i(n-1)}$ does not affect the force but a change of level type, from 1 to 2 or vice

versa, leads to a reversal of its direction. Let us consider a positive detuning, $\delta > 0$. As is observed from the expressions (46) and (47), a system travelling along the upper level E_{1n} is in the pure product state $|g>|n+1>$ at the field nodes where Ω equals δ, and therefore no decay can occur at these positions. The probability of decaying to the energy level $E_{2(n-1)}$ is proportional to the quantity $C^4(Z)$ which attains its maximum at the hill tops. Here, a decay to a level of type 2 brings the system from an energy maximum to an energy minimum. This does not only prevent the atom from rolling down the back-side of the hill just climbed and thereby from recovering its kinetic energy. It has to climb yet another hill, to boot! The fate of the particle is analogous to the punishment of Sisyphus [6]. For an atom following a level of type 2, the situation is similar: When the probability for a transition to a level of type 1 is a maximum, now at field nodes, the decay brings the system from a hill top to a valley bottom. The repeated transitions, predominantly from higher to lower positions on the energy curves, are compensated by a gradual decrease in kinetic energy of the atom.

A change of sign on the detuning δ leads to an interchange of the functions C and S in Eq. (47). For negative detuning, the spontaneous events therefore, in the main, bring the system upwards on the energy curves, and the atom is accelerated.

In the high velocity limit, $v > \Gamma/k$, the system on the average follows an energy curve over a large number of wavelengths of the laser field before decaying. This number is proportional to v, and the average force, equal to the energy change per unit path length, therefore becomes inversely proportional to v. Since the average force was found to be velocity proportional in the opposite limit $v < \Gamma/k$, we expect a maximum at velocities of the order of Γ/k. This expectation is borne out by the results in Fig. 2 which shows the magnitude of the average dipole force,

$$F(v) = -\frac{\hbar}{2} < (\Pi_1(Z,v) - \Pi_2(Z,v))\partial\Omega / \partial Z > \varepsilon_z \ . \tag{53}$$

Here, the brackets indicate an average over position. The populations were obtained from numerical solution of Eq. (52) under steady state conditions, i.e., corresponding to populations which have attained the periodicity of the field energy. With the parameter values applied, $\omega_R(0)/\Gamma = 1000$ and $|\delta|/\Gamma = 200$, the maximum average force equals 33 $\hbar k\Gamma$, which may be compared with the maximum scattering force $\hbar k\Gamma/2$ in a travelling wave. At a fixed value of the ratio $\delta/\omega_R(0)$, the dipole force is proportional to the Rabi frequency $\omega_R(0)$ in the high velocity region, $v > \Gamma/k$.

Figure 3 shows the result of a Monte Carlo simulation of longitudinal cooling of a velocity distribution with a similar width as in an experiment (Ref. [12]) to be discussed below [7]. The dressed atom was allowed to jump back and forth between upper and lower energy levels according to the position dependent transition probabilities. The detuning and the strength of the standing laser field had similar values as in Fig. 2, and the interaction time was 17.2 μs, corresponding to the passage of a neon atom with an energy of 200 eV through a distance of 75 cm. In the experiment, linearly polarized light induced transitions between lower and upper atomic states with angular momenta $J_g = 2$ and $J_e = 3$ and it was therefore necessary to generalize the dressed-atom description from a two-level system to a multilevel system, in which the different members of a Zeemann multiplet give rise to different Rabi frequencies. The field strength applied in the calculation corresponded to a standing wave of cross sectional area $4.2 \cdot 10^{-3}$ cm^2, maintained in a resonator by means of two 0.5 W cw lasers. The loss per round trip for a light beam in the resonator was assumed to be 5%. From Fig. 3, very significant cooling is expected under these conditions.

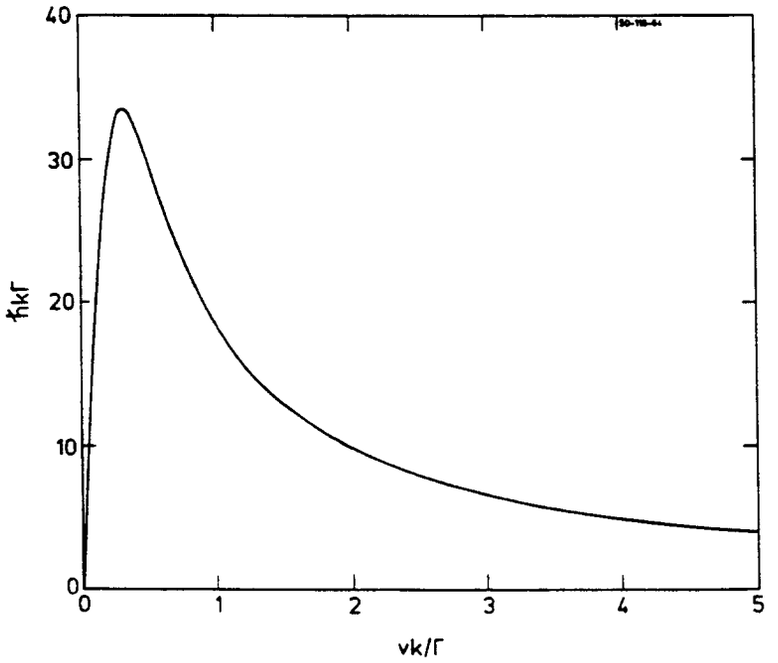


Fig. 2 Absolute value of the average dipole force, in units of $\hbar k\Gamma$, as a function of velocity, in units of Γ/k, at a Rabi frequency $\omega_R(0) = 1000\Gamma$ and a detuning $|\delta| = 200\Gamma$. The force is retarding for $\delta > 0$ and accelerating for $\delta < 0$.

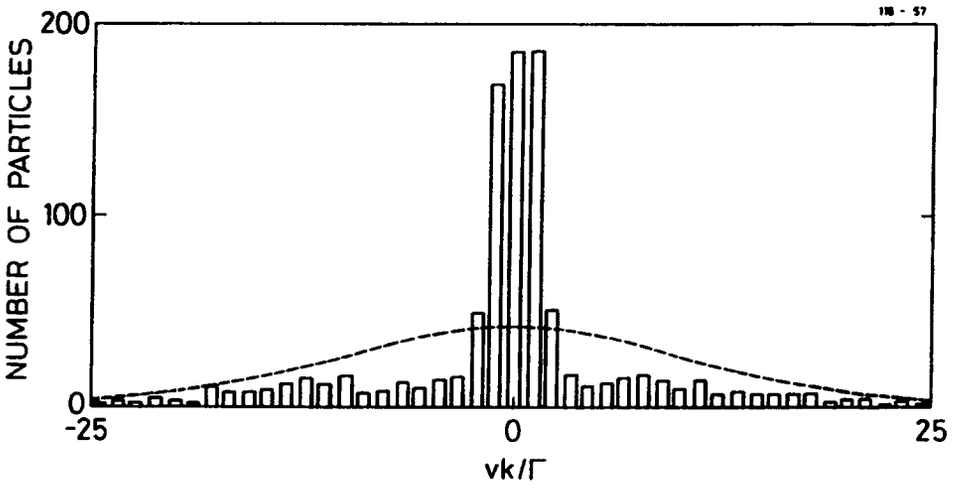


Fig. 3 Monte Carlo calculation of the transformation of an initial Gaussian distribution in atomic velocity as a result of interaction with a standing laser field during a time $886\Gamma^{-1}$. Linearly polarized light induced transitions between lower and upper atomic states with angular momenta $J_g = 2$ and $J_e = 3$. Quite similar results would be obtained for a two-level system with Rabi frequency $\omega_R(0) = 800\Gamma$ and a positive detuning $\delta = 200\Gamma$. In the experiment in Ref. [12] the velocity Γ/k equals 5.3 m/sec.

4.2 Nonadiabatic transitions

Spontaneous emission leads to transitions between dressed-atom states $|i, n; Z>$ and $|j, n\text{-}1; Z>$. At finite atomic velocity, transitions can also occur between the different dressed states belonging to the same value of n. These nonadiabatic processes are analogous to Landau-Zener transitions between molecular states applied in the treatment of colliding atoms.

The dressed system is described by a Hamiltonian of the type shown in Eq. (35), with only a single value for the indices $k\mu$ and with the electron coordinates inside the square brackets replaced by the atomic position, as in Eq. (37). When the atom moves along a classical trajectory, the Hamiltonian becomes a function of time. To solve the corresponding time-dependent Schrödinger equation, it is convenient to expand the wave function $\psi(t)$ for the dressed system in the eigenstates $|i, n; t>$ of the total Hamiltonian, corresponding to the atomic position at time t. The transitions to be discussed do not change the quantum number n, and we may therefore limit ourselves to functions of the form

$$\psi(t) = \sum_{j=1}^{2} a_j(t)\; e^{-i\int_0^t E_j(t')dt'/\hbar}\; |j;\; t>\;, \tag{54}$$

where the explicit reference to the fixed value n has been omitted in the symbols for the dressed states and the corresponding energies. For simplicity, we again consider a two-state atom. Inserting Eq. (54) into the time-dependent Schrödinger equation, we find for the expansion coefficients $a_k(t)$,

$$\frac{da_k(t)}{dt} = -\sum_{j=1}^{2} < k;t\left|\frac{\partial}{\partial t}\right|j;t>a_j(t)\; e^{-i\int_0^t \omega_{jk}(t')dt'}\;, \tag{55}$$

where the frequency $\omega_{12}(t) = (E_1(t) - E_2(t)) / \hbar$ equals the generalized Rabi frequency $\Omega(t)$. Both this quantity and the matrix element $< k;\, t \mid \partial/\partial t \mid j;\, t >$ are essentially unchanged during a time interval Δt short compared to the time $\sim(kv)^{-1}$ required for travelling a quarter of a wavelength. If the product $\Omega\Delta t$ is large compared to unity, we then obtain for the probability of finding the system in state j at time $t+\Delta t$, provided it was in state i at time t [4]

$$\left|a_j(t+\Delta t)\right|^2 \cong 2\left|< j;t\left|\frac{\partial}{\partial t}\right|i;t>\right|^2 / \Omega^2(t)\;, \quad \Delta t < (kv)^{-1}\;. \tag{56}$$

Here, a factor $\sin^2(\Omega\Delta t/2)$ has been replaced by its average value 1/2. At very high velocities, $kv > \Omega$, the exponential in Eq. (55) may be approximated by unity, and the formula (56) represents an upper limit since the time Ω^{-1} should be replaced by the shorter time interval Δt. To evaluate the time derivative of a dressed state in Eq. (56), we first note that according to their definitions in Eq. (47), the expansion coefficients C and S in Eq. (46) may be written in the form [4]

$$\begin{aligned} C &= \cos\Theta \quad \text{and} \\ S &= \sin\Theta\;, \end{aligned} \tag{57}$$

where the angle Θ is given by the expression

$$\cos 2\Theta = -\delta / \Omega\;. \tag{58}$$

The dressed states therefore fulfil the equations

$$\frac{\partial}{\partial t}|1;t>=|2;t>\frac{d\Theta}{dt} \quad \text{and}$$
$$\frac{\partial}{\partial t}|2;t>=-|1;t>\frac{d\Theta}{dt} \; . \tag{59}$$

The derivative $d\Theta/dt$ is determined through differentiation of Eq. (58). The generalized Rabi frequency is of the form

$$\Omega(t)=\left(\delta^2+\omega_R^2(t)\right)^{1/2}=\left(\delta^2+\omega_R^2(0)\cos^2(kvt)\right)^{1/2} \; , \tag{60}$$

and we find for the probability in Eq. (56),

$$\left|a_j(t+\Delta t)\right|^2 \cong \frac{1}{2}\left(\frac{\delta}{\Omega^3(t)}\frac{d\omega_R}{dt}\right)^2=\frac{1}{2}(\delta\omega_R(0)kv)^2\frac{\sin^2(kvt)}{\left(\delta^2+\omega_R^2(0)\cos^2(kvt)\right)^3} \; . \tag{61}$$

As expected, the probability for a nonadiabatic transition is largest around the electric field nodes where the distance between the dressed energy levels is a minimum. The condition for neglecting the nonadiabatic transitions in the evaluation of the average dipole force is that the maximum value in Eq. (61) be small compared to the probability for spontaneous emission during the passage of half a wavelength of the laser field. Since this probability is of the order of $\Gamma\lambda/(2v)$, the velocity of the atom should be lower than the critical value [4]

$$v_c=\left(2\pi\delta^4/(\Gamma\omega_R(0))^2\right)^{1/3}\Gamma/k \; . \tag{62}$$

Nonadiabatic transitions become more important for decreasing minimum distance $\hbar|\delta$ between energy levels of type 1 and type 2 and for increasing rate of change of the dressed states $|j;t>$, i.e., for increasing value of the ratio $\omega_R(0)/|\delta|$. In the examples shown in Figs. 2 and 3, the velocity v_c is ~20 Γ/k.

When nonadiabatic transitions must be included, together with spontaneous events, the evaluation of the average force is normally based on the solution of the so-called Optical Bloch Equations which govern the time evolution of the density matrix connected with the atomic states $|g>$ and $|e>$, and not with the dressed states [8]. However, also at these higher velocities, the force may be obtained within the dressed-atom picture [9], [10]. The populations and the off-diagonal elements of the density matrix in the basis of dressed-atom states are now coupled, and the interpretation of the average force becomes somewhat less straightforward than in the adiabatic case. We therefore only indicate the results. From Fig. 4 it is observed that as a function of velocity, the average force changes sign after the passage of a transition region in the vicinity of v_c with strong and rapid oscillations. These features result from interference between amplitudes for nonadiabatic transitions at successive field nodes. Beyond the critical velocity v_c, the force again becomes quite appreciable, provided the Rabi frequency exceeds the detuning, corresponding to saturation of the atomic transition. In fact, in the entire velocity range $v_c \lesssim v \lesssim |\delta|/k$ it is stronger than the saturated scattering force, and it is given by the expression $\frac{1}{2}\hbar k\Gamma|\delta|/(kv)$, apart from modulations.

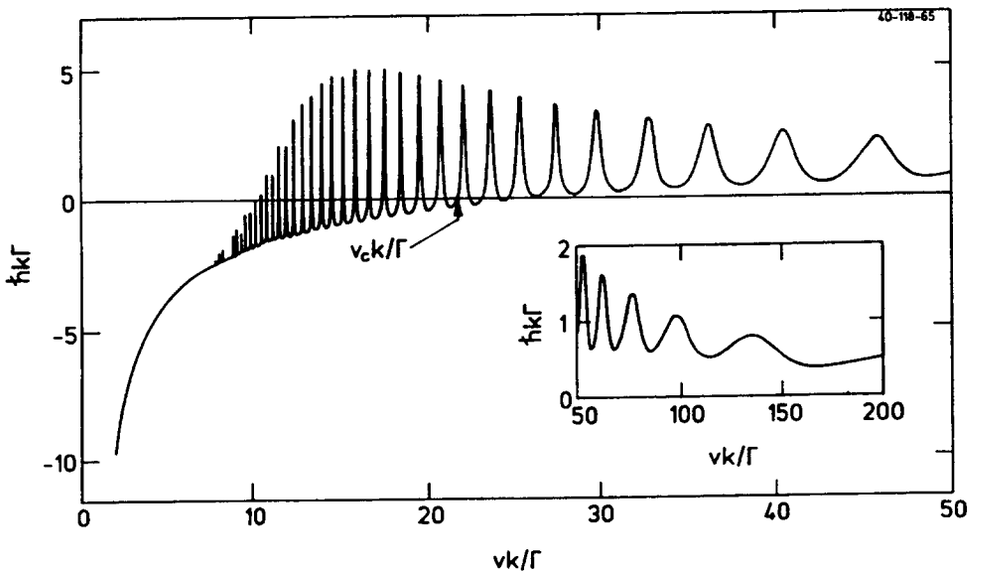


Fig. 4 Average force, in units of $\hbar k\Gamma$, as a function of velocity, in units of Γ/k, at a Rabi frequency $\omega_R(0) = 1000\Gamma$ and a positive detuning $\delta = 200\Gamma$. The critical velocity v_c is given by Eq. (62).

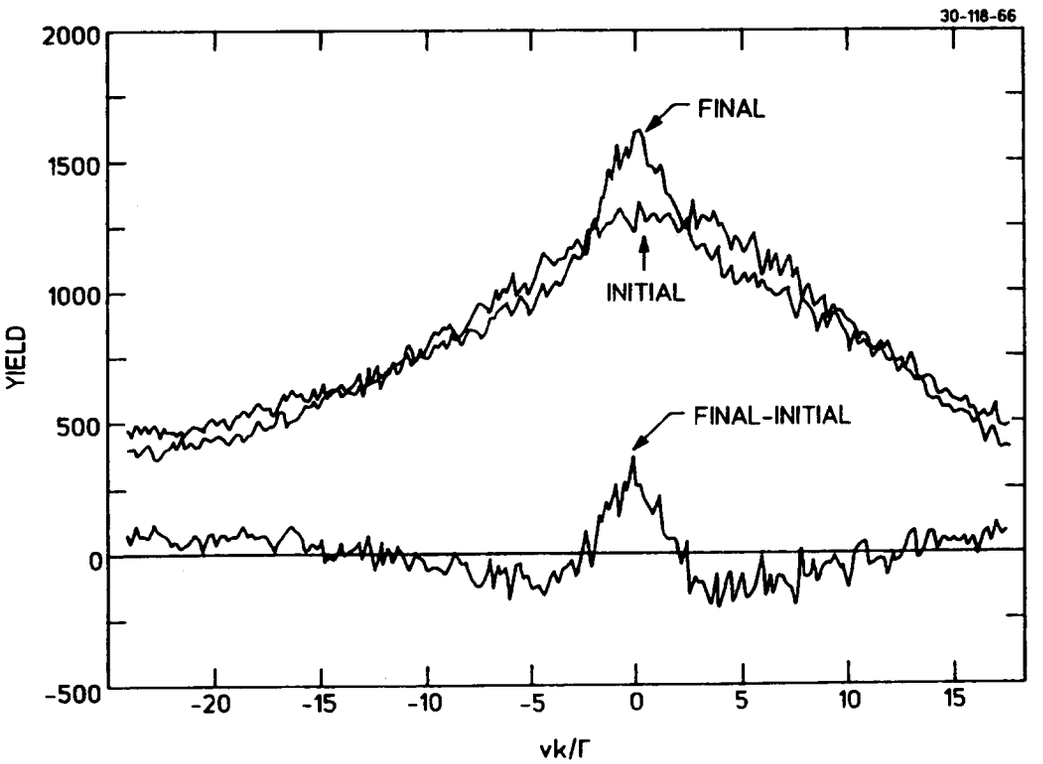


Fig. 5 Measured longitudinal velocity distribution for 1 keV neon atoms before and after interaction with a standing laser field during 6.5 μsec = $330\Gamma^{-1}$. The atomic states have angular momenta $J_g = 2$ and $J_e = 3$ and the light is linearly polarized and has a positive detuning of 60Γ. A closely equivalent two-level system has a Rabi frequency of $\omega_R(0) = 275\Gamma$. The velocity $v_c/2$ equals 5.1 Γ/k, where $\Gamma/k = 5.3$ m/sec. The distributions are inferred from the Doppler shift of a transition induced by a probe laser, and the resolution is $\sim \pm 2\Gamma/k$. The yield refers to spontaneous emission following the excitation. Published with kind permission of the author of Ref. [12].

4.3 Experiments

The change in direction of the force in a strong standing wave above the velocity v_c has been verified in two very recent experiments. The first measurement [11] was on a thermal beam of sodium atoms with an average velocity of ~500 m/sec and a velocity spread of $\sim \pm 200$ m/sec. The field was standing in the laboratory frame and virtually all of the particles had an initial velocity higher than v_c. The detuning was negative, and after penetration of the wave over a distance of 27 cm, the width of the distribution in longitudinal velocity had, in one case, diminished to 12 m/sec (FWHM). The laser power applied was 0.2 Watt.

In the second experiment one was directly concerned with the problems associated with laser cooling of a fast beam [12]. The laser field must now be standing in a frame moving with the average particle velocity. To obtain significant velocity changes in a single-pass experiment with an interaction time two orders of magnitude shorter then in Ref. [11], it was necessary to increase the energy density in the laser wave by means of a resonator. In the neon beam applied, the longitudinal velocity spread was on the order of the critical velocity v_c, owing to the velocity compression upon acceleration to an energy of 1 keV. In Fig. 5 are shown the velocity distributions before and after interaction with the field, as well as the difference between the two. Since the detuning is now positive, low and high velocity particles are forced towards the centre and towards the tails, respectively. The changes observed are in qualitative agreement with estimates based on a transformation of the initial distribution according to the calculated velocity-dependent average force. The intention is to carry the experiment to somewhat higher field strengths for a constant value of the ratio $\delta/\omega_R(0)$. For a positive detuning, this will lead to longitudinal cooling of essentially the entire beam, as in the calculation presented in Fig. 3.

5. SUMMARY

For atoms or ions in a beam, we have discussed the force from a strong collinear laser beam which forms a standing wave in a reference frame moving with the average velocity of the particles. In an experiment, the standing wave in the moving frame is obtained as the superposition of two travelling waves which are propagating in opposite directions and which have different frequencies in the laboratory.

When the Rabi frequency $\omega_R(0)$ becomes large compared to the detuning $|\delta|$, $\omega_R(0) > |\delta| > \Gamma$, the laser intensity is sufficient to saturate the atomic transition under consideration. In these circumstances, particles with velocities relative to the standing wave in the range $0 < v \lesssim |\delta|/k$ experience a force which is stronger than the maximum scattering force from a travelling wave. As an example, we may choose the detuning δ such that the upper velocity limit $|\delta|/k$ equals 1000 m/sec. In a typical case with a photon energy $\hbar\omega = 2$ eV and an atomic lifetime $\Gamma^{-1} = 10^{-8}$ sec, saturation is then obtained when the two travelling waves composing the standing wave have intensities of the order of 1 Watt/mm^2.

The average force changes sign with the detuning, and for given values of the detuning and the field strength, the force directions are opposite in the two velocity regimes below and above a complicated transition region around the critical velocity v_c, given in Eq. (62). For red detuning, the force is retarding above v_c which is of the order of 100 m/sec in the example considered above, provided the saturation is not too strong. Apart from modulations, the force equals $\frac{1}{2}\hbar\Gamma|\delta|/v$ in the velocity region $v_c < v \lesssim |\delta|/k$, and the time required for the average force to bring an atom of mass M from one end of the velocity region to the other is therefore approximately equal to $M|\delta|/(\hbar\Gamma k^2)$. In the example, we obtain a time of 0.3 msec for atomic

mass 20. After the change of direction around v_c, the average force again becomes inversely proportional to velocity, and it increases linearly with the Rabi frequency when the ratio $\delta/\omega_R(0)$ remains constant. As is observed from Fig. 2, the maximum force, attained at a velocity less than Γ/k, becomes quite strong for realistic field strengths.

In conclusion it would appear that for ions circulating in a storage ring, a strong standing laser field might be useful in a second cooling stage when a very accurate definition of longitudinal velocity is desired. First, electron cooling and/or stochastic cooling should reduce the velocity spread to a value of the order of a thousand meters per second. Then a millisecond of interaction with a laser field with red detuning could transform the velocity distribution into two peaks around the velocities $\pm v_c$ with respect to the standing wave. Finally, a change of sign of the detuning together with a moderate increase in its absolute magnitude would bring the particles into the low velocity region with a strong dipole force, and a rapid cooling to essentially vanishing velocity spread would result.

ACKNOWLEDGEMENTS

I gratefully acknowledge help and constructive criticism from J.U. Andersen, P. Jessen, K. Mølmer, and O. Poulsen.

REFERENCES

[1] For instance, J. Slater: Quantum Theory of Matter. McGraw-Hill, (1975), Chapter 13.

[2] Reproduced from lecture notes by P. Kristensen, Aarhus University, 1981.

[3] A. Yariv: Quantum Electronics. John Wiley, (1975), p. 90.

[4] J. Dalibard and C. Cohen-Tannoudji, J. Opt. Soc. Am. B 2 , (1985), 1707.

[5] C. Cohen-Tannoudji, in: Frontiers in Laser Spectroscopy, Vol. 1 (Les Houches, Session XXVII, 1975) North-Holland (1977), Ed. by R. Balian, S. Haroche, and S. Liberman.

[6] A. Aspect, J. Dalibard, A. Heidmann, C. Salomon, and C. Cohen-Tannoudji, Phys. Rev. Lett. 57 , (1986), 1688.

[7] E. Bonderup and K. Mølmer, J. Opt. Soc. Am. B 6 , (1989), 2125.

[8] For instance, J. Dalibard, Thèse, Université de Paris, 1986.

[9] A.P. Kazantsev, V.S. Smirnov, G.I. Surdutovich, D.O. Chudesnikov and V.P. Yakovlev, J. Opt. Soc. Am. B 2 , (1985), 1731.

[10] K. Mølmer, Physica Scripta 45 , (1992), 246.

[11] M. Prentiss and A. Cable, Phys. Rev. Lett. 62 , (1989), 1354.

[12] P. Jessen, Thesis, Aarhus University, 1989.

SCHOTTKY NOISE AND BEAM TRANSFER FUNCTION DIAGNOSTICS

D. Boussard
CERN, Geneva, Switzerland

Abstract
Following the analysis of Schottky signals for the unbunched and the bunched beam cases, a general study of electromagnetic detectors is presented. Here the image-current approach and the Lorentz reciprocity theorem will be used to evaluate the detector (or pick-up) performance for several typical examples. Then, signal-processing techniques, which play an important role in the Schottky signal analysis, will be reviewed. The beam transfer function which relates the beam response to an external excitation also provides very useful information about the accelerator behaviour. It requires an element to excite the beam (kicker) which will be shown to be equivalent to a detector working in reverse. With beam-transfer-function measurements an assessment of beam stability limits can be made, leading to the determination of the overall ring impedance.

1. INTRODUCTION

The noise generated in an old fashioned electron tube is governed by the Schottky formula which simply reflects the fact that the anode current is composed of individual electrons randomly emitted by the cathode. Very similarly, the beam current in a circular particle accelerator, also exhibits a random component, called the Schottky noise, which results from the large, but finite, number of particles in the beam. In the absence of random quantum emissions (i.e. for hadron machines) the analysis of Schottky noise signals (or Schottky signals, for brevity) is a very powerful tool to study the accelerator behaviour. Historically, Schottky signals have been observed first on unbunched beam machines (CERN ISR) [1,2], leading to the development of the very successful stochastic cooling technique. For bunched beams, the presence of strong "macroscopic" beam signals renders the observation of the tiny Schottky signals more difficult. However improved signal processing techniques have recently made their observation possible.

Following the analysis of Schottky signals for the unbunched and the bunched beam cases, a general study of electromagnetic detectors is presented. Here the image-current approach and the Lorentz reciprocity theorem will be used to evaluate the detector (or pick-up) performance for several typical examples. Then, signal-processing techniques, which play an important role in the Schottky signal analysis, will be reviewed.

The beam transfer function which relates the beam response to an external excitation also provides very useful information about the accelerator behaviour. It requires an element to excite the beam (kicker) which will be shown to be equivalent to a detector working in reverse. With beam-transfer-function measurements an assessment of beam stability limits can be made, leading to the determination of the overall ring impedance.

2. SCHOTTKY SIGNALS

2.1 Unbunched beam, longitudinal

For a single particle circulating in the machine (charge e, revolution period $T_i = 1/f_i$), the beam current, at a given location in the ring, is composed of an infinite train of delta pulses (Fig. 1a) separated in time by T_i. In frequency domain, this periodic waveform is represented by a line spectrum (Fig. 1b), the distance between lines being $f_i = \omega_i/2\pi$.

$$i_i(t) = ef_i \sum_{n=-\infty}^{+\infty} \exp\ jn\omega_i t \ . \tag{1}$$

Looking at positive frequencies only:

$$i_i(t) = ef_i + 2ef_i \sum_{n=1}^{\infty} \cos\ n\omega_i t \ . \tag{2}$$

The first term represents the DC component, the others are simply the successive harmonics of the revolution frequency.

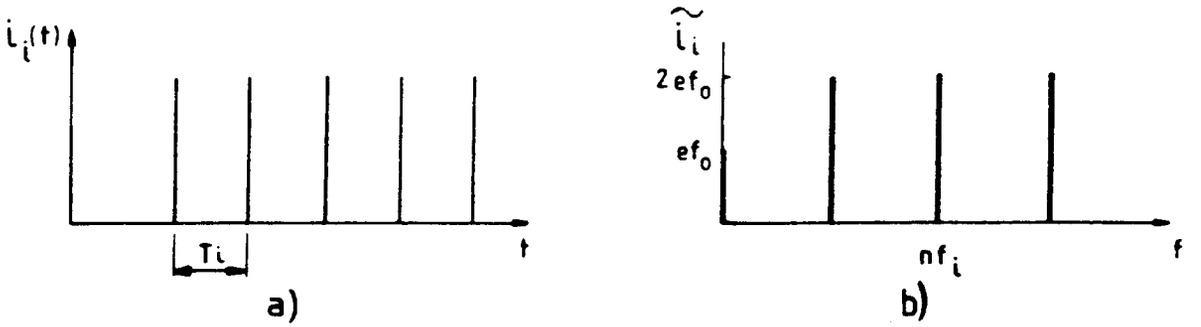


Fig. 1 a) Time domain δ pulses b) Frequency domain: line spectrum

For N particles, randomly distributed in azimuth along the ring circumference (debunched beam case) and having slightly different ω_i, each line at frequency nf_i which is infinitely narrow in the case of a single particle, will be replaced by a band of frequencies (Schottky band) whose width is simply:

$$\Delta f = n\ \Delta f_i = n\ f_0 \eta \frac{\Delta p}{p} \ , \tag{3}$$

Δf_i is the spread in particle's revolution frequencies resulting from the relative momentum spread $\Delta p/p$ and the machine parameter $\eta = \left(1/\gamma_t^2 - 1/\gamma^2\right)$. f_0 is the average revolution frequency.

When averaging equation (2) over N particles, only the DC terms remain ($i_{DC} = Nef_0$), the other components cancel due to the random azimuth phase factor. However, the r.m.s. current per band which is given by the sum:

$$\langle i^2 \rangle = \left[2ef_0\left(\cos\ \theta_1 + \cos\ \theta_2 + \ \ldots\ + \cos\ \theta_1 + \ \ldots\ + \cos\theta_N\right)\right]^2 \tag{4}$$

does not vanish because of the cos θ^2 terms. One obtains:

$$i_{rms} = \sqrt{\langle i^2 \rangle} = 2ef_0\sqrt{N\langle \cos\ \theta_i \rangle^2} \tag{5}$$

$$i_{rms} = 2ef_0\sqrt{\frac{N}{2}} \ . \tag{6}$$

The r.m.s. current per band (Schottky current) is independent of n (harmonic number) and proportional to the square root of the number of particles N.

As indicated on Fig. 2, the power spectral density, proportional to $\langle i^2 \rangle / \Delta f$, decreases with n until overlap occurs $(\Delta f > f_0)$. For a given band the local power density is obviously proportional to the number of particles per unit frequency. If the parameter η is known (η may be frequency dependent), the measurement of the power spectral density, in one particular Schottky band gives directly the $\Delta p/p$ distribution of the beam.

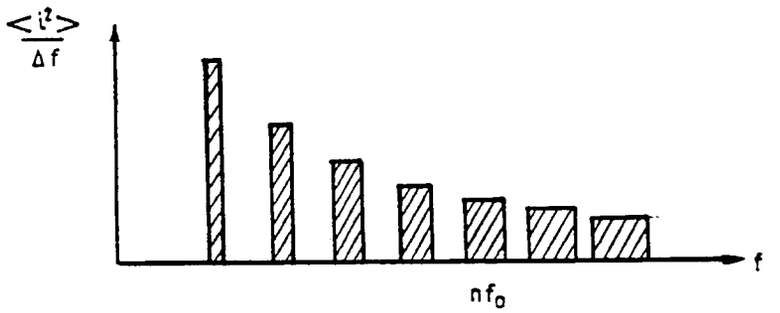


Fig. 2 Power spectral density of Schottky lines with increasing n

This forms the basis of $\Delta p/p$ beam distribution measurements in DC coasting machines (cooling and accumulation rings in particular).

Note that the noise signals pertaining to successive Schottky bands are not correlated because the random azimuthal phase factor is multiplied by n in Eq. (4).

2.2 Unbunched beam, transverse

For a single particle, the beam current $i_i(t)$ must be replaced by the dipole moment: $d_i(t) = a_i(t).i_i(t)$, where $a_i(t)$ is the transverse displacement. The i^{th} particle executes a sinusoidal betatron oscillation, of amplitude a_i, which can be written:

$$a_i(t) = a_i\ \cos\left(q_i\omega_i t + \varphi_i\right) \ . \tag{7}$$

Here $q_i f_i$ is the observed frequency, at a fixed location in the ring, q_i being the non-integer part of the betatron tune (Fig. 3a).

In frequency domain:

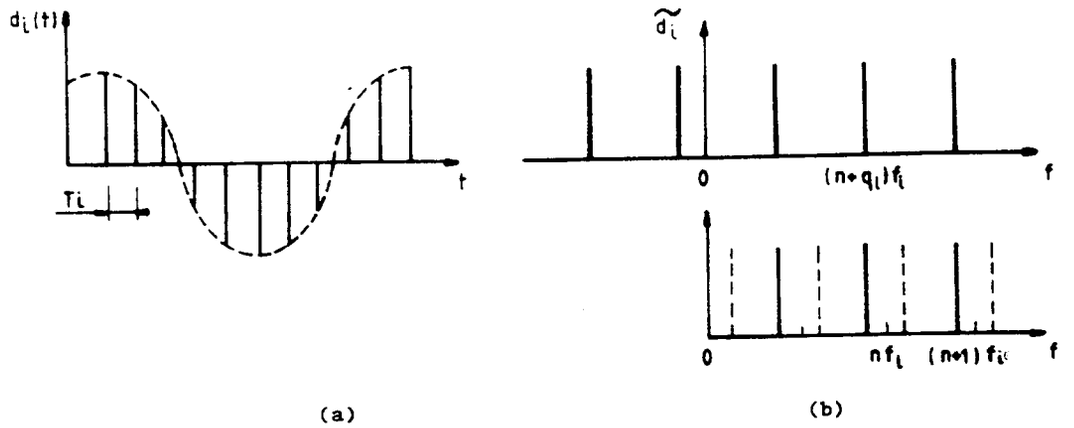


Fig. 3 Time (a) and frequency (b) domain representations of a single-particle transverse oscillation

$$d_i(t) = a_i \cos\left(q_i\omega_i t + \varphi_i\right) ef_0 \sum_{n=-\infty}^{+\infty} \exp\, jn\omega_i t \tag{8}$$

$$d_i(t) = a_i \; ef_0 \mathrm{Re}\left\{\sum_{n=-\infty}^{+\infty} \exp\, j\left[(n+q_i)\omega_i t + \varphi_i\right]\right\} . \tag{9}$$

The spectrum is again a series of lines spaced by the revolution frequency of the i^{th} particle, but shifted in frequency by $q_i f_i$. Looking at positive frequencies only (Fig. 3b) one obtains two betatron lines per revolution frequency band as in the case of an amplitude-modulated carrier which exhibits two symmetrical sidebands.

For N particles in the beam, again randomly distributed in azimuth and in betatron phases, averaging equation (9), for a given value of $n + q$, gives:

$$\langle d\rangle = 0 \; ; \quad \langle d^2\rangle = \langle a_i^2\rangle e^2 f_0^2 \frac{N}{2} \tag{10}$$

$$d_{rms} = ef_0 a_{rms}\sqrt{\frac{N}{2}} \; . \tag{11}$$

Again, the total power per Schottky band is independent of its location in the frequency spectrum; it is proportional to the number of particles in the beam and to the square of the r.m.s. oscillation amplitude.

Each Schottky band has now a finite width which results from the spread of revolution frequencies $\Delta f_i / f_0 = \eta \; \Delta p / p$ and from the spread of betatron frequencies Δq_i. The latter usually comes from the machine chromaticity ξ: $\Delta q_i = Q\xi\Delta p / p$, but may also result from space charge, beam-beam or nonlinear effects.

The line width of two adjacent Schottky bands $(n \pm q)$ is given by:

$$\Delta f = (n \pm q)\Delta f_i \pm 2\pi f_0 \Delta q_i \tag{12}$$

$$\Delta f = f_0 \frac{\Delta p}{p}\left[(n \pm q)\eta \pm Q\xi\right] \tag{13}$$

if only chromaticity contributes to the betatron frequency spread.

Equation (13) shows that the width of the two Schottky bands is not the same, due to the machine chromaticity. However, by comparing the two bands $n \pm q$, one can determine the Δq_i of the beam. Even more, if one can identify similar points on the distribution (resonances, for instance), their q can be determined by the formula:

$$q = \frac{1}{2}\left(1 + \frac{\Delta f_c}{f_0}\right) \tag{14}$$

Δf_c being the measured frequency difference between them. This technique was extensively used in the ISR to monitor the working line of the machine distribution in transverse tunes.

Comparing equations (10) and (6) gives a direct measure of the r.m.s. betatron amplitude:

$$\frac{d_{rms}}{i_{rms}} = \frac{a_{rms}}{2} \ . \tag{15}$$

Equation (15) can be used to measure directly the transverse beam emittance, if the beam distribution is known. This obviously requires well calibrated longitudinal and transverse detectors to measure accurately d_{rms} and i_{rms} unless only relative measurements are sought (evolution of AA transverse emittance, for instance).

2.3 Bunched beam, longitudinal

In the bunched beam case, every individual particle executes synchrotron oscillations at the frequency $\Omega_s / 2\pi$. The time of passage of the particle in front of the detector is modulated according to:

$$\tau_i(t) = \hat{\tau}_i \ \sin\left(\Omega_s t + \psi_i\right) \tag{16}$$

$\tau_i(t)$ is the time difference with respect to the synchronous particle (frequency f_0) and $\hat{\tau}_i$ is the amplitude of the synchrotron oscillation, assumed to be linear. In time domain, the beam current is represented in Fig. 4, as a series of delta pulses, with a modulated time of passage. It can be written:

$$i_i(t) = ef_0 + 2ef_0 \mathrm{Re}\left\{\sum_{n=1}^{\infty} \exp\ jn\ \ \omega_0\left(t + \hat{\tau}_i\ \sin\left(\Omega_s t + \psi_i\right)\right)\right\} \ . \tag{17}$$

Using the relation:

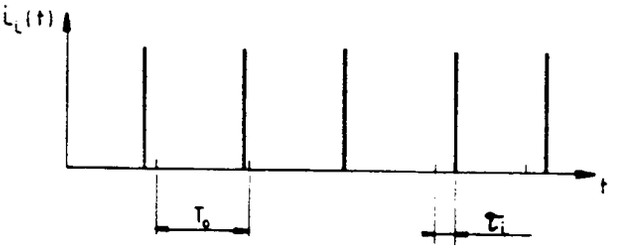


Fig. 4 Time domain representation of a single-particle current in a bunched beam

$$\exp\left(j(z\ \sin\ \theta)\right) = \sum_{p=-\infty}^{+\infty} J_p(z)\ e^{jp\theta} \tag{18}$$

where J_p is the Bessel function of order p, one can expand the nth harmonic in equation (17) and obtain:

$$i_n = 2ef_0 \mathrm{Re}\left\{\sum_{p=-\infty}^{+\infty} J_p(n\omega_0\hat{\tau}_i)\ \exp\ j(n\omega_0 t + p\Omega_s t + p\psi_i)\right\}. \tag{19}$$

Each revolution frequency line (nf_0) now splits into an infinity of synchrotron satellites, spaced by $\Omega_s / 2\pi$, the amplitudes of which being proportional to the Bessel functions of argument $n\omega_0\hat{\tau}_i$ as shown in Fig. 5.

The amplitudes of the synchrotron satellites become negligible beyond a certain value of p. This is because $J_p(x) \cong 0$ for $p > x$ if x is large. Therefore, the synchrotron satellites are, in practice, confined into a limited bandwidth:

$$2p\Omega_s = 2n\ \omega_0\ \hat{\tau}_i\Omega_s\ . \tag{20}$$

The spread in the instantaneous revolution frequency of the i^{th} particle due to the synchrotron oscillation is simply:

$$2\Delta\omega_i = 2\Omega_s\ n\omega_0\hat{\tau}_i\ . \tag{21}$$

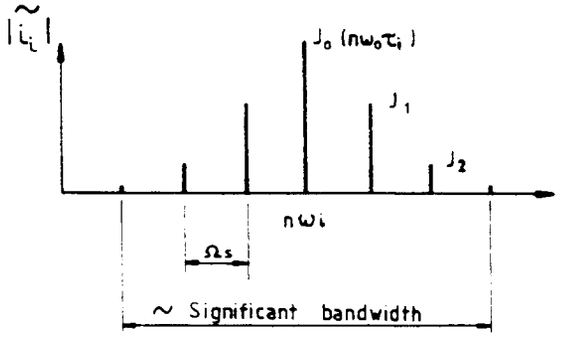


Fig. 5 Decomposition of each revolution line into synchrotron satellites

Consequently, for large values of n, the significant bandwidth around line n is the same as that of a beam of many particles having the same $\Delta\omega_i$ and therefore the same $\Delta p / p$.

Consider now the case of many particles, with randomly distributed synchrotron phases ψ_i and $\hat{\tau}_i$ ranging from 0 to $\hat{\tau}_m$ ($2\hat{\tau}_m$ being the total bunch length).

For a given n, the central line ($p = 0$) shows the same phase factor ($\exp jn\omega_0 t$) for all particles: the current in the central line is therefore proportional to N and not $\sqrt{N}$; this is simply the macroscopic RF current of the bunch. On the contrary, the synchrotron satellites ($p \neq 0$) add r.m.s. wise because of the random phase factor $\exp j(n\omega_0 t + p\Omega_s + p\psi_1)$ (Fig. 6).

Each line is infinitely narrow if the synchrotron oscillation is purely linear (Ω_s is the same for all particles) and if the machine has no imperfections. However, magnet and RF fluctuations broaden in practice each individual line. In addition a spread in synchrotron frequency within the bunch $\Delta\Omega_s$ transforms each satellite ($p \neq 0$) into a band of width $p\Delta\Omega_s$. For large values of n, overlap between successive synchrotron satellites $(p\Delta\Omega_s > \Omega_s)$ can occur within the significant width of the Schottky band of order n. (Fig. 6b).

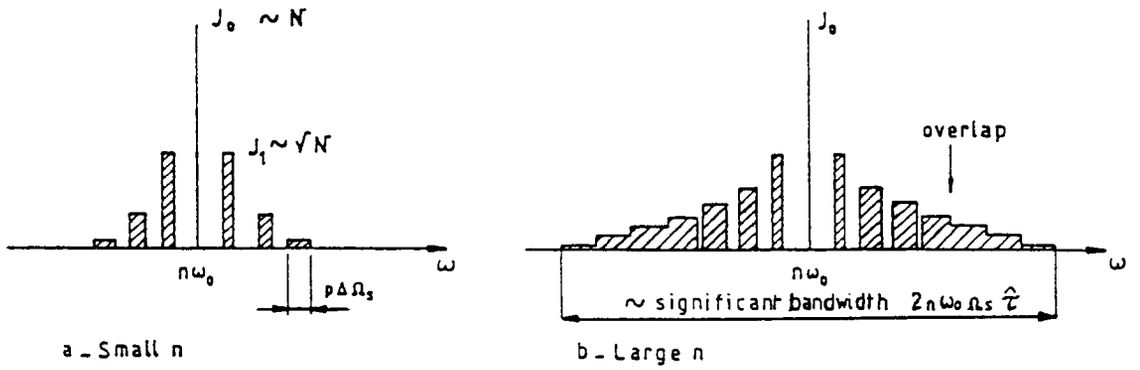


Fig. 6 Longitudinal Schottky spectrum of a bunched beam

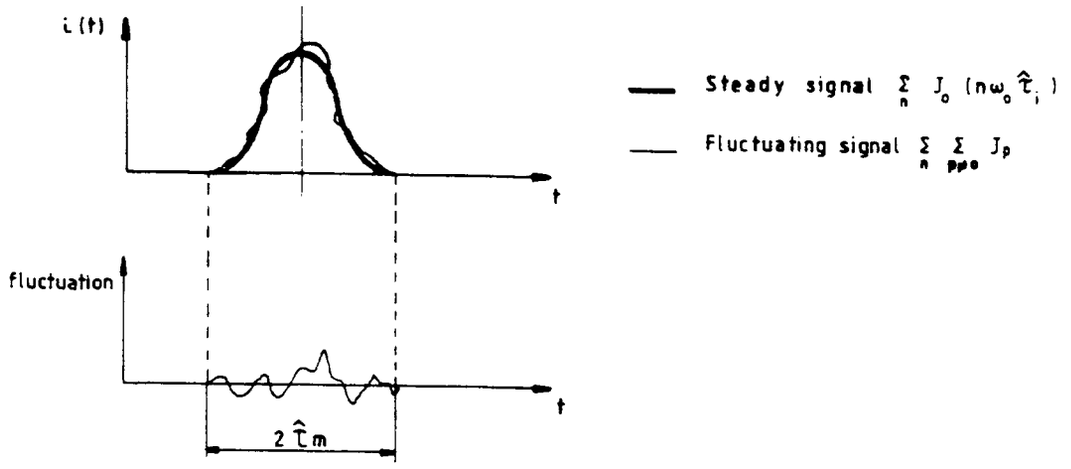


Fig. 7 Time domain of representation of bunched beam Schottky signal

If we consider two Schottky bands with different values of n, their corresponding synchrotron satellites (of order p) are correlated. This results from Eq. (19), where the random phase factor $p\psi_i$ is the same, even for different values of n.

Another way to look at the coherence between successive Schottky bands is to examine the bunch signal in time domain (Fig. 7). It is composed of a steady component (macroscopic signal resulting from the terms: $\sum_n J_0(n\omega_0\hat{\tau}_i)$) and a fluctuating Schottky signal $\left(\sum_n \sum_{p\neq 0} J_p(n\omega_0\hat{\tau}_i)\right)$.. The fluctuating signal extends in time over $2\hat{\tau}_m$, and can be Fourier decomposed into components at multiples of the fundamental bunch frequency $f_b = 1/2\hat{\tau}_m$. All information concerning the Schottky signal is contained into those components (in the limit $\Omega_s << \omega_0$). In other words significant information about the Schottky signal only appears every f_b frequency interval, the other spectral lines in (19) (every f_0) simply give redundant information, i.e., they are correlated.

As a consequence, sampling of Schottky signals at f_0, which folds many nf_0 bands on top of each other and only gives one Schottky signal, does not introduce any loss of information, if the bandwidth before sampling is limited to $\pm f_b/2$.

2.4 Bunched beam, transverse

Here we have to combine the amplitude modulation (betatron oscillation) and the time modulation (synchrotron oscillation). One obtains:

$$d_i(t) = a_i \cos(q_i\omega_o t + \varphi_i) ef_0 \,\mathrm{Re}\left\{\sum_{n=-\infty}^{+\infty} \exp\, jn\omega_0\left(t + \hat{\tau}_i \sin(\Omega_s t + \psi_i)\right)\right\} . \qquad (22)$$

If q_i is independent of ω_i, the n^{th} sum becomes:

$$d_n = ef_0 a_i \,\mathrm{Re}\left\{\sum_{p=-\infty}^{+\infty} J_p\left((n\pm q)\omega_0\hat{\tau}_i\right)\exp j\left[\left((n\pm q)\omega_0 + p\Omega_s\right)t + p\psi_i + \varphi_i\right]\right\} . \qquad (23)$$

Again, each betatron line splits into an infinite number of synchrotron satellites (Fig. 8). The significant bandwidth, as in the longitudinal case, approaches that of coasting beams with the same $\Delta p/p$, for large values of n. On the contrary, for small values of n, most of the energy is concentrated in the $p = 0$ line.

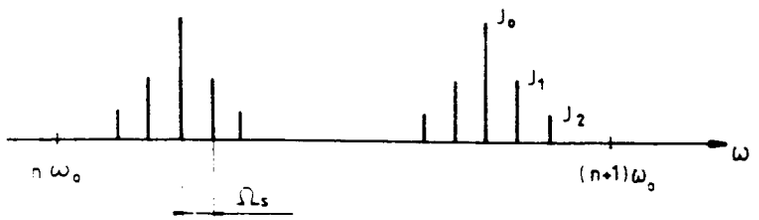


Fig. 8 Decomposition of each betatron line into synchrotron satellites

For a non-zero chromaticity, the argument of the Bessel function $(n\pm q)\omega_o\hat{\tau}_i$ should be replaced by $\left[(n\pm q) - Q\xi/\eta\right]\omega_0\hat{\tau}_i$. In this case, the relative amplitudes of the synchrotron satellites also depend on the chromaticity. In particular, for the chromatic frequency:

$$\omega_\xi = Q\frac{\xi}{\eta}\omega_0 \tag{24}$$

only the term J_0 is significant: all the energy of the nth Schottky band is concentrated in the central line.

With many particles, we should average over the two random variables φ_i and ψ_i. Unlike the longitudinal case, the central lines ($p = 0$) add up r.m.s. wise due to the random betatron phase factor φ_i, the consequence being that there is no transverse macroscopic signal. Successive bands are correlated as in the longitudinal case, again, because all the signal is concentrated in the time interval $2\hat{\tau}_m$ and not $T_0 = 1/f_0$ as if the beam were unbunched.

The width of the central line is determined by RF and magnetic field fluctuations, but also by transverse nonlinearities (tune spread due to octupole fields, beam-beam or space charge forces). In addition, the synchrotron satellites are broadened by the spread in synchrotron frequencies within the bunch (width $p\Delta\Omega_s$ as in Fig. 6).

The total power per band (for a given n) is given by:

$$\langle d_n \rangle^2 = e^2 f_0^2 \langle a^2 \rangle \frac{N}{2} \sum_p J_p^2\left((n \pm q)\omega_0 \hat{\tau}_i\right) . \tag{25}$$

With the identity:

$$\sum_{p=-\infty}^{+\infty} J_p^2(x) = 1 \tag{26}$$

one obtains

$$\langle d_n \rangle^2 = e^2 f_0^2 \langle a^2 \rangle \frac{N}{2} . \tag{27}$$

The total power per band is the same as in the coasting beam case, for the same total number of particles and the same transverse oscillation amplitude (Fig. 9).

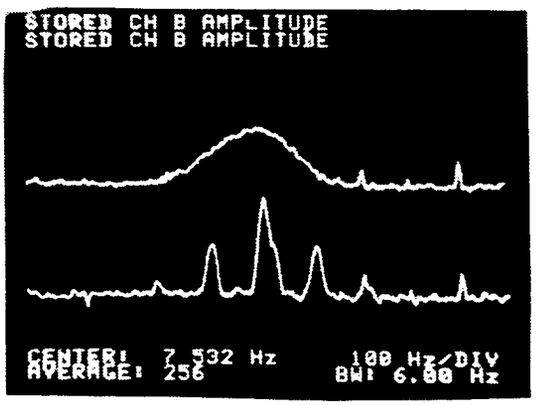


Fig. 9 Horizontal Schottky signals in the SPS.
Top: debunched beam, bottom: bunched beam

3. BEAM DETECTORS

3.1 The image-current approach

Consider the very simple geometry of Fig. 10a, where a round beam circulates in the centre of a cylindrical smooth vacuum chamber. This is a two-dimensional problem, and it is well known that the electromagnetic fields are purely transverse, as in a coaxial line, in the limit $v = c$. It follows that for all frequencies the beam and wall currents are opposite:

$$i_b = -i_w \tag{28}$$

Equation (28) is only valid up to some upper frequency, depending on the particle relativistic factor γ and the transverse dimensions of the vacuum chamber. However, for most practical cases (high energy storage rings) this is not a limitation.

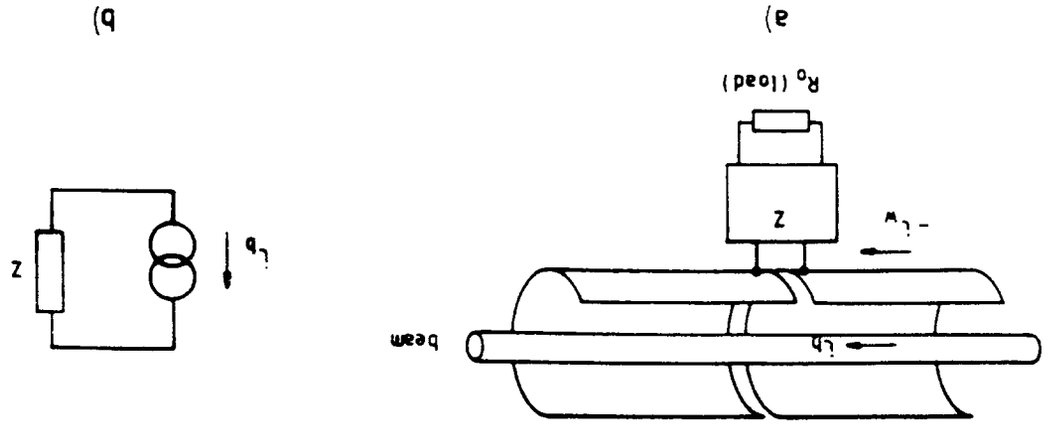


Fig. 10 The beam is equivalent to a current source flowing into the detector impedance

If now we cut a gap in the circular wall we introduce a coupling between the inside and the outside of the vacuum pipe. The latter is characterised by the impedance Z which we can measure between the two sides of the gap. As the energy lost by the beam when passing through the detector is much smaller than the particle's energy, the current i_b, and hence i_w is independent of the gap voltage: it means that the wall current i_w which flows through Z can be represented by a pure current source (Fig. 10b).

The detector, which seen from the gap appears like an impedance Z, delivers its output signal in the load R_0 (Fig. 10a). The sensitivity of the detector (longitudinal in this case) is defined by:

$$S = \frac{V_{\text{out}} \text{ in } R_0}{i_b}$$

For a lossless network between gap and R_0, one can easily obtain, from power considerations:

$$S = \sqrt{R_0 \cdot \text{Re}\, Z} \ .$$

The following examples will illustrate the image current approach for the evaluation of beam detectors (or beam pick-ups).

a) The resistive-gap kick-up

In this case the load resistor R_0 is simply connected to the vacuum chamber gap. However, to provide a low impedance DC return path for the wall current, a short-circuited coaxial line is built around the vacuum chamber, as shown on Fig. 11. The line is filled with lossy material (ferrites) such that, for the operating frequency of the pick-up, it appears as a terminated line. This introduces a high-pass characteristic in the detector response.

The upper frequency limit is determined by the parasitic capacitance at the gap. Making R_0 small (several parallel resistors) will increase the upper frequency limit, at the expense of sensitivity.

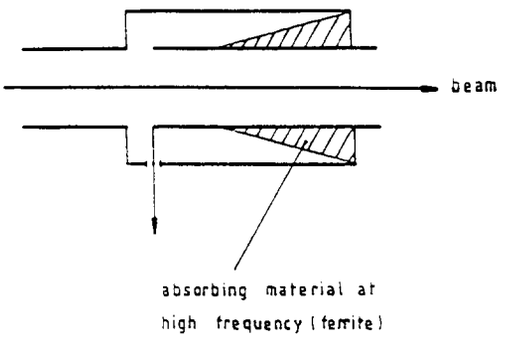


Fig. 11 Resistive-gap pick-up

The SPS wide-band longitudinal detector [3] uses eight parallel 50 Ω strip lines symmetrically connected to the gap, and a ferrite loaded coaxial line with 25 Ω characteristic impedance. This arrangement gives $Z = 5\ \Omega$. The eight gap signals are combined in an eight-port power combiner giving an overall sensitivity, in a 50 Ω load:

$$S = 5\sqrt{8} = 14\Omega$$

instead of the maximum $S = \sqrt{6.25 \times 50} = 17.6\Omega$ if no power would be lost in the ferrites (very high impedance coaxial line).

The bandwidth extends from 4 MHz to 4 GHz with almost no resonances. To improve the low-frequency response the inductance of the short-circuited line can be increased by lossless ferrites, but high-frequency resonances may be difficult to suppress.

b) The directional-coupler pick-up

As shown on Fig. 12a, there are two gaps in this detector, joined together by a piece of coaxial line of characteristic impedance R_0, surrounding the vacuum chamber, With the two load resistors R_0 which are connected to each gap, one can draw the equivalent circuit of Fig. 12b. The two beam current sources, at each gap, are in opposite direction, and are shifted in phase by the beam transit time.

The current flowing in the load R_0 on the right is the sum of the contributions from the two current sources:

$$\frac{i_b}{2}\exp\left(-j\omega\ell / \upsilon_\varphi\right) \qquad \text{left source}$$

$$-\frac{i_b}{2}\exp\left(-j\omega\ell / \upsilon_p\right) \qquad \text{right source}$$

υ_φ and υ_p being the wave and particle velocities and ℓ the distance between gaps. The total current:

$$\frac{i_b}{2}\left(\exp\left(-j\frac{\omega\ell}{\upsilon_\varphi}\right)-\exp\left(-j\frac{\omega\ell}{\upsilon_p}\right)\right) \tag{30}$$

vanishes if υ_p and υ_φ are equal.

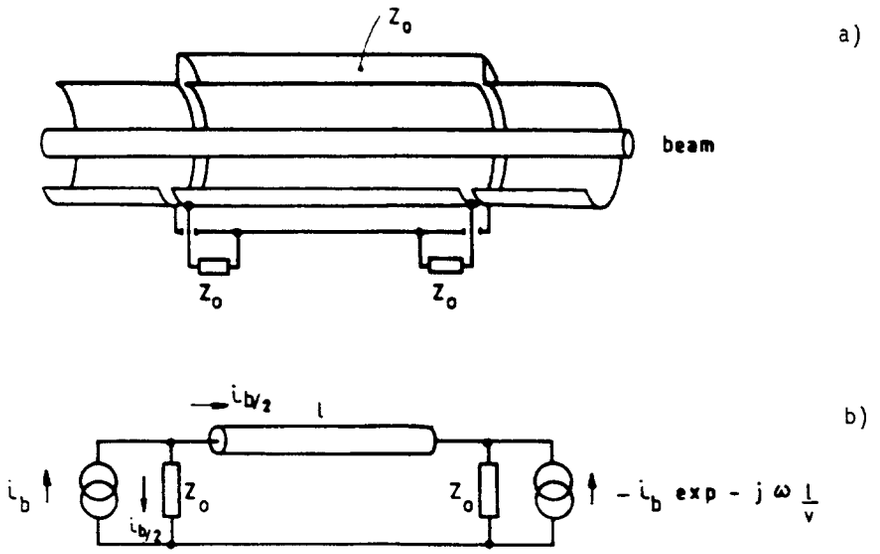


Fig. 12 Directional-coupler pick-up. a) Schematic, b) equivalent circuit

For the load R_0 on the left, one finds easily the current:

$$\frac{i_b}{2}\left(1-\exp\left(-j\omega\left(\frac{1}{\upsilon_\varphi}+\frac{1}{\upsilon_p}\right)\ell\right)\right) \tag{31}$$

and the corresponding sensitivity:

$$S=\frac{R_0}{2}\left|\left(1-\exp\left(-2j\omega\frac{\ell}{\upsilon}\right)\right)\right| \tag{32}$$

for $\upsilon_p=\upsilon_\varphi=\upsilon$.

If this synchronous condition is fulfilled, for instance if $\upsilon_p=c$ and the coaxial line is in vacuum, this detector is directional: the signal only appears at the upstream port (with respect to beam velocity). With counter rotating beams (p and pbars for instance) the directional pick-up can separate the signals from the two types of particles. In practice the directivity is of the order of 30 to 35 dB. Note that directivity can, in principle, be obtained also by combining the signals of several identical detectors.

The sensitivity of the detector, given by Eq. (32) is frequency dependent (Fig. 13). It shows a succession of zeros and maxima corresponding to:

$$\ell = \frac{\lambda}{2},\ \lambda,\ \frac{3\lambda}{2}: \qquad \text{zeros}$$

$$\ell = \frac{\lambda}{4},\ \frac{3\lambda}{4},\ \frac{5\lambda}{4}: \qquad \text{maxima}$$

the sensitivity being simply R_0 at the maxima.

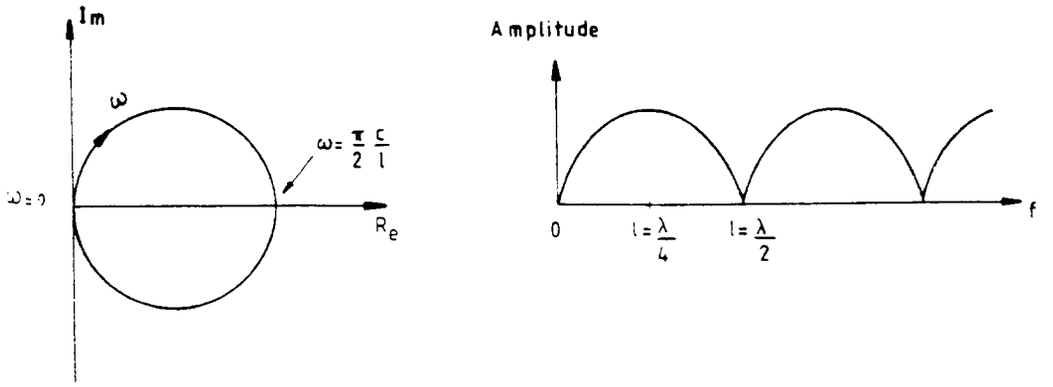


Fig. 13 Transfer functions of the directional-coupler pick-up

The transient response of the detector can be obtained by making the inverse Fourier transform of Eq. (32), but it is obvious from the equivalent circuit of Fig. 12b that it is composed of two opposite delta pulses separated in time by twice the transit time ($2\ell / c$) (Fig. 14a).

Several identical pick-ups can be combined to increase the overall sensitivity. With power combiners, the output signals are added power wise giving an overall sensitivity $S = R_0\sqrt{n}$ for n identical detectors, and the same frequency response. One can also combine several directional coupler detectors in cascade and obtain, with the proper delays, a transient response as in Fig. 14b. There the maximum sensitivity is proportional to n, but the frequency response now shows a $\sin f/f$ curve peaked at $\ell = \lambda / 4$. In other words, the higher sensitivity (proportional to n) results in a narrower bandwidth.

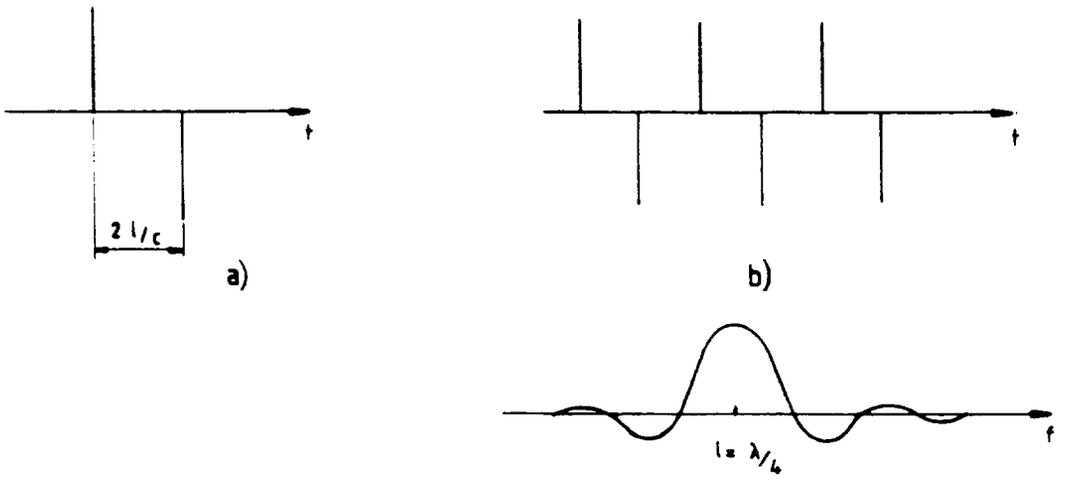


Fig. 14 Transient response of directional coupler. a) single, b) multiple (with the associated frequency response)

Directional coupler pick-ups are in fact mostly used as transverse detectors. With several strips symmetrically arranged in the vacuum chamber, as in Fig. 15a, the total wall current i_w should be replaced by $i_w\theta/2\pi$ for each strip, provided the beam is in the centre. For a non-centred beam the problem is truly three dimensional near the gaps. By approximating the electromagnetic field by that of a pure TEM wave one can obtain the wall current distribution along the vacuum chamber azimuth which obviously depends on the beam position. For small beam displacement, Δx, the difference of the signals of two opposite strip lines is proportional to Δx:

$$\Delta V = V_2 - V_1 = S_\Delta i_b \Delta x \tag{33}$$

S_Δ being defined by Eq. (33) as the transverse sensitivity of the detector (in ohms/meter).

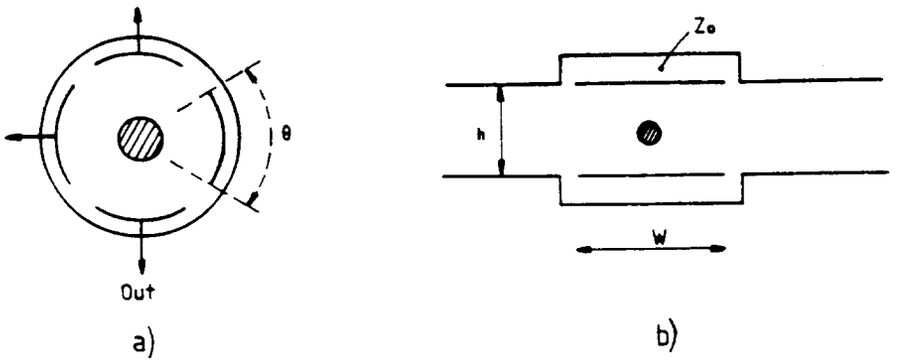


Fig. 15 Cross section of transverse directional-coupler pick-up. a) circular, b) rectangular.

In the case of a rectangular geometry, often used in wide-aperture cooling rings for instance, the maximum sensitivity S_Δ is given by:

$$S_\Delta = \frac{R_0}{h}\left(\tanh\frac{\pi w}{h}\right). \tag{34}$$

The form factor $\tanh(\pi w/h)$ simply reflects the fact that some fraction of the wall current flows outside the strip-line gaps.

This type of pick-up (sometimes called loop coupler) is widely used in cooling systems. It offers a good compromise between bandwidth (of the order of one octave) and sensitivity. The signals of many couplers are often added power wise on a combiner board, inside vacuum, to increase the overall sensitivity. If only one type of particle is present, the downstream resistor R_0, where no current flows, can be replaced by a short circuit (hence the name of loop coupler), but microwave resonances may be harmful in this case.

c) The electrostatic pick-up

If the coaxial line of Fig. 12a is much shorter than the wavelength ($\ell << \lambda$), it can be represented by a simple capacitor $C = \ell / R_0 v_\varphi$ (Fig. 16a). For a very high load resistor, the equivalent circuit of Fig. 16b represents the electrostatic detector, with the two current sources phase shifted by $\omega\ell / v_p$.

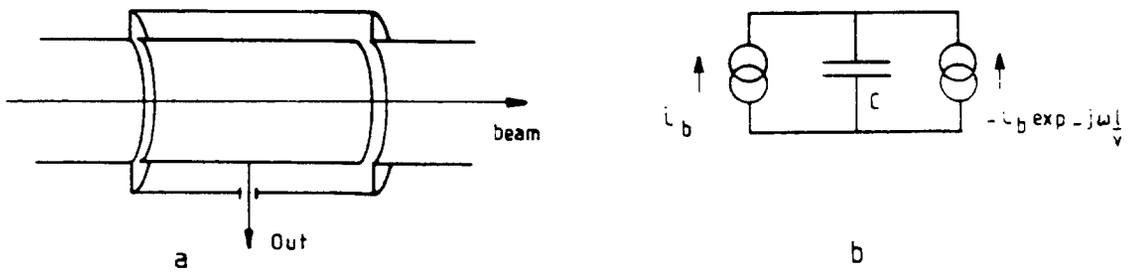


Fig. 16 The electrostatic pick-up

The voltage developed on the line (or the electrode) is simply:

$$V = \frac{1}{jC\omega}\left[i_b - i_b \exp\left(-j\omega\ell / \upsilon_p\right)\right] \tag{35}$$

$$V \cong \frac{i_b}{jC\omega} \cdot j\frac{\omega\ell}{\upsilon_p} \ . \tag{36}$$

The quantity $\ell i_b / \upsilon_p$ is the beam charge q contained in the detector length, (assuming a slowly varying charge distribution with respect to the electrode length). It follows:

$$V = q / C \tag{37}$$

as the electrostatic theory would have given immediately.

For $\upsilon_\varphi = \upsilon_p$ and in the approximation of a high load resistor, Eq. (36) combined with $C = \ell / R_0 \upsilon_\varphi$ leads to the very simple result:

$$S = R_0 \ .$$

The sensitivity is independent of the frequency and of the length of the detector. Of course this is only true at medium frequencies. The non-infinite load resistor (usually an amplifier with high input impedance) introduces a low frequency cut off whereas at high frequencies the approximation $\ell << \lambda$ is no longer valid.

The transverse version of the electrostatic pick-up can be obtained by splitting the electrode cylinder in two halves along a linear cut (Fig. 17). Electrostatic theory shows that the difference in voltage between the two plates is a linear function of the beam displacement. Many versions of the transverse electrostatic pick-up with various shapes could be found in the literature (circular, rectangular, elliptical) [4]. They are mostly used for closed orbit measurements (sometimes horizontal and vertical pick-ups are combined in a single unit).

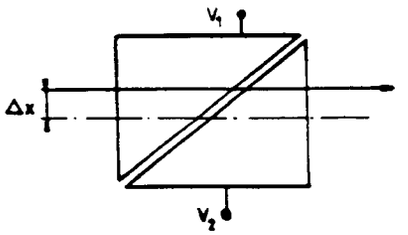


Fig. 17 Transverse electrostatic pick-up linearly cut

If the linearity requirement is less important, the linear cut could be abandoned, for instance in the so-called "buttons" used in LEP (Fig. 18). There, only the high-frequency response is important, and consequently the load resistor is a 50 Ω cable. The linearity can be restored by a proper algorithm at the signal processing level.

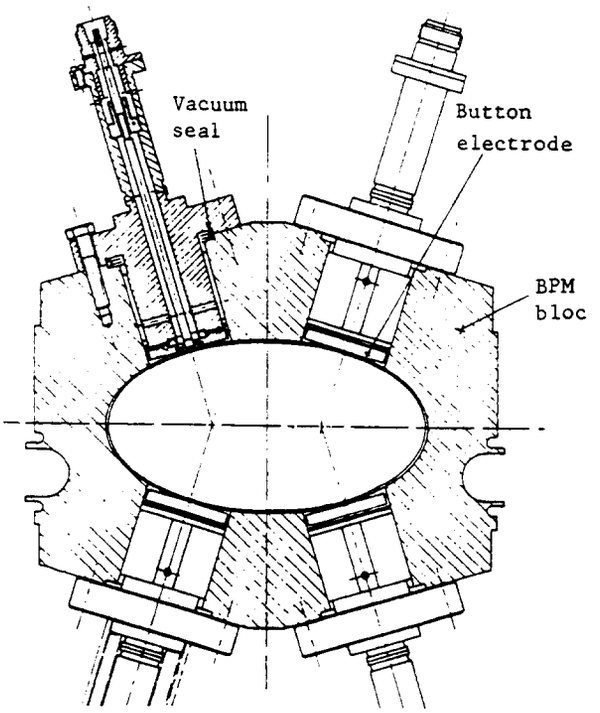


Fig. 18 Cross section of the LEP position detector

The electrostatic detector can be made resonant, with a coil (or transformer) connected to the electrode. A transverse version is sketched in Fig. 19a, with the equivalent circuit of Fig. 19b.

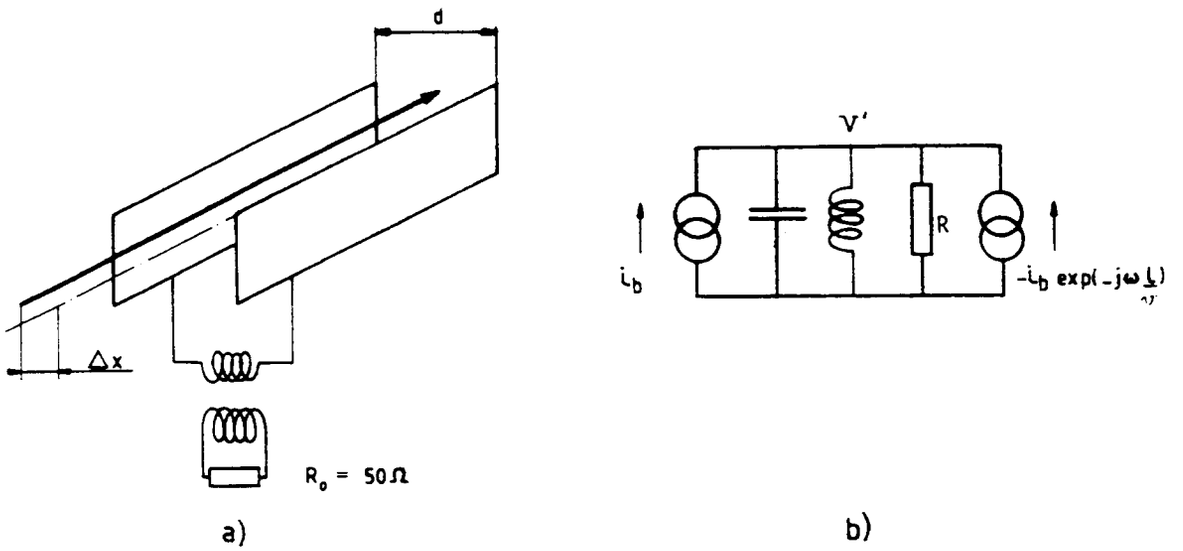


Fig. 19 a) Resonant electrostatic pick-up, b) Equivalent circuit

At resonance the voltage across the plates V' is given by:

$$V' = j\omega \frac{\ell}{v_p} R \frac{2\Delta x}{d} i_b \tag{38}$$

which is transformed in the output load into:

$$V_{out} = \sqrt{R_0 R}\ \omega \frac{\ell}{v_p} \frac{2\Delta x}{d} i_b \tag{39}$$

for a lossless transformer.

Taking into account the ohmic losses of the coil (Q_0 = quality factor of the resonant circuit, Q_L = loaded quality factor), one obtains:

$$S_\Delta = \frac{2}{d} \sqrt{\frac{R_0 \omega Q_L}{C}} \frac{\ell}{v_p} \left(1 - \frac{Q_L}{Q_0}\right). \tag{40}$$

This technique has been used in the CERN SPS, for a dedicated, very sensitive Schottky detector [5] (sensitivity: 75 Ω/mm).

2.2 Pick-up evaluation using the reciprocity theorem

The reciprocity theorem, well known in antenna theory, results from Maxwell equations applied to a linear, isotropic system. If we have two sets of current sources in the system J' and J'' which produce the electric fields E' and E'' and the magnetic fields H' and H'', the following relation is valid:

$$\iint_s (E'' \times H' - E' \times H'') \cdot \vec{n} ds = \iiint_v (E' \cdot J'' - E'' \cdot J') dv \tag{41}$$

where the volume v is enclosed by the surface s ($\vec{n}$ is the unity vector on that surface).

For the application of the reciprocity theorem (Fig. 20), we take $\vec{J}' = \vec{i}_b$, ($\vec{i}_b$ is the beam current along the detector axis), and $J'' = I_1$ (I_1 is a pure current source applied across the load resistor R_0).

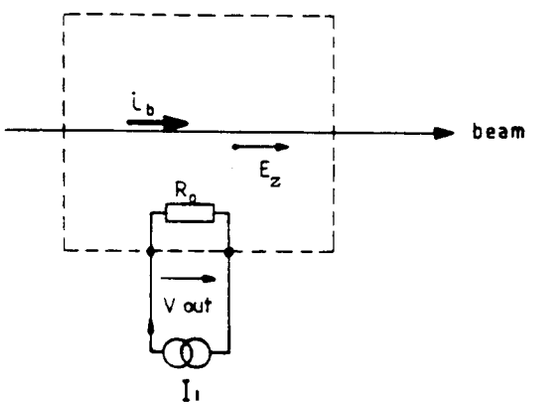


Fig. 20 Application of the reciprocity theorem to a beam detector

We consider an integration volume limited by the metallic enclosure of the pick-up, where the electric fields are normal to the surface, which makes the left side of Eq. (41) vanish and leads to:

$$\iiint_v \vec{E}_b \cdot \vec{I}_1 \; dv = \iiint_v \vec{E}'' \cdot \vec{i}_b \; dv \tag{42}$$

$$I_1 \cdot V_{out} = \int_z E_z \cdot i_b \; dz \tag{43}$$

where V_{out} is the output voltage of the detector when excited by i_b, and E_z is the on axis component of the field in the pick-up structure when excited by I_1. For a given geometry and a given field configuration, E_z can be related to I_1, from power considerations. Then application of Eq. (43) directly gives the detector sensitivity $S = V_{out} / i_b$, for cases where the image current approach would fail (e.g. microwave structures).

Note that the reciprocity theorem, transposed in circuit theory, simply states that, for a passive quadrupole, the determinant of its transfer matrix is unity.

Application of the reciprocity theorem will be illustrated in the following by two examples: the slow-wave and the slot-line pick-ups.

The slow-wave pick-up is essentially an electromagnetic wave guide in which the phase velocity has been slowed down to match the velocity of the particles. Dielectric slabs (Fig. 21a) or corrugations (Fig. 21b) have been considered for this purpose [6]. A description of the field in the structure will be given by standard wave guide theory. With respect to the transverse dimension, the E_z field configuration is either symmetrical (even mode) or antisymmetrical (odd mode), leading to a longitudinal or a transverse detector respectively.

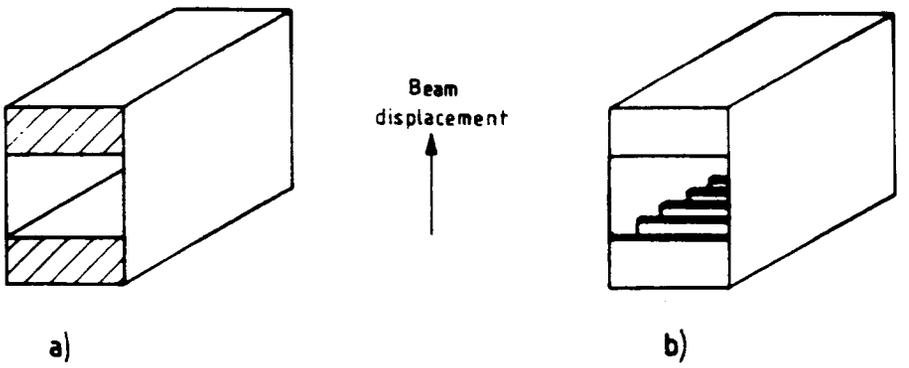


Fig. 21 Slow-wave pick-ups. a) dielectric slab, b) corrugated wall.

In the case of a pure travelling-wave structure, terminated at both ends by resistors R_0 via matched transitions, the power flow P_0 in the waveguide is related to I_1 by:

$$P_0 = \frac{1}{2} R_0 \left(\frac{I_1}{2} \right)^2 \tag{44}$$

Note that only $I_1/2$ flows towards the waveguide, the rest being dissipated in the load resistor R_0.

The sensitivity is given by:

$$\int_z E_z i_b \ \exp\left(jk_0 z / \beta_p\right) dz = I_1 V_{\text{out}} \ . \tag{45}$$

It is found to be proportional to the transit time factor:

$$\frac{\sin k_0\left(\frac{1}{\beta_p} - \frac{1}{\beta_\varphi}\right)\frac{\ell}{2}}{k_0\left(\frac{1}{\beta_p} - \frac{1}{\beta_\varphi}\right)\frac{\ell}{2}} \tag{46}$$

k_0, β_φ, β_p: propagation constants in free space, waveguide and beam respectively.

The sensitivity is optimum for $\beta_p = \beta_\varphi$ (synchronism condition) as expected. For a given frequency, optimum dimensions of the waveguides are given by the synchronism condition (as in Fig. 14b). Making the detector longer increases the sensitivity (proportional to ℓ) but reduces its bandwidth according to (46).

The slot-line pick-up [7,8] offers another interesting example, in which the waves propagate in a direction perpendicular to that of the beam (Fig. 22). A thin slot in a metallic plane on a dielectric substrate can support quasi TEM waves in the upper region. The electric field, not too close to the slot, is purely tangential: its amplitude is given by:

$$E_\emptyset = V_0 \frac{k_c}{2j} H_1^{(1)}(k_c r) \tag{47}$$

with:

$$k_c = \frac{2\pi j}{\lambda}\sqrt{\left(\frac{\lambda}{\lambda'}\right)^2 - 1} \tag{48}$$

V_0 is the voltage across the slot, λ' the wavelength along the slot and $H_1^{(1)}$ the Hankel function of first order.

The longitudinal field E_z, along the beam (at a distance d from the metallic plane) is simply:

$$E_z = \frac{d}{r} E_\phi \ . \tag{49}$$

For $\lambda' \sim \lambda$ (ε_r not too large) one can replace $k_c H_1^{(1)}(k_c r)$ by $2j\,\pi r$, which gives:

$$E_s = \frac{V_0}{\pi}\frac{d}{r^2} \tag{50}$$

V_0 is related to the power flow P_0 along the gap by the slot-line impedance Z: $P_0 = V_0^2 / 2Z$. Combined with (44) and (45), one obtains:

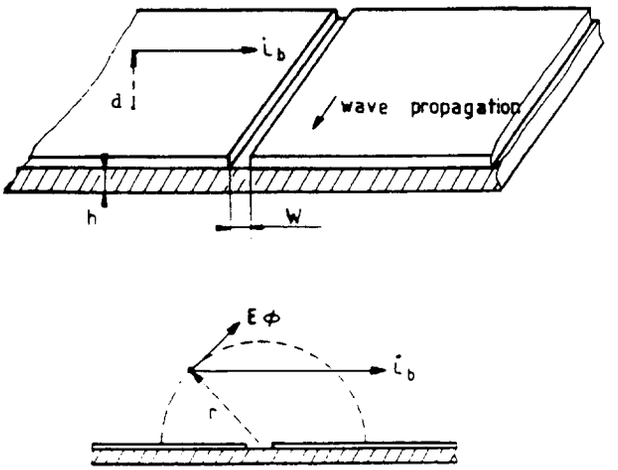


Fig. 22 Schematics of slot-line pick-up

$$S = \frac{\sqrt{R_0 Z}}{2} \int_z \frac{1}{\pi} \frac{d}{r^2} \cos\frac{2\pi z}{\lambda} dz \tag{51}$$

$$S = \frac{\sqrt{R_0 Z}}{2} \exp\left(-\frac{2\pi d}{\lambda}\right). \tag{52}$$

Equation (52) can be shown to be valid also even if λ and λ' are not very close: λ should then be replaced by λ' in (52).

It is interesting to remark that in the limit $d << \lambda$, Eq. (52) reduces to $S = \sqrt{R_0 Z}/2$ which is the result given by the image current approach. With the reciprocity theorem, transverse propagation which was previously neglected can be taken into account.

If the signals of two symmetrical plates with two slots are combined, a transverse detector can be built. Its sensitivity would be:

$$S_\Delta = \frac{\sqrt{R_0 Z}}{h} \frac{1}{\sinh(\pi h / \lambda')} \tag{53}$$

h being the distance between plates.

Slot-line pick-ups would be interesting, because they can be easily produced by standard printed-circuit techniques, even in the microwave region. Their bandwidth is only limited by that of the slot-line to strip-line transitions (the wave on the slot is coupled to outside via a strip line deposited on the opposite side of the dielectric). Because of the transverse propagation, the inherent delay of the detector depends on the transverse beam position. This could be useful for some stochastic cooling schemes.

3.3 Impulse response

Consider again a travelling wave detector like, for instance, the corrugated wall waveguide, where a number of cells (or individual rectangular boxes) are coupled together via the beam pipe. When excited by a short beam pulse, the response of the detector is, in first

approximation, an RF burst (Fig. 23) of amplitude V_{out} and duration τ. After the time τ, all the energy deposited in the detector has been transported with the group velocity v_g to the end of the structure and then to the terminating resistor R_0.

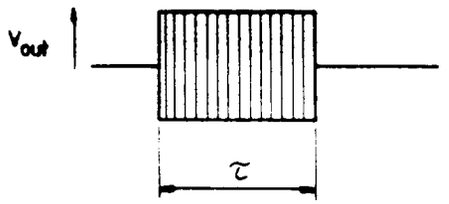


Fig. 23 Impulse response of a travelling-wave detector

What is the relation between the detector sensitivity S and its output voltage V_{out} in this case? If we assume a periodic train of short beam pulses (charge q), separated in time by τ, the RF component of the beam current i_b at the central frequency of the pick-up is simply $i_b = 2q / \tau$. Obviously the output amplitude V_{out} is constant, with that particular beam input, which gives:

$$V_{out} = S \; i_b = 2Sq / \tau \tag{54}$$

for the longitudinal case and:

$$V_{out} = 2S_\Delta \; q \; \Delta x / \tau \tag{55}$$

for the transverse case.

The energy W deposited by the charge q in the detector is related to the geometry of the structure via its "loss parameter" defined by:

$$W = k \; q^2 \; . \tag{56}$$

The k factor is also the R/Q of the structure ($k = \frac{1}{2}\omega_c \; R/Q$, ω_c: frequency of the mode considered).

Combining (54), (56) and the relation:

$$W = \frac{1}{2}\frac{{V_{out}}^2}{R_0}\tau \tag{57}$$

valid for a lossless detector one finds:

$$S = \sqrt{\frac{R_0 k \tau}{2}} \tag{58}$$

and a similar equation for the transverse case.

The loss parameter k depends essentially upon the cell geometry, and can be calculated analytically in some simple cases (neglecting the effect of the beam hole) or evaluated by computer codes like SUPERFISH for instance. On the other hand τ characterises the cell-to-cell coupling via v_g.

From Eq. (58), the maximum sensitivity is again proportional to ℓ (detector length) as both k and τ are themselves proportional to ℓ. Of course the bandwidth decreases correspondingly as was shown in the example of the multiple directional coupler (Fig. 14b). Note that this multiple directional coupler can be considered as a backward-travelling wave structure with $v_g = c$.

In the following example, we shall evaluate the k factor for the simple geometry of Fig. 24: a chain of coupled cylindrical cavities. We consider the mode E_{011} (transverse detector) where the electric field is only longitudinal:

$$E_x = E_\phi = 0 : \quad E_z = E_0 \; J_1(2\pi x / \lambda)\cos\,\phi \tag{59}$$

The energy lost by charge q is given by:

$$W = \frac{1}{2} q \int_\ell E_z \exp\, j\frac{\omega_c z}{c} dz \; . \tag{60}$$

The factor 1/2 simply reflects the fact that the charge q only sees one half of its own induced voltage (fundamental theorem of beam loading). W is also obtained by integrating E_z^2 over the whole cavity volume:

$$W = \frac{1}{2} \iiint \varepsilon_0 \; E_z^2 \; dv \; . \tag{61}$$

Eliminating E_0 between (59), (60) and (61) finally gives:

$$W = \frac{\mu_0}{4\pi} \frac{\ell \omega_c^2}{2.38} \left(\frac{2\pi}{\lambda}\right)^2 \left(\frac{\sin(\pi L / \lambda)}{\pi L / \lambda}\right)^2 q^2 x^2 \; . \tag{62}$$

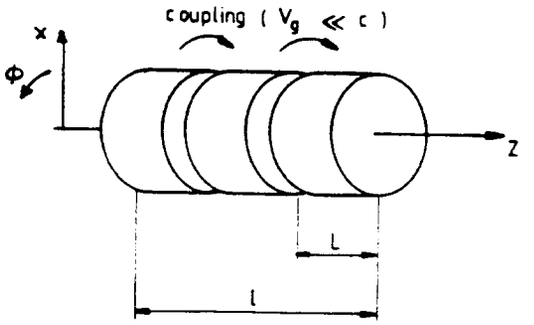


Fig. 24 A chain of coupled cylindrical cavities as transverse detector

Equation (62) shows the interest of high-frequency detectors as far as sensitivity is concerned (factors ω_c^2 and $1/\lambda^2$). But the influence of the beam hole which has been neglected

in this simplified analysis will become more and more important. An example of this type of detector, using the first transverse mode of the accelerating cavities in the CERN SPS is given in Ref. 9.

4. OBSERVATION OF SCHOTTKY SIGNALS

4.1 Spectral analysis

As already mentioned in section 2 the measurement of the power spectral density of the Schottky signals gives the particle distribution in either momentum or betatron tune (or a combination of both). Therefore, spectral analysis is the natural technique for observing Schottky signals.

The frequency span of interest is of the order of the revolution frequency, or even less (in most cases below 100 kHz). Consequently, the Fast Fourier Transform (FFT) or, more precisely, the Digital Fourier Transform (DFT) techniques which operate at low frequencies, can be used to evaluate in real time the signal spectrum. The Schottky band to be analysed must be translated at low frequency prior to FFT analysis, as in a conventional spectrum analyser. This may require a careful prefiltering to reject the unwanted image frequencies.

In the DFT technique, the signal is sampled and digitised at frequency f_S. Each digital word is stored in a memory with M locations (typically $2^{10} = 1024$ locations): the duration of the signal sample to be analysed is then $T = M/f_S$. The frequency content (frequency span) of the sampled signal extends only up to $f_S/2$ (Nyquist theorem), and the resolution of the frequency analysis is of the order of $1/T$ (Fig. 25). Depending on the choice of the signal processing "windowing", the resolution varies a little: $1/T$ for the rectangular window; $1.4/T$ for the "Hamming window", better optimised for noise signals.

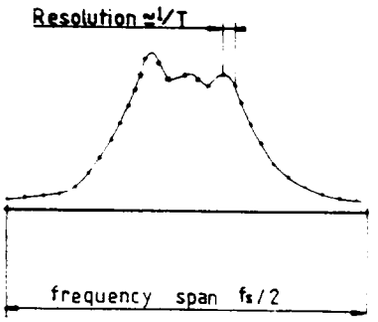


Fig. 25 Spectral analysis (DFT) of Schottky signals

For a given resolution of the beam distribution measurement (in $\Delta p/p$, or $\Delta Q/Q$), T is minimum for the largest width of the Schottky band. For instance, in the longitudinal, debunched beam case, one would minimise T by looking at the highest frequency Schottky bands (width $n\Delta f$) limited by either $f_S/2$, the detector sensitivity, or the overlap condition. This is of particular interest for the observation of "pseudo" Schottky signals in pulsed machines to measure the beam momentum spread during debunching. (T is there strictly limited by the duration of the magnetic cycle flat top.) The beam develops, during debunching at high intensity, a very complicated structure which is more or less equivalent to random noise, but of macroscopic nature ("pseudo" Schottky signal). Its spectrum analysis provides an estimate of the momentum spread of the beam during debunching.

Even if T can be made very long, the result of the DFT on a noise signal does not give a good estimate of its spectral density. This is because the variance of the power measurement is comparable to its mean value: it does not decrease when T is made longer.

A better "estimation" of the true power density is obtained by averaging several spectra taken at different time intervals. The "degree of confidence" of the measurement increases with the number of averaged spectra (Fig. 26), at the expense of the total analysis time (which may be distributed over several machine cycles in the previous example).

4.2 Parasitic signals of the Schottky spectrum

Due to the very low level of the Schottky signals, many sources of disturbance can be harmful and should be eliminated whenever possible.

Parasitic signals may come from the beam itself; if there is a coherent excitation (i.e. transverse) it will appear as a betatron signal, but with an amplitude proportional to N and not $\sqrt{N}$ as for the Schottky signal. The longitudinal line in a transverse Schottky scan can be suppressed by careful centering of the pick-up on the beam axis. In the bunched beam case, additional sharp filtering with a crystal filter is necessary [5].

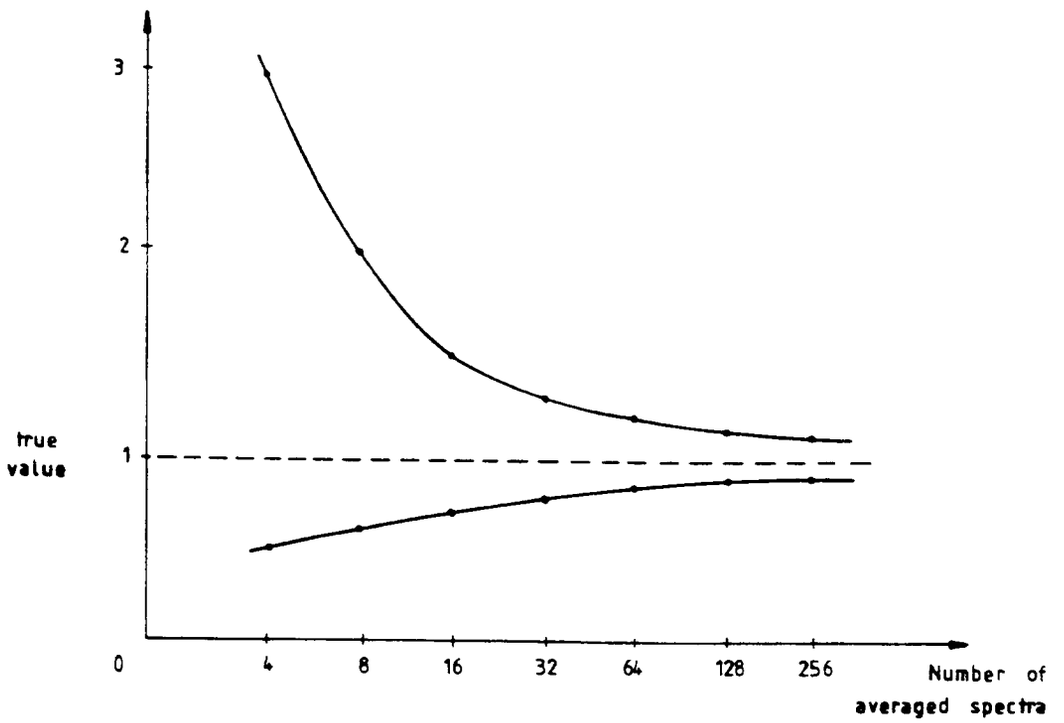


Fig. 26 90% confidence level of the noise spectral density measurement

The line-related components are reduced by a careful design of the amplifiers, power supplies and earth connections. If this is not sufficient, narrow-band synchronous filtering (locked to the mains frequency) can also be employed.

Of more fundamental nature is the disturbance due to the thermal noise of the first preamplifier, after the detector. The amplifier is characterised by its noise factor F (excess noise with respect to a simple resistor R_0). The available noise spectral density resulting from the amplifier and which is given by: $Fk_0t_0R_0$ (k_0: Boltzmann constant, t_0: temperature) must be considerably smaller than the Schottky noise spectral density. One possibility is to cool the

preamplifier and the terminating resistors of the pick-ups (ACOL). From Fig. 2, it is clear that the best signal-to-noise ratio is obtained at low frequencies (for the case of unbunched beams). Unfortunately, observation of Schottky signals at high frequency is more favourable as far as sensitivity and analysis time are concerned.

4.3 Bunched-beam signal processing

As all Schottky lines in a frequency interval f_b are correlated (see section 2), it is interesting to sample the beam signal at the revolution frequency. All lines will be folded in the base band giving a much better signal-to-noise ratio as will be shown in the following.

In the case of a detector with a bandwidth larger than f_b, the Schottky signal appears like a noise burst of length $2\hat{\tau}_m$. Its peak power is the same as that of an unbunched beam having the same line density, i.e. with a total number of particles $N_{eff} = N / 2\hat{\tau}_m f_0$. Therefore, when sampling the Schottky signal at the revolution frequency the resulting signal-to-noise is the same as if it were given by an unbunched beam of $N_{eff}\left(N_{eff} >> N\right)$ particles.

With a detector having a slower response like a travelling wave pick-up, the output RF burst (Fig. 23) can be much longer than the bunch itself $\left(\tau >> 2\tau_m\right)$, but nevertheless very short comapred to $1/f_0$. Take, for instance a transverse detector giving an output voltage:

$$v = 2S_\Delta e\frac{x}{\tau}\cos\left(\omega_\beta t + \Psi\right)\cos\omega_{pu}t \tag{63}$$

for a single particle which performs a betatron oscillation $x \cos\left(\omega_\beta t + \Psi\right)$. ω_{pu} is the centre frequency of the detector.

The average power, per particle is therefore:

$$\langle v^2\rangle = S_\Delta^2 e^2\frac{x^2}{\tau^2}$$

and for N particles:

$$\langle v^2\rangle = NS_\Delta^2 e^2\frac{\langle x^2\rangle}{\tau^2}\ . \tag{64}$$

The thermal noise power of the amplifier, referred to the input, for an amplifier bandwidth B can be written as:

$$\langle v_{th}^2\rangle = F\ \ k_0 t_0 R_0 B\ .$$

The power signal-to-noise ratio, during the time interval τ is therefore:

$$\frac{1}{U} = \frac{N\ \ e^2 S_\Delta^2\langle x^2\rangle}{F\ \ \tau^2 k_0 t_0 R_0 B}\ . \tag{65}$$

This is also the signal-to-noise ratio after sampling. We can select $B\left(B = B_{\text{opt}}\right)$ to optimise $1/U$. B_{opt} is the minimum bandwidth for which the useful signal is not reduced significantly. This happens if the rise time of the band limited RF burst is of the order of its

length: $1/B = \tau$, as illustrated in Fig. 27. More precisely B_{opt} is that of the so-called "optimum filter" (radar terminology) for which the impulse response is the time reversed image of the RF burst. With that condition (65) becomes:

$$\frac{1}{U} = \frac{1}{\tau f_0} N \frac{e^2 f_0 S_\Delta^2 x^2}{F k_0 t_0 R_0} \tag{66}$$

which is the same as for the debunched beam case, except for the enhancement factor $1/\tau f_0$ which can be much larger than unity [10].

The overall signal processing system for a bunched beam transverse Schottky signal is displayed on Fig. 28. Frequency translation down to the base band frequency can be done by peak detection, as indicated, or with a synchronous detector driven by the sum signal of the pick-up. In this case, it is interesting to remark that the odd synchrotron satellites are rejected for an in-phase detection (like for a peak detection), whereas for a quadrature detection, it is the even synchrotron satellites which are rejected. This feature may be useful if one wants to isolate the central J_0 line of the Schottky band.

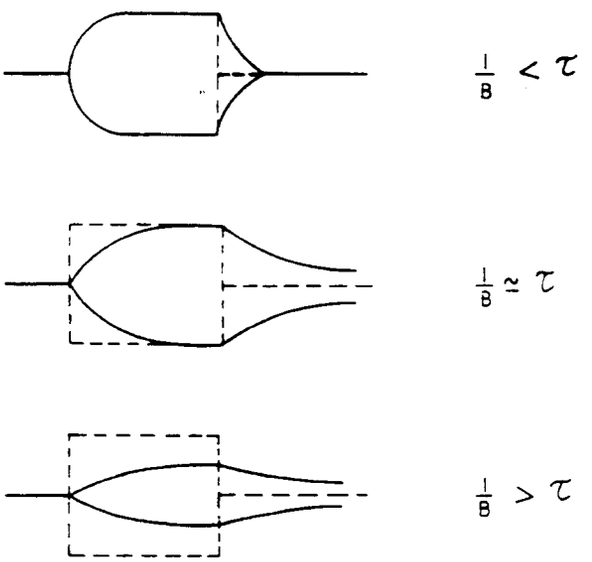


Fig. 27 Optimum filtering of an RF burst from a beam detector

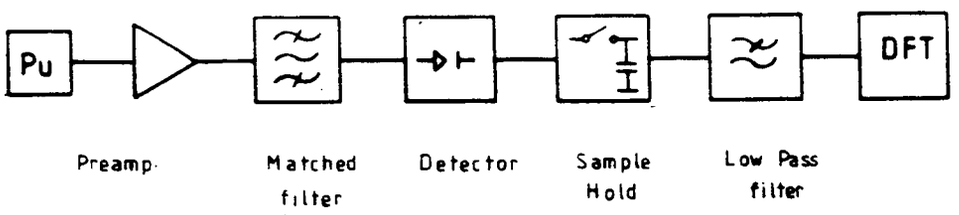


Fig. 28 Bunched beam signal processing system

Although the thermal noise of the preamplifier is of less importance for bunched beam signal processing, the effect of spurious coherent excitation of the beam may be more of a problem. This is because, even a low frequency excitation, near the first betatron line, appears everywhere in the spectrum, contrary to the debunched beam case, and may spoil even a high frequency Schottky system. A solution to that problem is to reject that part of the detector signal which is coherent from one bunch to the next [9].

5. BEAM TRANSFER FUNCTIONS

5.1 Principle of beam transfer functions

The name of beam transfer function almost speaks for itself: it relates the response of the beam (amplitude and phase) to a known excitation. In the case of a transverse excitation by a deflector (or kicker), the beam response is measured by a transverse pick-up as indicated on Fig. 29a, whereas Fig. 29b shows the arrangement for the measurement of a longitudinal transfer function [11].

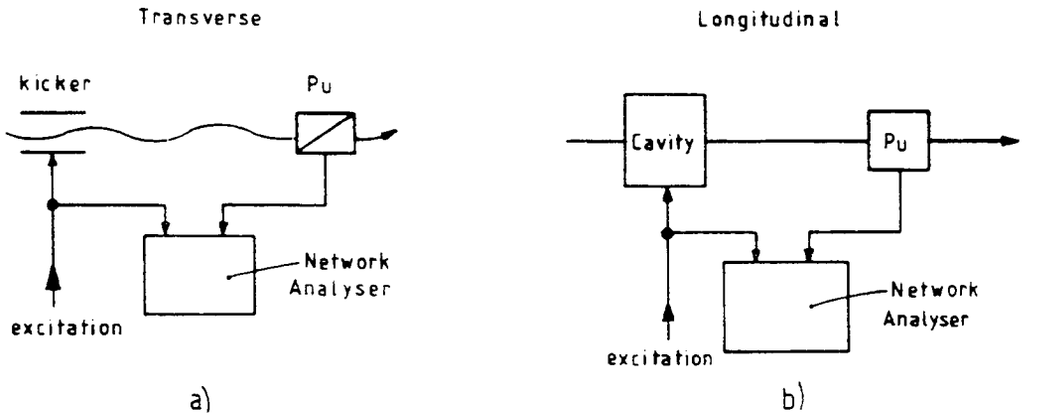


Fig. 29 Principle of beam transfer function measurement

To minimise the analysis time and the disturbance to the beam, it is interesting to excite the beam with a white noise spectrum (all frequencies are present in the band of interest). There the output will also be a noise signal, similar to the Schottky noise, and for which similar processing techniques can be applied. To extract the phase information spectral density measurements are not sufficient and a dual channel DFT instrument is needed. Again averaging many transfer functions reduces the variance of the estimate (Fig. 26). In Fig. 29a and b, a new element appears, namely the kicker (either transverse or longitudinal) which will be examined more in detail in the following.

5.2 Kickers

A longitudinal kicker is a fairly straightforward device in which a longitudinal electric field E_z is produced. The particle gains an energy ΔW, when crossing the kicker (or cavity), which is simply given by:

$$\Delta W = \int_z e \; E_z \; dz \; . \tag{67}$$

The application of the reciprocity theorem to a longitudinal beam detector has led us to Eq. (43), which combined with (67) results in:

$$\Delta W = e \; S \; I_1 \tag{68}$$

showing that the energy gain of the kicker and the sensitivity of the pick-up are simply proportional. In other words a longitudinal kicker is nothing but a longitudinal detector working in reverse. This is almost obvious for cavity like detectors, but is also true for a directional coupler type of pick-up for instance, where a quasi TEM wave propagates. There, only the field at the ends of the coupler are useful for beam excitation.

Consider now the case of a transverse deflection produced by the Lorentz force:

$$\Delta \vec{p} = \frac{1}{v_p} \int_z e\left(\vec{E} + \vec{v}_p \times \vec{B}\right) \; dz \tag{69}$$

which projected on the x axis (Fig. 30) can be written:

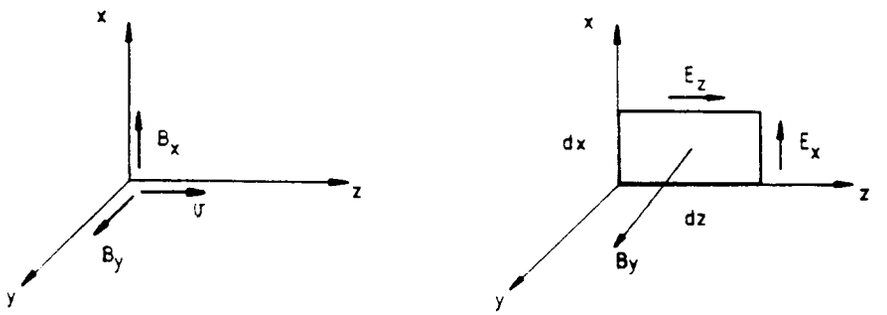


Fig. 30 Application of the induction law to the evaluation of the transverse force

$$\Delta p_x = \frac{1}{v_p} \int_z e\left(E_x + v_p \cdot B_y\right) \; dz \; . \tag{70}$$

To evaluate the quantity $E_x + v \cdot B_y$, we apply the induction law to a small rectangle in the $x0z$ plane:

$$\oint_C \vec{E} \cdot d\vec{\ell} = -\frac{\partial}{\partial t} \phi_B \tag{71}$$

ϕ_B being the flux of the magnetic field B on the contour C. One obtains

$$\frac{dE_z}{dx} dx \; dz + \frac{dE_x}{dz} dz \; dx = -j\omega \; B_y \; dx \; dz \tag{72}$$

$$\frac{dE_z}{dx} + \frac{dE_x}{dz} = -j\omega \; B_y \tag{73}$$

With:

$$\frac{dE_z}{dz} = \frac{dE_x}{dt} \frac{dt}{dz} = j\omega \; E_x \frac{1}{v_p} \tag{74}$$

(73) becomes:

$$E_x + v \cdot B_y = \frac{v_p}{j\omega} \frac{dE_z}{dx} \tag{75}$$

Equation (75) shows that only the longitudinal field E_z (more precisely dE_z/dx) is important for transverse deflection. This is a well known result (linac theory for instance) which has a few interesting corollaries. For instance, one cannot deflect a beam neither with a pure TEM wave nor with a pure H mode in a cavity if the end effects are neglected. A transverse kicker must show a longitudinal electric field, in the same way as a transverse pick-up extracts energy from the longitudinal velocity of the particles. There is complete equivalence between pick-ups and kickers even in the transverse plane. This will be illustrated in the following example.

The "TEM" travelling-wave kicker has the same geometry as the transverse directional coupler pick-up (Fig. 31a). The field is that of a TEM wave along the two lines, except at the two ends where a longitudinal component E_z exists (Fig. 31b). Assume, for simplicity $v_p = v_\varphi$: the particles receive successively two opposite transverse kicks at either end of the kicker, the result being a zero deflection (another way of saying the same thing is that the electric and magnetic deflections along the line exactly cancel each other). On the contrary, for $v_p = -v_\varphi$ (beam in the opposite direction) the two kicks add exactly if they are separated by half a period of the RF wave $(\ell = \lambda / 4)$. This gives a variation of the type $\sin 2\pi \ell/\lambda$.

More generally the transverse kicker sensitivity defined by:

$$K_\perp = \frac{1}{V_k} \frac{v_p}{e} \Delta p_\perp \tag{76}$$

where V_k is the voltage applied to the kicker, can be related to the pick-up sensitivity S_Δ of the same structure.

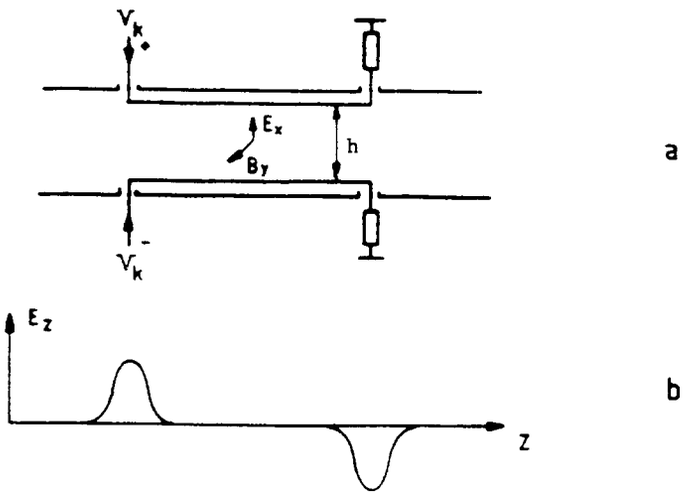


Fig. 31 "TEM" travelling-wave kicker

Combining Eqs. (70) and (75), gives:

$$\left| \frac{v}{e} \Delta p_x \right| = \int_z \frac{v_p}{\omega} \frac{dE_z}{dx} dz \ . \tag{77}$$

When applying the reciprocity theorem to a transverse pick-up, we already obtained (Eq. (43), modified for the transverse case):

$$I_1 V_{out} = \int_z \frac{dE_z}{dx}\, i_b\ \Delta_x\, dz \tag{78}$$

I_1 being the current source producing the field E_z. For a matched pick-up or kicker, I_1 is split equally between the structure and the load resistor, which gives:

$$V_k = I_1\ R_0 / 2\ . \tag{79}$$

Combining Eqs. (77), (78), (79) and (33) one obtains:

$$K_\perp = 2\frac{v_p}{\omega}\frac{S_\Delta}{R_0} \tag{80}$$

which relates pick-up and kicker sensitivities.

For the "TEM" travelling wave kicker, which is the same structure as the one shown in Fig. 15b, combining Eqs. (32), (34) and (80) gives:

$$K_\perp = 2j\frac{\ell}{h}\ \tanh\left(\frac{\pi w}{h}\right)\frac{\sin\ 2\pi\ \ell/\lambda}{2\pi\ \ell/\lambda}\ . \tag{81}$$

5.3 Debunched beam transfer function

The beam is composed of a collection of particles, each having its own transverse oscillation frequency $q_i\omega_i$, submitted to a common driving force $F(\omega)$. The equation of motion, for each individual particle is, in linear approximation:

$$\ddot{x}_i + (q_i\omega_i)^2 x = F(\omega) \tag{82}$$

with a forced solution of the form:

$$\begin{aligned} x_i &= X_i\ \exp\ j\omega t \\ X_i &= \frac{F(\omega)}{(q_i\omega_i)^2 - \omega^2} \cong \frac{1}{2\omega}\frac{F(\omega)}{\omega_i q_i - \omega}\ . \end{aligned} \tag{83}$$

The average beam response $\langle X_i \rangle / F(\omega)$ is given by the integral:

$$\frac{\langle X_i \rangle}{F(\omega)} = \frac{1}{2\omega}\int_{q_1\omega_1}^{q_2\omega_2} \frac{\rho(q_i\omega_i)}{q_i\omega_i - \omega} d(q_i\omega_i) \tag{84}$$

where $\rho(q_i\omega_i)$ is the normalised distribution of the betatron frequencies within the beam, ($q_1\omega_1$ and $q_2\omega_2$ being the two extreme frequencies).

This is a singular integral because of the pole at $q_i\omega_i = \omega$. It can be decomposed into its Cauchy principal value, which is real, and its residue at the pole (imaginary):

$$\frac{\langle X_i \rangle}{F(\omega)} = \frac{1}{2\omega}[\text{Princ. Value} - j\pi\rho(\omega)] \ . \tag{85}$$

We now replace <X_i> by $j\omega\langle X_i \rangle$ to obtain a real transfer function $B(\omega)$ when energy is absorbed (force and displacement in quadrature) and obtain:

$$B(\omega) = \frac{1}{2}(\pi\rho(\omega) + j \ \text{Princ. Value}) \ . \tag{86}$$

The real part of the transfer function gives the particle distribution in tune like the spectral power density of the Schottky signal. Outside the frequency band $(q_1\omega_1, q_2\omega_2)$ the real part of $B(\omega)$ vanishes (pure imaginary response). The fact that a collection of lossless oscillators responds like a damped resonator is the basis of Landau damping and is illustrated in Fig. 32.

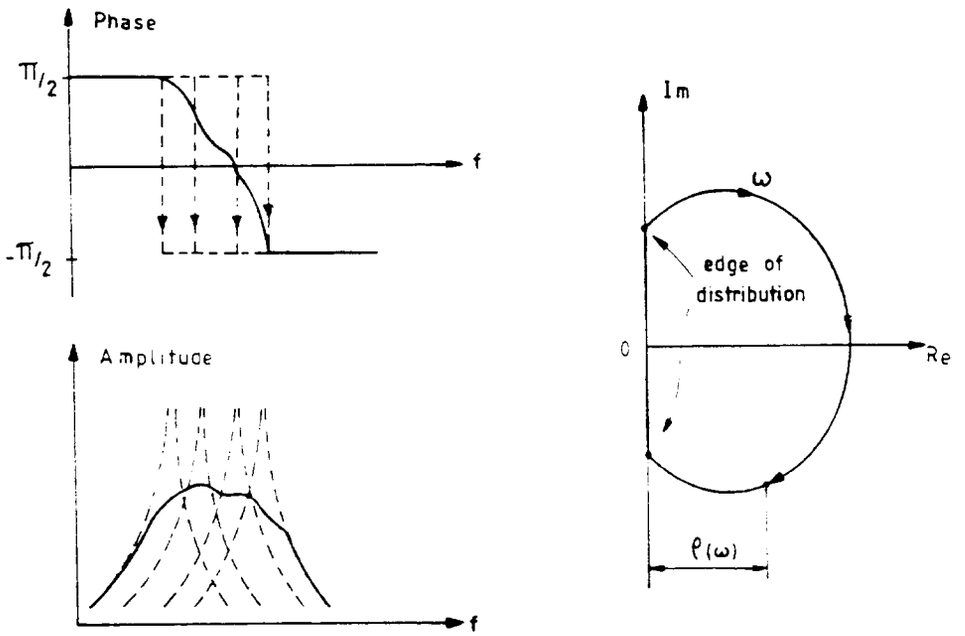


Fig. 32 Response of a large number of lossless resonators.
--- individual particles, ___ average

The evaluation of the stability of the beam certainly corresponds to the most interesting application of beam transfer function measurements. Collective effects (and in particular beam instabilities) result from the presence of parasitic impedances in the machine which generate a deflecting force (in the transverse case), when excited by a collective displacement of the beam. In other words the excitation $F(\omega)$ in Eq. (82) should be combined with a term proportional to the beam response $j\omega\langle X_i \rangle$. This leads to the well known feedback loop of Fig. 33, where $H(j\omega)$ is linked to machine parameters and is proportional to the impedance of the machine $Z(\omega)$. For instance in the transverse case:

$$H(j\omega) = j\frac{e\omega_0 i_b}{2\pi m_0 c\gamma} Z(\omega) \tag{87}$$

m_0 is the rest mass of the particle.

From Fig. 33 the new transfer function becomes:

$$B'(\omega) = \frac{j\omega\langle X_i\rangle}{F} = \frac{B(\omega)}{1 - B(\omega)H(\omega)} \tag{88}$$

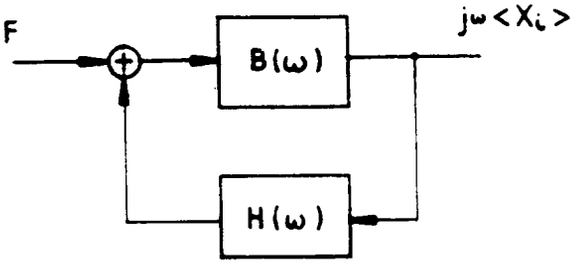


Fig. 33 Feedback loop due to the machine impedance

$$\frac{1}{B'(\omega)} = \frac{1}{B(\omega)} - H(\omega)\ . \tag{89}$$

By plotting the curve $1/B(\omega)$ for different beam intensities i_b one obtains a family of curves shifted in the complex plane by the quantity $H(\omega)$ (Fig. 34). This shift being proportional to $Z(\omega)$, the machine impedance can be directly measured at any frequency [11].

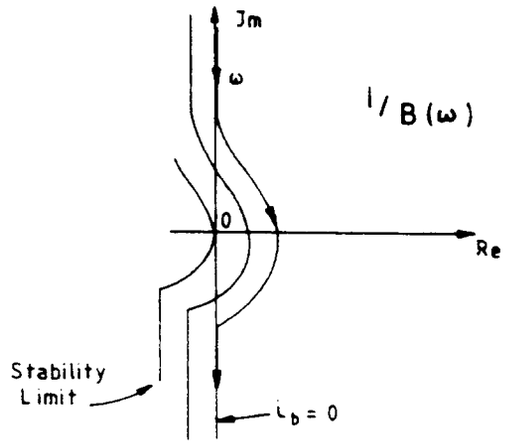


Fig. 34 Evaluation of the beam stability with transfer function measurements

When the shifted $1/B(\omega)$ curve reaches the complex plane origin, stability of the beam is lost $(B(\omega) \to \infty)$, this means that the distance of the curve to the origin is a measure of beam stability. If a feedback system is employed to stabilise the beam, its effect which should be to shift the curve towards the right side of the complex plane could also be evaluated.

With very sensitive detectors and provided long analysis times are available (DC storage rings), beam transfer function is a very powerful technique, almost non-disturbing to the beam; it can also be used in a similar way for the longitudinal plane.

5.4 Bunched-beam transfer function

The main difference with respect to the unbunched beam case is that an excitation of the beam at a given frequency ω, not only results in a beam response at ω, but also at all frequencies $n\omega_0 \pm \omega$. (This is because the bunched beam samples the ω waveform at the revolution frequency ω_0.) The process is therefore fundamentally nonlinear, and as a consequence, the beam transfer function is not defined in general, unless additional conditions are imposed [12]. For instance, if bunch-to-bunch coupling can be neglected, one can define unambiguously the beam transfer function of a single bunch, for a given mode of oscillation (dipole, quadrupole, etc.), i.e. within an f_b frequency interval. Another interesting case is when the bunched beam behaves like an unbunched beam: many equal bunches, frequency range from DC up to $f_{RF}/2$ and negligible effects beyond.

In the transverse plane, the measurement of the machine tune is nothing but a beam transfer function measurement. Many descriptions of tune measurement systems exist in the literature; excitation can be sinusoidal or random (band-limited noise) near a betatron line, or pulsed; beam measurement could be at the same or at different frequency. In general the machine impedance $Z(\omega)$ cannot be measured directly, as a function of frequency; on the other hand if the shape of $Z(\omega)$ is known (e.g. resistive wall) one can determine its magnitude by measuring the tune shift as a function of beam intensity.

The RF system and its associated feedback loops strongly perturbs the longitudinal transfer function of a bunched beam. This is particularly true for the dipole mode; fortunately the quadrupole mode is easier to analyse and can provide meaningful measurements of the machine impedance. Amplitude modulation of the RF waveform at around twice the synchrotron frequency excites the quadrupole mode of a single bunch; the quadrupole oscillation can be observed in a very simple way by peak detecting the bunch signal from a wide-band longitudinal detector.

The measured beam transfer function, at low intensity shows a sharp phase discontinuity, at the bunch centre, where the particle density is maximum, and a smooth phase curve near the bunch edge (Fig. 35a). This corresponds to the $1/B(\omega)$ plot in Fig. 35b and provides a direct measurement of the center synchrotron frequency. At higher intensities, the inductive wall effect shifts the $1/B(\omega)$ curve along the imaginary axis (real frequency shift) and the phase curve of Fig. 35a shows a sharper transition. From those measurements, the magnitude of $Z(\omega)/n$ for the inductive-wall case can be determined over a frequency interval of the order of f_b.

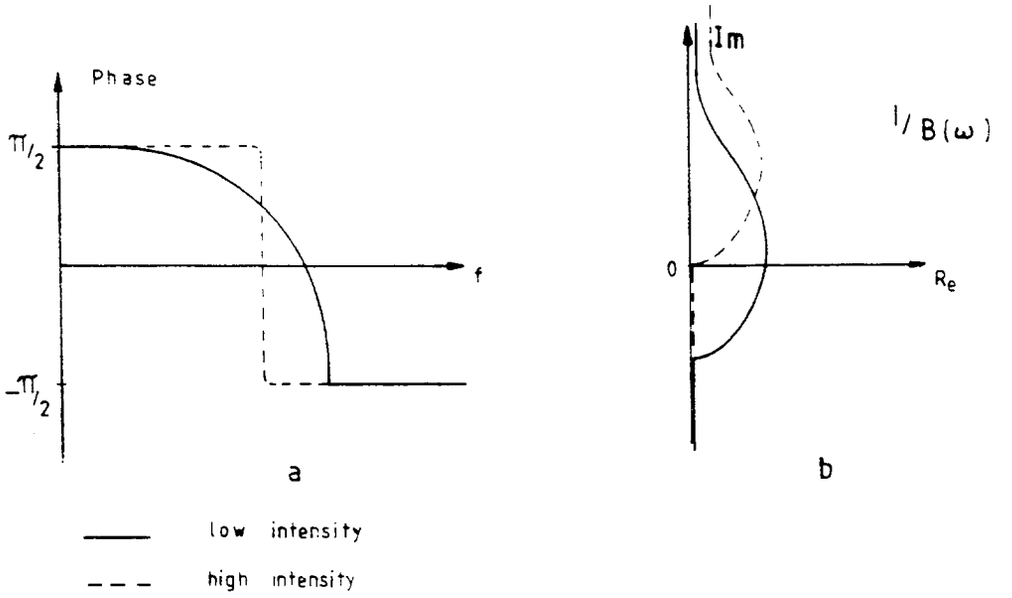


Fig. 35 Quadrupole-mode transfer function (bunched beam)

REFERENCES

[1] J. Borer, P. Bramham, H.G. Hereward, K. Hübner, W. Schnell, L. Thorndahl, Non-destructive diagnostics of coasting beams with Schottky noise. IXth Int. Conf. on High Energy Accelerators, SLAC, May 1974 (SLAC, Stanford, 1974).

[2] H.G. Hereward, W. Schnell, Statistical phenomena, Proc. of the first course of the Int. School of particle Accelerators, Erice, Sicily, 1976 (CERN 77-13, 1977).

[3] T. Linnecar, The high frequency longitudinal and transverse pick-ups used in the SPS, CERN SPS/ARF/78-17 (1978).

[4] J. Borer, R. Jung, CERN Accelerator School, Antiprotons for colliding beam facilities, CERN, 1983 (CERN 84-15, 1984).

[5] T. Linnecar, W. Scandale, A transverse Schottky noise detector for bunched proton beams, IEEE Trans. Nucl. Sci., NS-28, p. 2147 (1981).

[6] D. Boussard, G. Di Massa, High frequency slow wave pick-ups, CERN SPS/86-4 (1986).

[7] F. Caspers, Planar slotline pick-ups and kickers, CERN PS/AA Note 85-46 (1985).

[8] D. Boussard, Evaluation of slot line pick-up sensitivity, CERN SPS/ARF Note 85-9 (1985).

[9] D. Boussard, T. Linnecar, W. Scandale, Recent developments on Schottky beam diagnostics at the CERN SPS collider, IEEE Trans. Nucl. Sci. NS-32, p. 1908 (1985).

[10] D. Boussard, S. Chattopadhyay, G. Dome, T. Linnecar, Feasibility study of stochastic cooling of bunches in the SPS, CERN Accelerator School, Antiprotons for colliding beam facilities, CERN, 1983 (CERN 84-15, 1984).

[11] J. Borer et al., Information from beam responses to longitudinal and transverse excitation, IEEE Trans. Nucl. Sci. NS-26, p. 3405 (1979).

[12] S. Chattopadhyay, Some fundamental aspects of fluctuations and coherence in charged particle beams in storage rings, CERN 84-11 (1984).

SYNCHROTRON RADIATION AND EQUILIBRIUM BEAM SIZES

A. Ropert
ESRF, Grenoble, France

Abstract
Synchrotron radiation is one of the specific features of electron accelerators and storage rings. The main properties of this radiation such as radiated power, spectrum of the radiation, photon distribution, can be derived from the calculation of the electromagnetic fields from moving particles using classical theory. The main consequences of the synchrotron radiation energy losses on the dynamics of charged particles (radiation damping, equilibrium beam dimensions...) are outlined. The impact of synchrotron radiation on machine design is also discussed.

1. INTRODUCTION

Charged particles emit an electromagnetic radiation when they are accelerated transversally in a magnetic field. This radiation is known as synchrotron radiation. This radiation depends on the angle between the field and the particle's velocity.

The main difference between electrons and protons (or other heavy particles) is related to the synchrotron radiation energy loss since the lighter are the particles the more intense is the radiation. This means that this phenomena is negligible for protons up to the 100 GeV range but that it is an essential mechanism for electrons or positrons.

Synchrotron radiation has a significant impact on the dynamics of particles. Longitudinal or transverse oscillations are damped or antidamped. The energy is emitted as quanta of discrete energy and generates a kind of stochastic excitation in the beam, which leads to a diffusion phenomenon. The equilibrium beam dimensions depend on the two previous processes.

The effects of radiation also have significant consequences on the design of the machine and its components, in particular influencing the size of the accelerator. The following figures (5 MeV energy loss per turn for the 6 GeV ESRF and 260 MeV for LEP at 55 GeV) show the importance of the losses. The design of the RF system required to compensate for the energy loss might also give a limitation on the maximum energy attainable. On the engineering side, the vacuum system must be designed to cope with the gas desorption due to the impinging photons. The evacuation of the heat load deposited by the radiation in the vacuum chamber is also a major concern.

Synchrotron radiation was observed for the first time in an electron synchrotron and considered as a parasitic effect by the high energy users up to the 1975's. It has now a wide range of applications: basic research in atomic, molecular, solid state fields as well as in biology and other fields. Dedicated synchrotron radiation sources are now being built in many places all around the world for exploiting this domain of reserach. Beam diagnostics in accelerators often use the properties of the synchrotron radiation.

In the first part of this paper, the basic properties of synchrotron radiation are derived, using the classical electromagnetic theory. Then some of the relevant effects on the dynamics of particles will be considered.

Numerous reports and textbooks deal with the theory of synchrotron radiation. Most of the material used in this lecture can be found in the following list of references [1, 8]. This list is by far non-exhaustive.

2. LIENARD-WIECHERT POTENTIALS

2.1 Basic principles of the calculations

The objective is to compute the electromagnetic fields and potentials of a moving charge. Fields E and H are measured in the laboratory frame and are therefore time dependent. The particle is moving in this frame. For simplicity, it is convenient to compute the fields in a frame which is fixed with respect to the particle but moves with time.

We therefore consider two referentials:

i) the laboratory frame in which the observer P observes the electromagnetic fields at the time t. The position of the observer is given by r_p (Fig. 1).

ii) the one of the moving particle q which emitted the radiation at the time t'. The position of the charge is given by the vector R. The distance between the point of emission and the observer is given by the vector r, which fulfils the relation:

$$R\,(t') + r\,(t') = r_p \tag{1}$$

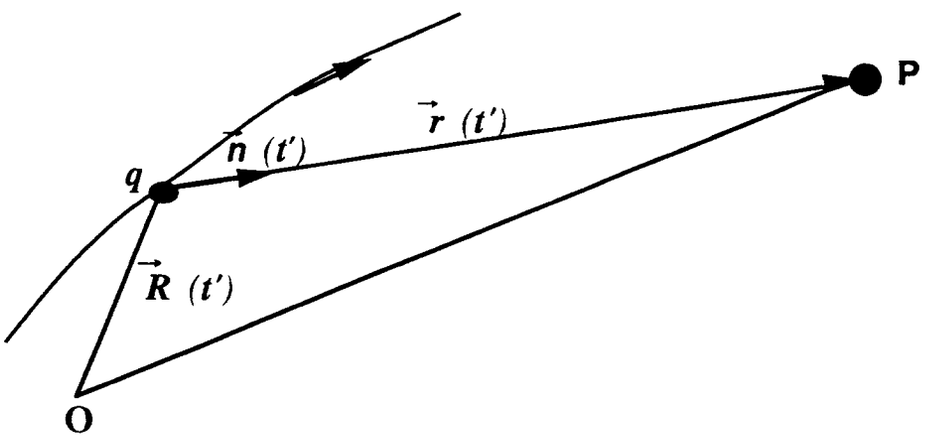


Fig.1 Geometry of the observer and the particle

The time difference $t - t'$ between emission and observation is the time it takes the radiation to travel the distance between the point of emission and the charge:

$$t - t' = \frac{r\,(t')}{c} \quad \text{with } \mathrm{r}\,(t') = r_P - R\,(t') \tag{2}$$

The motion of the charge q and its velocity v at time t' are given by:

$$R(t') \text{ and } \mathrm{v}\,(t') = \beta(t')\ c = \frac{dR(t')}{dt'} \tag{3}$$

Since the observer does not move, $r_P = \mathrm{C^{te}}$. Therefore, by differentiating Eq. (1), one gets:

$$\frac{dR\,(t')}{dt'} = v\,(t') = c\,\beta = -\,\frac{dr\,(t')}{dt'} \tag{4}$$

In order to get the time derivative of $r\ (t')$, we define the unit vector pointing to the observer and we write:

$$\vec{r}\frac{dr\ (t')}{dt'} = \frac{1}{2}\frac{d\ (\overrightarrow{r^2})}{dt'} = r\frac{dr}{dt'} = -\vec{r}\ .\ \vec{v}\ ;\ \frac{dr}{dt'} = -\ n\ .\ v \tag{5}$$

The relation between the time at the observer and the time of emission is then derived:

$$dt = \left(1 + \frac{1}{c}\frac{dr}{dt'}\right) dt' = (1 - n\ \beta)\ dt' \tag{6}$$

To illustrate the importance of this formula, let's take the example of the radiation emitted by a relativistic particle in the forward direction. Two photons received by the observer within a small time interval Δt have been emitted within a much longer interval $\Delta t'$.

2.2 Calculations of the fields

In order to calculate the electric and magnetic fields $E(t)$ and $B(t)$ of the radiation received by the observer, we will start from Maxwell equations:

$$c^2\ rot\ \vec{B} = \frac{\vec{j}}{\varepsilon_0} + \frac{\partial \vec{E}}{\partial t} \tag{7}$$

The fields are deriving from scalar and vector potentials φ and A through:

$$\vec{B} = rot\ \vec{A} \tag{8}$$

$$\vec{E} = -grad\ \varphi - \frac{\partial}{\partial t}\vec{A}$$

When φ and A are given, E and B are determined in a unique way; however different potentials could correspond to a unique field. Therefore, we must introduce an extra condition, by using the Lorentz convention:

$$\text{div}\ \vec{A} + \frac{1}{c^2}\frac{\partial \varphi}{\partial t} = 0 \tag{9}$$

We obtain the wave equations:

$$\Delta \vec{A} - \frac{1}{c^2}\frac{\partial^2 \vec{A}}{\partial t^2} = -\frac{1}{c^2}\frac{\vec{j}}{\varepsilon_0} \quad \text{and} \quad \Delta \phi - \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\varepsilon_0} \tag{10}$$

In the case of a stationary charge distribution with a charge density $\rho\ (x, y, z)$, the resulting electrostatic potential is simply given by Coulomb's law and the vector potential A is obtained in the same way:

$$\phi\ (\vec{x}) = \int \frac{1}{4\ \pi\ \varepsilon_0}\frac{\rho(\vec{x})}{r}\ dx\ dy\ dz \tag{11}$$

$$A(\vec{x}) = \int \frac{1}{4\pi\varepsilon_0 r} \frac{v\,\rho(\vec{x})}{r} \; dx\,xy\,dz$$

where r is the distance between the charge and the observer. If the charge moves, the charge density ρ becomes a function of time. In order to compute the potentials of the moving charge, we need to evaluate the charge density at the earlier time t' such that the potential created by the charge reaches the observer at the time t. By using the Green's function of the wave equations, we obtain:

$$\phi(\vec{x}, t) = \int \frac{1}{4\pi\varepsilon_0} \frac{\rho(\vec{x}, t')}{r(\vec{x}, t')} \; \delta\left[t' - (t - \frac{r}{c})\right] dx\,dy\,dz\,dt' \tag{12}$$

$$\vec{A}(\vec{x}, t) = \int \frac{1}{4\pi\varepsilon_0 c^2} \frac{\vec{j}(\vec{x}, t')}{r(\vec{x}, t')} \; \delta\left[t' - (t - \frac{r}{c})\right] dx\,dy\,dz\,dt'$$

Equations (12) are called the retarded potentials. In order to obtain the Lienard - Wiechert potentials, we will use the approximation that, in the domain occupied by the charge, the distance r between the charge and the observer can be considered as a constant, which makes the integration with respect to the spacial variables easier. We then introduce:

$$\theta(t') = t' + \frac{r(t')}{c} \tag{13}$$

as a new variable and we integrate the two Eqs (12):

$$\phi(\vec{x}, t) = \frac{e}{4\pi\varepsilon_0} \left[\frac{1}{r(1 - \vec{n}\vec{\beta})} \right]_{ret} \tag{14}$$

$$\vec{A}(\vec{x}, t) = \frac{e}{4\pi\varepsilon_0 c} \left[\frac{\vec{\beta}}{r(1 - \vec{n}\vec{\beta})} \right]_{ret}$$

$[\ ]_{ret}$ has to be evaluated at the time t' in order to get the potential at the later time $t = t'' + \frac{r(t'')}{c}$. For a non relativistic particle ($1 - \vec{n} \cdot \vec{\beta}$ --> 1), Eq. (14) yields the classical non-relativistic results.

The fields of the moving charge at the observer location are obtained by introducing Eq. (14) into Eq. (1):

$$B(\vec{x}, t) = \frac{e}{4\pi\varepsilon_0 c} \,\mathrm{rot} \left[\frac{\vec{\beta}}{r(1 - \vec{n}\vec{\beta})} \right]_{ret} \tag{15}$$

$$\vec{E}(\vec{x},t) = -\frac{e}{4\pi\varepsilon_0}\,\mathrm{grad}\left[\frac{1}{r(1-\vec{n}\vec{\beta})}\right]_{ret} - \frac{e}{4\pi\varepsilon_0 c}\frac{\partial}{\partial t}\left[\frac{\vec{\beta}}{r(1-\vec{n}\vec{\beta})}\right]_{ret}$$

An electron with a trajectory $R(t')$ which is observed at a time t has emitted an electromagnetic wave at the former time t' which depends on the position of the observer r. Rather tedious calculations are necessary to establish the relationships which enable the quantities in brackets (computed at time t') to be evaluated and expressed at the time t of the observer. The detailed derivation is beyond the scope of this lecture. The final results are as follows:

$$\vec{E}(\vec{x},t) = \frac{e}{4\pi\varepsilon_0 r^2}\frac{(\vec{n}-\vec{\beta})(1-\beta^2)}{(1-\vec{n}\vec{\beta})^3} + \frac{e}{4\pi\varepsilon_0 r c}\frac{\vec{n}\wedge\left[(\vec{n}-\vec{\beta})\wedge\dot{\vec{\beta}}\right]}{(1-\vec{n}\vec{\beta})^3} \qquad (16)$$

$$\vec{B}(\vec{x},t) = \frac{e}{4\pi\varepsilon_0 c r^2}\frac{\vec{n}\wedge(\vec{n}-\vec{\beta})}{\gamma(1-\vec{n}\vec{\beta})^3} + \frac{e}{4\pi\varepsilon_0 r c^2}\frac{\vec{n}\wedge\left\{\vec{n}\wedge\left[(\vec{n}-\vec{\beta})\wedge\dot{\vec{\beta}}\right]\right\}}{(1-\vec{n}\vec{\beta})^3}$$

Some important properties can be drawn from Eq. (16):

i) $\vec{B} = \frac{1}{c}(\vec{n}\wedge\vec{E})$, which shows that the fields E and B are perpendicular to each other and to the the unit vector n pointing to the observer

ii) for a stationary charge, $\beta = \dot{\beta} = 0$ and we find again Coulomb's law $\vec{E} = \frac{e}{4\pi\varepsilon_0 r^2}\vec{n}$

iii) the general expression given in Eq. (16) has two terms. The first one is proportional to $1/r^2$ and characterizes the acceleration independent part of the field. The second term depends on $1/r$. At large distances (which is the typical case of observation of synchrotron radiation), the second term will dominate and the field can be approximated as:

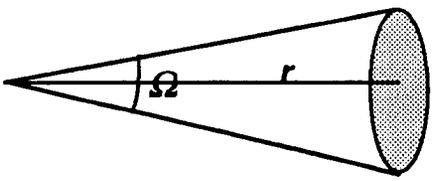
$$\vec{E}(\vec{x},t) = \frac{e}{4\pi\varepsilon_0 r c^2}\frac{\vec{n}\wedge\left[(\vec{n}-\vec{\beta})\wedge\dot{\vec{\beta}}\right]}{(1-\vec{n}\vec{\beta})^3} \qquad (17)$$

The remaining term is proportional to the acceleration $\dot{\beta}$ and perpendicular to the direction of the unit vector n pointing from the charge to the observer. From now on we will consider only this field which is often called the "far field".

3. POWER RADIATED BY AN ACCELERATED CHARGE

3.1 Definitions

The power lost by a charged and moving particle is radiated in the whole space. The energy radiated by unit time and by unit surface is characterized by the Poynting's vector S:



$$\vec{S} = \frac{1}{\mu_0} \vec{E} \wedge \vec{B} = \frac{dW}{ds\, dt} \tag{18}$$

Using the previously established relation $\vec{B} = \frac{1}{c} (\vec{n} \wedge \vec{E})$, we get

$$\vec{S} = \frac{1}{\mu_0 c} \vec{E} \wedge (\vec{n} \wedge \vec{E}) = \frac{1}{\mu_0 c} \left[\vec{n} E^2 - \vec{E} (\vec{E} \,.\, \vec{n}) \right] \tag{19}$$

In the case of the "far field", the electric field is perpendicular to the unit vector n and we therefore obtain:

$$\vec{S} = \frac{1}{\mu_0 c} E^2 \vec{n} \tag{20}$$

The Poynting vector is parallel to the vector n. By integrating the Poynting's vector over a sphere around the particle, we obtain the power radiated by the particle:

$$P = \frac{dW}{dt'} = \int (n \,.\, S) \frac{dt}{dt'} r^2 \, d\Omega = \int S^2 (1 - n \,.\, \beta) \, r^2 \, d\Omega \tag{21}$$

3.2 Radiated power

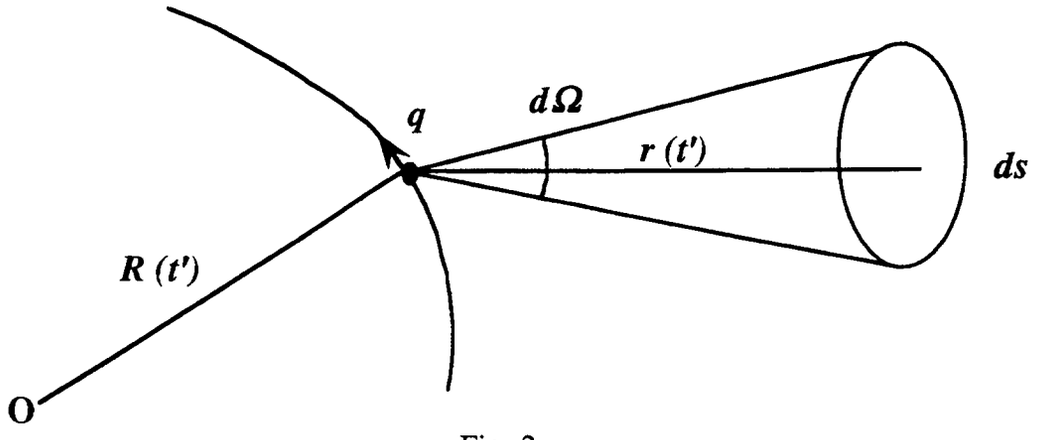


Fig. 2

In order to compute $\frac{dP}{d\Omega}$ which is the instantaneous power per unit of solid angle in the frame of the particle, we start with:

$$S = \frac{dW}{r^2 \, d\Omega \, dt} = \frac{1}{r^2} \frac{dW}{d\Omega \, dt'} \frac{dt'}{dt} = \frac{1}{r^2} \frac{dP}{d\Omega} \frac{1}{1 - \vec{n} \,.\, \vec{\beta}} = \varepsilon_0 \, c \, E^2 \tag{22}$$

Replacing E by its expression (17) yields:

$$\frac{dP}{d\Omega} = \frac{e^2}{(4\pi)^2 \varepsilon_0 c} \frac{\left[\vec{n} \wedge \left((\vec{n} - \vec{\beta}) \wedge \dot{\vec{\beta}}\right)\right]^2}{(1 - \vec{n} \cdot \vec{\beta})^5} \qquad (23)$$

3.2.1 General case

In the general case, the acceleration is perpendicular to the velocity and pointing in the $-x$ direction. This corresponds to the case of synchrotron radiation emitted by a particle going through a magnetic field B pointing in the direction y. The geometry of the problem is indicated in Fig. 3.

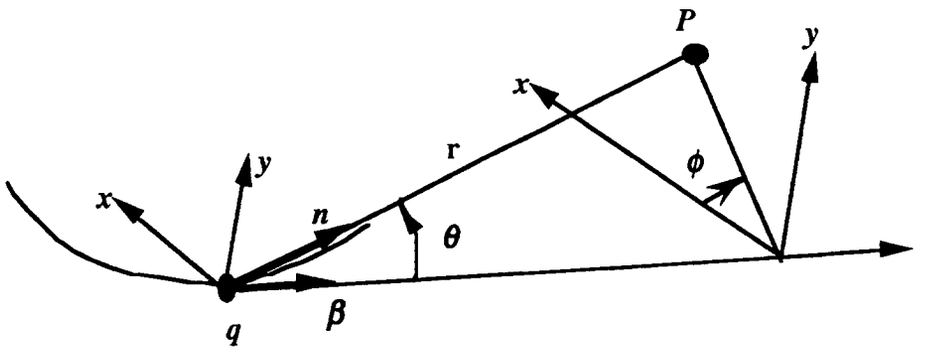


Fig. 3 Definition of angles θ and ϕ

In the particle frame, the coordinates of the three vectors (normalized velocity, normalized acceleration and unit vector pointing to the observer are expressed as:

$$\vec{\beta} \ (0, 0, 1)$$

$$\dot{\vec{\beta}} \left(-\frac{\beta^2 c}{\rho}, 0, 0\right)$$

$$\vec{n} \ (\sin\theta \cos\phi \ , \ \sin\theta \sin\phi \ , \ \cos\theta)$$

ρ being the curvature in the magnetic field: $\frac{1}{\rho} = = \frac{e B}{\beta \gamma m_0 c}$

The angular distribution of the power radiated by the particle is then given by:

$$\frac{dP}{d\Omega} = \frac{e^2}{(4\pi)^2 \varepsilon_0 c} \frac{\dot{\beta}^2}{(1 - \beta\cos\theta)^3} \left[1 - \frac{(1 - \beta^2)}{(1 - \beta\cos\theta)^2} \sin^2\theta \cos^2\phi\right] \qquad (24)$$

with using the classical radius $r_0 = \frac{e^2}{4\pi\varepsilon_0 m_0 c^2} = 2.818\ 10^{-15}$ for electrons

3.2.2 Case of a non-relativistic particle

In this section, we treat the case of a particle with $\vec{\beta} << 1$ but with any acceleration value. Starting from Eq. (23), we can easily express the different terms, with Θ being the angle

between $\vec{\dot{\beta}}$ and $\vec{n}$.

$$\frac{dP}{d\Omega} = \frac{1}{\varepsilon_0 c}\left(\frac{e\,\dot{\beta}}{4\pi}\sin\Theta\right)^2 \tag{25}$$

The total radiation power is obtained by integrating over the solid angle:

$$P = \int\int \frac{dP}{d\Omega}\,d\Omega = \frac{e^2}{6\pi\varepsilon_0 c}\,\dot{\beta}^2 \tag{26}$$

In the particle frame, the electron is non-relativistic. Since the emitted power is independent of the frame, this is also the power in the laboratory frame. We need to express $\dot{\beta}$ in the laboratory frame, which is done through a Lorentz transformation:

$$P = \frac{e^2}{6\pi\varepsilon_0 c}\,\gamma^4\left(\gamma^2\,\dot{\beta}_{//}^{\,2} + \dot{\beta}_{\perp}^{\,2}\right) \tag{27}$$

with $\dot{\beta}_{//}$ being the acceleration component parallel to $\vec{v}$ and $\dot{\beta}_{\perp}$ being the component perpendicular to $\vec{v}$.

This represents the power radiated by an accelerated electron of any speed. The first derivation was given by Lienard in the 1898's. If the acceleration $\dot{\beta}_{\perp}$ is due to the curvature of the orbit, Eq. (27) can be rewritten as:

$$P = \frac{2\,r_0}{3}\frac{E_0}{c}\,\gamma^4\left(\frac{\dot{\gamma}^2}{\beta^2} + \frac{c^2\,\beta^4\,\gamma^4}{\rho^2}\right) \tag{28}$$

with the following relations:

$$\gamma = \left(1 - \beta^2\right)^{\frac{1}{2}} \tag{29}$$

$$\dot{\gamma} = \beta\,\dot{\beta}\left(1 - \beta^2\right)^{\frac{3}{2}} = \beta\,\dot{\beta}\,\gamma^3$$

$$\dot{\beta}_{//} = \beta^{-1}\,\gamma^{-3}\,\dot{\gamma} \quad \text{and} \quad \dot{\beta}_{\perp} = \frac{\beta^2\,c}{\rho}$$

We can also write the power in terms of momentum, by differentiating $p^2\,c^2 = E^2 - E_0^2$ and using the basic relations, $\dot{p}_{//} = \frac{1}{\beta c}\,\dot{E} = \frac{1}{\beta c}\,E_0\,\dot{\gamma}$ and $\dot{p}_{\perp} = m_0\,\gamma\dot{v}_{\perp} = m_0\,\gamma c\,\dot{\beta}_{\perp}$:

$$P = \frac{2\,r_0\,c}{3\,E_0}\left(\dot{p}_{//}^{\,2} + \gamma^2\,\dot{p}_{\perp}^{\,2}\right) \tag{30}$$

Equation (30) shows that the same accelerating force $\dot{p}_{//} = \dot{p}_{\perp}$ produces a γ^2 times stronger radiation power, if applied in a perpendicular direction rather than in a parallel direction. This explains why, above a certain energy, linear accelerators are advantageous compared to storage rings for colliding electrons and positrons.

3.2.3 Case of an electron moving in a constant magnetic field

This electron moves at a constant speed and therefore $\dot{\gamma} = 0$. Equation (28) reduces to:

$$P = \frac{2}{3} r_0 \, c \, E_0 \, \beta^4 \, \frac{\gamma^4}{\rho^2} \tag{31}$$

This is the practical case prevailing in circular accelerators where only the power radiated due to the curvature of the orbit has to be considered.

3.2.4 Total power and energy loss

The emission of radiation from each circulating electron is independent from the emission by other particles because most of the particles in a bunch are in the far zone of the radiation emitted by the electron. Since the radiation is incoherent, the total power is proportional to the number N of circulating electrons.

$$P = N \, e = \frac{4 \, \pi}{3} \frac{r_0 \, (m_0 \, c^2) \, \gamma^4 \, I}{\rho \, c} \tag{32}$$

The energy loss by turn is obtained by integrating the power over the circumference:

$$U_0 = \int_0^L P \, dt = \int_0^L \frac{P}{\beta \, c} \, ds = \frac{2}{3} r_0 \, (m_0 \, c^2) \, \beta^3 \, \gamma^4 \int_0^L \frac{ds}{\rho^2} \tag{33}$$

Since ρ is a constant for an isomagnetic machine, we get:

$$U_0 = \frac{4 \, \pi}{3} r_0 \, (m_0 \, c^2) \, \frac{\beta^3 \, \gamma^4}{\rho} \tag{34}$$

For an ultra-relativistic electron, Eq. (34) becomes:

$$U_0 = 88.4 \, \frac{E^4}{\rho} \tag{35}$$

with U_0 expressed in eV, E in GeV and ρ in m.

Table 1 gives practical values of the energy losses for a few typical machines, two storage rings used for synchrotron radiation and operated with positrons (DCI at Orsay) or electrons (ESRF at Grenoble) and two colliders (LEP anf the hadron project LHC).

As already stated, synchrotron radiation is negligible for proton machines up to high energies. On the other hand, it has a strong impact on the design of electron or positron circular machines. For large machines, a powerful RF system is needed to compensate for the energy

Table 1
Energy losses for a few machines

	E (GeV)	ρ (m)	U_0 (MeV)
DCI	1.8	3.8	0.24
ESRF	6.0	25	4.8
LEP	55	3100	260
LHC	8100	3100	0.018

loss and the choice of a large bending radius is required to limit the amount of synchrotron radiation. The comparison of the RF system of the ESRF and LEP illustrates the importance of the RF system. Both machines are using the same 352.2 MHz RF cavities and 1 MW klystrons: while the ESRF needs four cavities and two transmitters, the figures are 128 cavities and 16 transmitters for LEP. There is therefore a strong tendency at high energy to replace ambient temperature copper cavities by superconducting cavities which provide a high accelerating gradient.

The design of the vacuum system is also strongly influenced by the emitted power. It requires a careful selection of vacuum materials and preconditioning of the vacuum chambers to minimize photon-induced desorption. The design of a powerful pumping system is another prerequisite. Other engineering problems are also raised: vacuum surfaces on which synchrotron radiation is incident must be water-cooled; sophisticated design of absorbers is needed to dissipate the power (for example, at the ESRF, power density of 500 W/mm^2 are involved).

3.3 Angular distribution

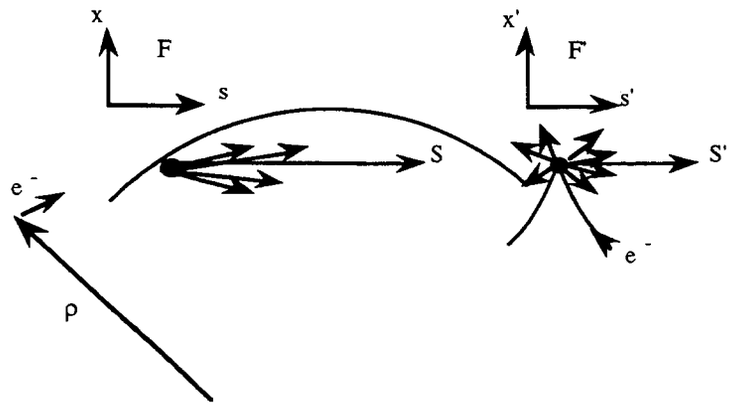


Fig. 4

We will estimate the opening angle of the radiation in a qualitative way. As shown in Fig. 4, the electron is moving in the laboratory frame F on a circular orbit. When going to a moving frame F' which moves at one instant with the same velocity as the particle, the motion in this frame is mainly transverse and the trajectory of the particle is a cycloid. In the moving frame, the electron emits radiation more or less uniformely as indicated by the Poynting' vector S'. Going back to the laboratory frame, because of Lorentz transformation, an emission angle of 90° in the moving frame is transformed into an opening angle $\theta \sim 1/\gamma$ with respect to the direction of motion. The radiation is peaked forward and contained mainly in a cone with an opening angle of $\sim 1/\gamma$. Therefore for ultra-relativistic particles with $\gamma >> 1$, the radiation is confined in very small angles around the direction of motion. For instance the numerical value for the ESRF is 85 μrad.

3.4 Frequency spectrum

Due to the small opening angle, the observer can detect only the radiation emitted from the part of the electron trajectory which has an angle smaller than $1/\gamma$ with respect to the direction of observation.

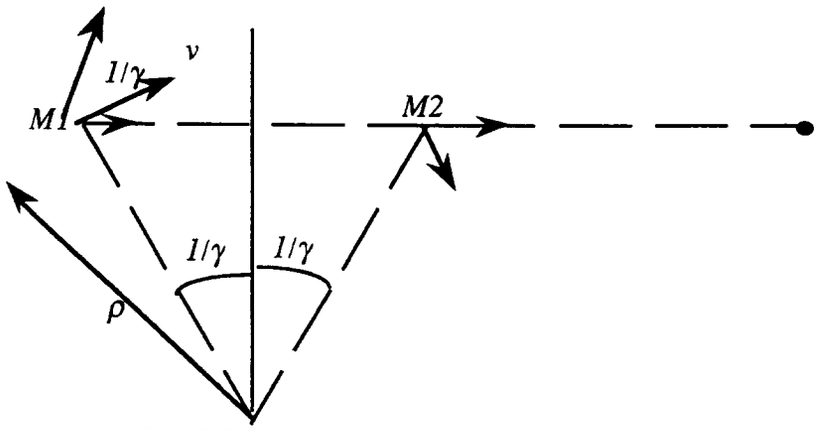


Fig. 5 Length of the radiation pulse

As indicated in Fig. 5, the length of the radiation pulse received by the observer is equal to the difference in arrival time of photons emitted at M1 and M2. This is also the difference in travel time between the electron and the photon going from M1 to M2.

$$\Delta t = t_e - t_\gamma = \frac{2\rho}{\beta c \gamma} - \frac{2\rho \sin\left(\frac{1}{\gamma}\right)}{c} \tag{36}$$

We consider only the case of ultra-relativistic electrons with $\gamma >> 1$:

$$\Delta t \cong \frac{2\rho}{\beta c \gamma}\left(1 - \beta + \frac{\beta}{6\gamma^2}\right) \cong \frac{2\rho}{\beta c \gamma}\left(\frac{1}{\gamma^2}\right) + \frac{2\rho}{6\gamma^3 c} \tag{37}$$

$$\Delta t \cong \frac{4\rho}{3 c \gamma^3}$$

The spectrum of the radiation is wide (see Fig. 6) with a typical frequency:

$$\omega_c \sim \frac{3}{2}\frac{c\gamma^3}{\rho} \tag{38}$$

where ω_c is called the critical frequency. It is also often convenient to express the spectrum of the radiation in terms of critical wavelength:

$$\lambda_c = \frac{2\pi c}{\omega_c} = \frac{18.64}{B E^2} \tag{39}$$

using the following units: λ_C in Å, B in T, E in GeV

As an example, for the 6 GeV ESRF synchrotron, the radiation emitted from the main bending magnets (B = 0.8 T, ρ = 25 m) consists of X-rays of critical energy 20 keV.

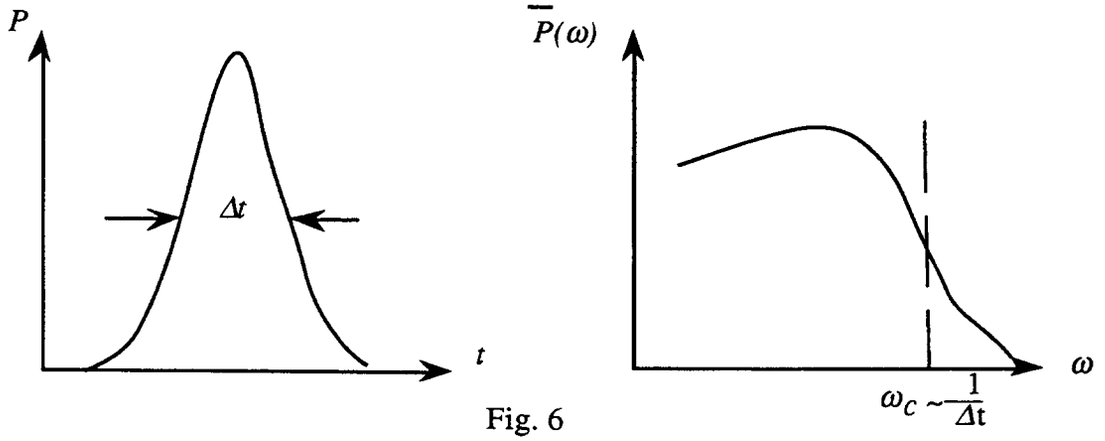


Fig. 6

3.5 Spectral angular power distribution of the radiation

The observer receives a flash of radiation which determines the form of the spectrum. In order to analyze the signal in the frequency domain, we compute the total energy received by the observer during a single passage of the particle, which consists of integrating the energy per unit solid angle:

$$W = \varepsilon_0 c \int_0^{4\pi} \int_{-\infty}^{+\infty} E^2 r^2 \, d\Omega \, dt \tag{40}$$

The field $E(t)$ is obtained from the Fourier transformed field $E(\omega)$:

$$E(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathcal{E}(\omega) \, e^{i\omega t} \, d\omega \tag{41}$$

Using the integral representation of the Dirac function, we get:

$$W = \varepsilon_0 c r^2 \int d\Omega \int_{-\infty}^{+\infty} \mathcal{E}(\omega) \, \mathcal{E}(\omega') \, \delta(\omega + \omega') \, d\omega \, d\omega' \tag{42}$$

Since $E(t)$ is a real function, its Fourier transform is symmetrical and we obtain:

$$W = 2 \varepsilon_0 c r^2 \int d\Omega \int_0^{+\infty} |\mathcal{E}(\omega)|^2 \, d\omega \tag{43}$$

The spectral angular distribution can then be derived:

$$\frac{d^2W}{d\Omega \, d\omega} = 2 \varepsilon_0 c r^2 |\mathcal{E}(\omega)|^2 \tag{44}$$

From the computation of $\mathcal{E}(\omega)$, one gets:

$$\frac{d^2W}{d\omega\, d\Omega} = \frac{e^2}{12\,\pi^3\,\varepsilon_0\,c}\left(\frac{\omega\rho}{c}\right)^2 \frac{(1+\gamma^2\,\theta^2)^2}{\gamma^4}\left[K^2_{\frac{2}{3}}(\xi) + \frac{\gamma^2\,\theta^2}{1+\gamma^2\,\theta^2} K^2_{\frac{1}{3}}(\xi) \right] \quad (45)$$

with $\xi = \frac{\omega}{\omega_c}\left(1+\gamma^2\,\theta^2\right)^{\frac{3}{2}}$ being the argument of the modified Bessel function K; θ is the angle of observation with respect to the speed and ω_c is the critical frequency defined by Eq. (38). For $\xi >> 1$, $K \approx \frac{e^{-\xi}}{\xi}$, therefore the radiation is confined in a decreasingly smaller range in θ for increasing frequencies, as shown in Fig. 7.

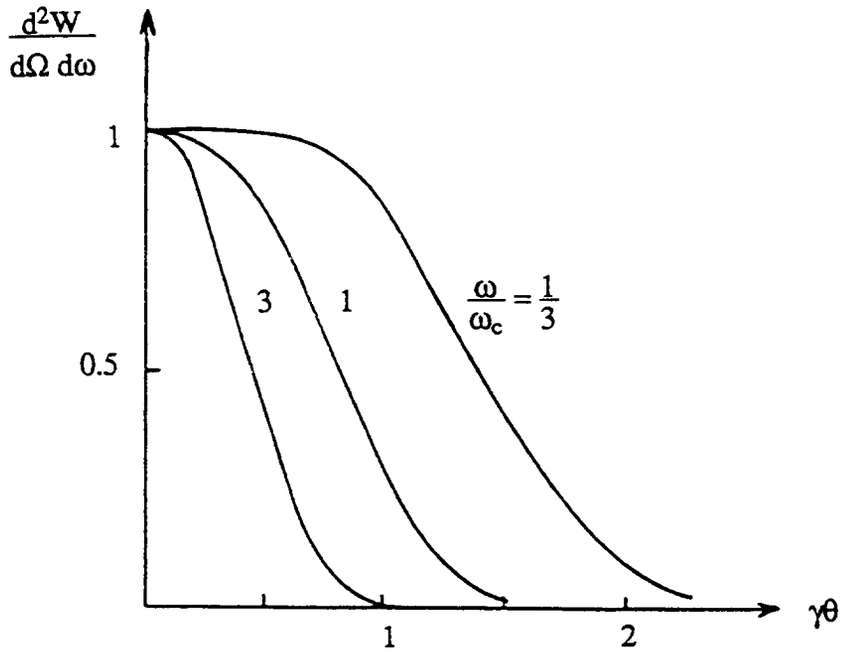


Fig. 7 Energy received by the observer

Equation (45) shows that the radiation is polarized. The first term corresponds to the polarization in the plane of the orbit, whilst the second one is related to the polarization in the plane perpendicular to the orbit. Integrating over all frequencies yields:

$$\frac{dW}{d\Omega} = \frac{7\,e^2}{64\,\pi\,\varepsilon_0}\,\frac{1}{\rho}\,\frac{\gamma^5}{(1+\gamma^2\,\theta^2)^{\frac{5}{2}}}\left[1+\frac{5}{7}\,\frac{\gamma^2\,\theta^2}{(1+\gamma^2\,\theta^2)^2} \right] \quad (46)$$

Figure 8 shows the relative contributions of the two components of the polarization. The polarization in the plane of the orbit is the dominant term.

4. RADIATION DAMPING

4.1 Energy oscillations

Synchrotron radiation has important consequences on particle dynamics. For electrons in an accelerator or a storage ring, γ is large and dipoles give an accelerating force which is perpendicular to the speed. Because of the losses due to synchrotron radiation, electrons would

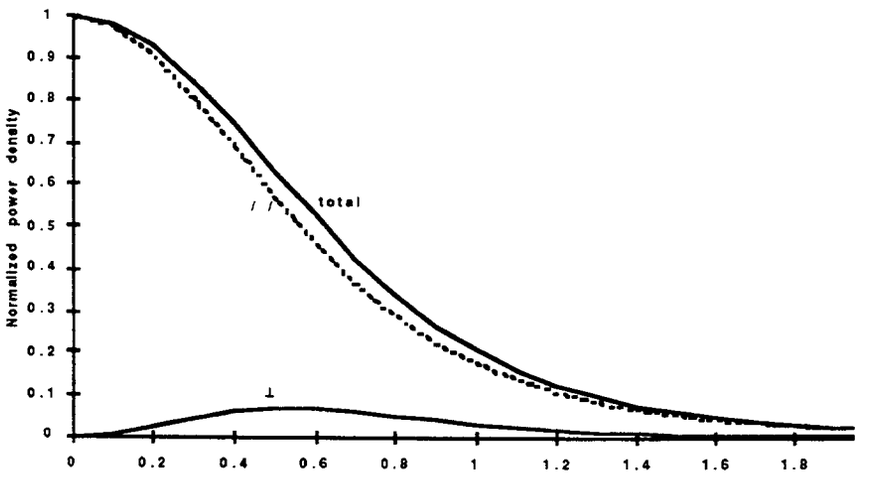


Fig. 8 Relative contributions of the two polarizations

spiral towards the walls of the vacuum chamber. Therefore, RF cavities are needed to compensate for these losses.

The energy loss is obtained by integrating the average power radiated with respect to time over an off-momentum closed orbit (see Fig. 9).

$$U = \int P\, dt = \int P \frac{dt}{ds}\, ds \tag{47}$$

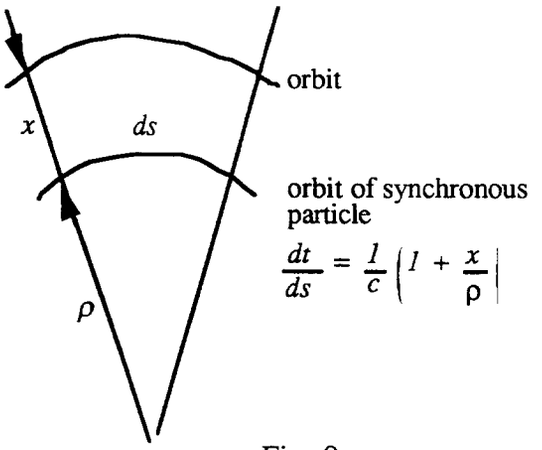


Fig. 9

Taking into account the different path length of an off-momentum particle and the fact that a difference in energy ε (deviation from the design energy E_0) corresponds to a variation in radial position $x = \eta \frac{\varepsilon}{E_0}$, the energy radiated per turn can be rewritten as:

$$U = \frac{1}{c}\int \left(1 + \frac{\eta}{\rho}\frac{\varepsilon}{E_0}\right) P\, ds \tag{48}$$

Since we are interested in small oscillations, we differentiate U and evaluate the derivative at $\varepsilon = 0$:

$$\frac{dB}{dE} = \frac{dx}{dE}\frac{dB}{dx} = \frac{\eta}{E_0}\frac{dB}{dx} \tag{49}$$

When E changes, the orbit moves to a different location so that B changes and we can use the following derivations to get the final expression of the energy loss:

$$\frac{dP}{dE} = 2\frac{P}{E_0} + 2\frac{P}{B_0}\frac{dB}{dE}$$

$$\frac{dB}{dE} = \frac{dx}{dE}\frac{dB}{dx} = \frac{\eta}{E_0}\frac{dB}{dx}$$

$$\frac{dU}{dE} = \frac{U_0}{E_0}\left(2 + \frac{1}{c\,U_0}\int\left\{\eta P\left[\frac{1}{\rho} + \frac{2}{B}\frac{dB}{dx}\right]\right\}ds\right) \tag{50}$$

U_0 being the energy radiated in one revolution by an electron circulating on the design orbit with the nominal energy E_0. We can now define the damping constant as:

$$\alpha_\varepsilon = \frac{1}{2\,T_0}\frac{dU}{dE} = \frac{U_0}{2\,T_0\,E_0}\left(2 + \mathcal{D}\right) \tag{51}$$

with

$$\mathcal{D} = \frac{1}{c\,U_0}\int\left\{\eta P\left[\frac{1}{\rho} + \frac{2}{B}\frac{dB}{dx}\right]\right\}ds \tag{52}$$

Replacing U_0 and P by their expressions yields:

$$\mathcal{D} = \frac{\displaystyle\int \frac{\eta}{\rho^2}\left[\frac{1}{\rho} + \frac{2}{B}\frac{dB}{dx}\right]ds}{\displaystyle\int \frac{\eta}{\rho^2}\,ds} \tag{53}$$

Equation (53) shows that $\mathcal{D}$ is a property of the magnetic guide field configuration which involves only the guide field function and the dispersion. The expression for $\mathcal{D}$ becomes simpler for a separated function machine with an isomagnetic guide field:

$$\mathcal{D} = \frac{\alpha R}{\rho} \tag{54}$$

α being the momentum compaction and R the mean orbit radius. As an example, for the ESRF, the strong focusing required to achieve a low emittance yields a small momentum compaction $\alpha = 3\ 10^{-4}$; the ratio of the mean radius to the magnetic radius is of the order of 5 and the resulting value of $\mathcal{D}$ is 0.0015. In this typical example, $\mathcal{D}$ is sufficiently small so that Eq. (51) can be approximated to:

$$\alpha_\varepsilon = \frac{U_0}{T_0\,E_0} \tag{55}$$

4.2 Damping of betatron oscillations

4.2.1. Vertical oscillations

As shown in Fig. 10, photons are emitted essentially in the direction of the motion since the emission angle of the photon with respect to p is very small ($\theta_{rms} \sim 1/\gamma$), which means that the synchrotron radiation does not change the displacement z nor the slope of the trajectory. The amplitude of oscillations given by $A^2 = z^2 + (\beta z')^2$ (using the smooth approximation for simplicity) is therefore unchanged by the radiation.

When a photon is emitted, the electron loses by radiation a small amount of energy δE and its momentum is changed by $dp = c\ \delta E$. On the other hand, at each passage of the electron in the cavity, the RF will provide some energy to the particle. The accelerating force is parallel to the design orbit. Therefore z is unchanged whilst the slope becomes:

$$z' + \delta z' = \frac{p_\perp}{p + \delta p} = z' \left(1 - \frac{\delta p}{p} \right) \tag{56}$$

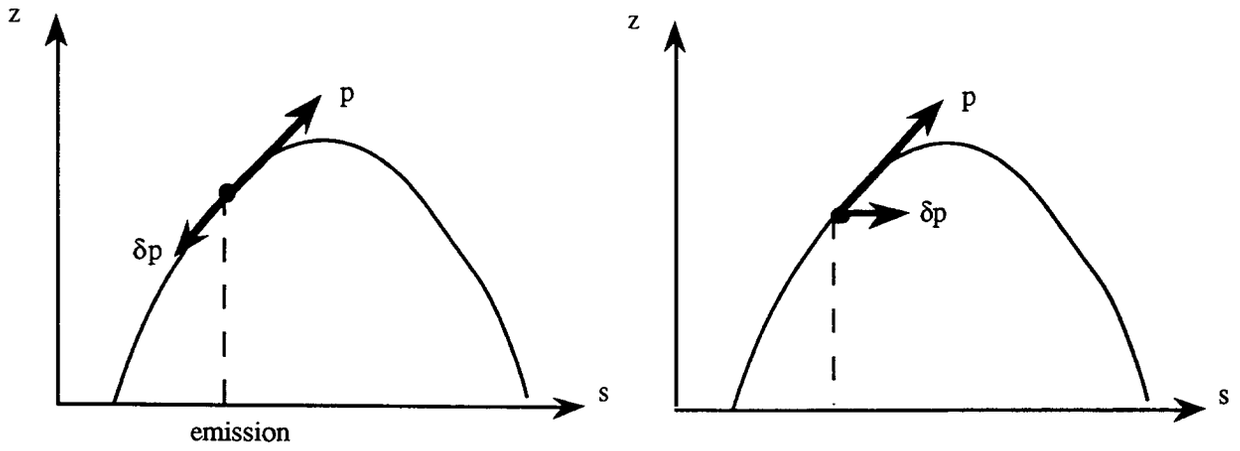


Fig. 10 Effect on an energy change on vertical betatron oscillations

The change in z' is:

$$z' = - z' \frac{\delta p}{p} = -z' \frac{\delta E}{E} \tag{57}$$

There is a corresponding change in the amplitude A:

$$A\ \delta A = (\beta^2 z')\ dz' = - \beta^2\ z'^2 \frac{\delta E}{E} \tag{58}$$

Since the phase of the oscillation when the particle arrives in the RF cavity is arbitrary (any value between 0 and 2 π is equally probable), we can average over all phases:

$$A < \delta A > = < z^2 - A^2 > \frac{\delta E}{E} = - \frac{A^2}{2} \frac{\delta E}{E} \tag{59}$$

Over one revolution, all the elements of energy gain add up to the radiation loss U_0, so that the change in the amplitude A due to the RF acceleration during a revolution is given by:

$$\frac{\Delta A}{A} = - \frac{U_0}{2 E_0} \tag{60}$$

Since ΔA in each revolution time T_0 is proportional to A, the motion is exponentially damped:

$$\frac{1}{A} \frac{dA}{dt} = - \frac{U_0}{2 E_0 T_0} \tag{61}$$

and the damping coefficient is given by:

$$\alpha_z = \frac{U_0}{2 E_0 T_0} \tag{62}$$

It is interesting to mention that α_z is 2 times larger than the damping coefficient of energy oscillations. "Radiation damping" does not come from the radiation process but is linked to the process of energy gain from the RF system and is similar to adiabatic damping during acceleration. It should also be remembered that quantum excitation gives rise to weak excitation of the vertical motion, but that this effect is usually very small and is masked by the much stronger coupling effects between horizontal and vertical motions.

4.2.2 Horizontal oscillations

The total horizontal displacement is the sum of the betatron displacement x_β with respect to the closed orbit and the displacement x_ε of the off-energy closed orbit. In a first approach, there is the same damping mechanism as for vertical oscillations. But the difference comes from the fact that one must take into account the change in the reference orbit due to the energy change. At each emission of a photon, E changes by δE and there is a change in x_ε given by:

$$\delta x_\varepsilon = \eta \frac{\delta E}{E_0} \tag{63}$$

Since the position in space of the particle does not change, there is a change in x_β and x'_β:

$$\begin{aligned} \delta x &= \delta x_\varepsilon + \delta x_\beta = 0 \;\Rightarrow\; \delta x_\beta = -\eta \frac{\delta E}{E_0} \\ \delta x' &= \delta x'_\varepsilon + \delta x'_\beta = 0 \;\Rightarrow\; \delta x'_\beta = -\eta' \frac{\delta E}{E_0} \end{aligned} \tag{64}$$

The change in the amplitude A can then be expressed as:

$$A\, \delta A = x_\beta\, \delta x_\beta + \beta^2\, x'_\beta\, \delta x'_\beta = - \left(\eta\, \delta x_\beta + \eta'\, \beta^2\, x'_\beta \right) \frac{\delta E}{E_0} \tag{65}$$

The energy radiated in a path δl is :

$$\delta E = - \frac{P}{c}\, \delta l \tag{66}$$

Taking into account that $\delta l = c\, \delta t = \delta s \left(1 + \frac{x_\beta}{\rho} \right)$, we can deduce the expectation value of

the amplitude δA after having averaged over all phase angles and used the relationship between a change in energy δE and a change in magnetic field δB. Equation (65) becomes then:

$$\frac{<\delta A>}{A} = \frac{1}{2}\frac{\eta P}{c E_0}\left(\frac{2}{B}\frac{\delta B}{dx} x_\beta + \frac{1}{\rho} \right) \delta s \tag{67}$$

The final step consists in integrating Eq. (67) over one revolution and adding the contribution from the RF. We obtain the damping coefficient of the radial oscillations as:

$$\alpha_x = (1 - \mathcal{D})\frac{U_0}{2 E_0 T_0} \tag{68}$$

Equation (68) shows that there is an antidamping effect of the radiation. However, provided the term $\mathcal{D}$ which counteracts the positive damping of the RF is small (which is the case in separated function machines), $\alpha_x > 0$ and the radial oscillations are damped.

4.2.3 Damping partition numbers

Radiation damping effects have been considered for the three degrees of freedom in a bunch, the two transverse betatron motions and the energy oscillations. Each of these oscillation modes has an exponential decay with damping coefficients:

$$\alpha_i = J_i \frac{U_0}{2 E_0 T} \quad with\ i = \varepsilon, x, z \tag{69}$$

The numbers J_i are known as the damping partition numbers. They are given by:

$$\begin{aligned} J_\varepsilon &= 2 + \mathcal{D} \\ J_x &= 1 - \mathcal{D} \\ J_z &= 1 \end{aligned} \tag{70}$$

Their sum is a constant: $\sum J_i = 4$. This yields a particular result that $J_\varepsilon + J_x = 3$ which applies only under the restrictive assumption that the design orbit lies in a plane and that the magnetic field is symmetrical with respect to this plane. The damping rates can be modified if additional forces are introduced. Such forces may come from image currents in the vacuum chamber walls which could excite unstable transverse coherent oscillations or from currents induced by the beam in RF cavities which are the cause of longitudinal oscillations.

The damping times τ_i are the inverse of α_i. For an isomagnetic machine, they can be expressed as:

$$\tau_i = \frac{4\pi}{C_\gamma} \frac{R \rho}{J_i E_0^3} \tag{71}$$

which shows that the damping time varies as the inverse cube of the energy.

5. RADIATION EXCITATION

5.1 Introduction

Up until now, it has been assumed that the energy loss due to synchrotron radiation was a continuous process. From this classical view, the synchrotron radiation is emitted with a continuous spectrum of frequencies. If we consider the radiation emitted by an electron in some

time interval Δt and make a Fourier analysis of the radiation field, the frequency spectrum will be different for each direction in space. We may average this spectrum over all directions to define a radiated power spectrum $\mathcal{P}(\omega)$ such as $\mathcal{P}(\omega)\, d\omega\, \Delta t$ is the energy radiated in Δt with frequencies between ω and $\omega + d\omega$. The instantaneous radiated power is given by the integral of $\mathcal{P}$ over ω:

$$P = \int_0^\infty \mathcal{P}(\omega)\, d\omega \tag{72}$$

The power spectrum is written as:

$$\mathcal{P}(\omega) = \frac{P}{\omega_c} S\left(\frac{\omega}{\omega_c}\right) \tag{73}$$

The spectral function S is a pure algebraic function shown in Fig. 11. It is expressed by:

$$S = \frac{9\sqrt{3}}{8\pi}\left(\frac{\omega}{\omega_c}\right) \int_{w_c}^{\infty} K_{5/3}\left(\frac{\omega}{\omega_c}\right) d\left(\frac{\omega}{\omega_c}\right) \tag{74}$$

where $K_{5/3}$ is a modified Bessel function.

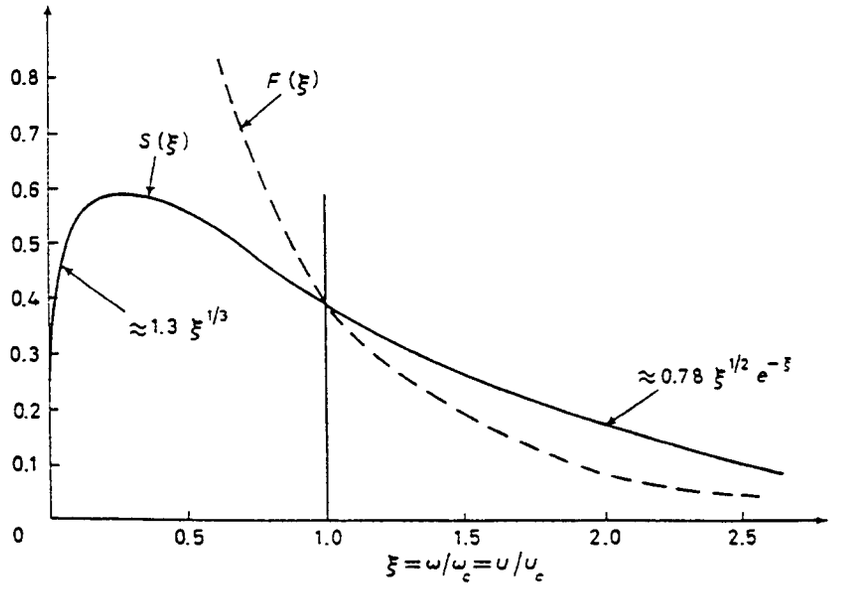


Fig. 11 Normalized power spectrum S

In fact, electromagnetic radiation occurs in quanta of discrete energy; this quantization has significant effects on the behaviour of electrons. Each time a quantum is emitted, the energy of the electron makes a small discontinuous jump. We may consider this emisson as instantaneous since it is much smaller than other times such as betatron or synchrotron periods.

The emission of a given quantum is statistically independent from the other ones and therefore the phenomena can be considered as a purely random process. The cumulative effect of many such disturbances generates a kind of "noise". Electrons undergo a diffusion process which enlarges the oscillation amplitudes until the radiation excitation is balanced by the damping. As already seen, the damping only depends on the mean emitted power, while the diffusion process depends on the statistical properties of the emitted radiation.

5.2 Quantum emission

The electromagnetic radiation at frequency ω is emitted in quanta of energy:

$$u = \frac{h}{2\pi}\omega \tag{75}$$

h being the Planck's constant

Let us define $n\,(u)\,du$ as the number of quanta emitted per unit time with energies between u and $u + du$. Then $n(u)$ can be written as:

$$n\,(u) = \frac{P}{u_c^2} F\left(\frac{u}{u_c}\right) \tag{76}$$

with

$$u_c = \frac{h}{2\pi}\omega_c = \frac{3}{2}\frac{h}{2\pi}\frac{c\,\gamma^3}{\rho} \tag{77}$$

$$F\left(\frac{\omega}{\omega_c}\right) = \frac{\omega}{\omega_c} S\left(\frac{\omega}{\omega_c}\right)$$

Like the frequency spectrum, the quantum spectrum is a universal function of the ratio u/u_c. The total rate of emission of quanta of all energies is given by:

$$\mathcal{N} = \int_0^\infty n\,(u)\,du = \frac{15\sqrt{3}}{8}\frac{P}{u_c} \tag{78}$$

We can also define the mean number of quanta emitted per radian of trajectory which depends only on the electron energy:

$$\mathcal{N}_r = \mathcal{N}\frac{\rho}{\beta c} \approx \frac{5}{2\sqrt{3}}\frac{\gamma}{137} \tag{79}$$

The mean quantum energy is given by:

$$< u > = \frac{1}{\mathcal{N}}\int_0^\infty u\,n\,(u)\,du = \frac{P}{\mathcal{N}} = \frac{8}{15\sqrt{3}}u_c \tag{80}$$

The quantity of interest for the quantum excitation of oscillations is the product of the rate of emission with the mean square quantum energy. As shown from Eq. (81), at a fixed radius the quantum excitation varies as the 7^{th} power of the energy

$$\mathcal{N} < u^2 > = \frac{55}{24\sqrt{3}} P\,u_c = \frac{55}{24\sqrt{3}} r_0 \frac{h}{2\pi} m_0 c^4 \frac{\gamma^7}{\rho^3} \tag{81}$$

5.3 Energy fluctuations and bunch length

The effect of quantum emission on the energy oscillations of an electron is sketched in Fig. 12.

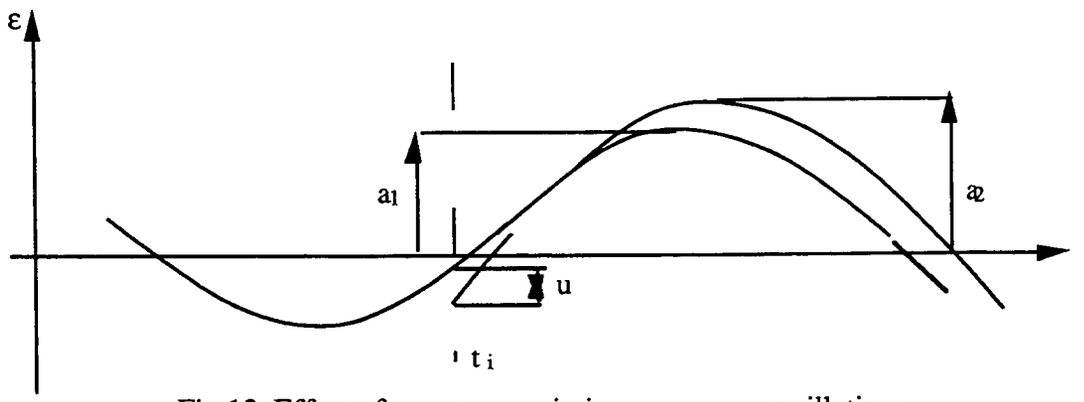


Fig.12 Effect of quantum emission on energy oscillations

The energy deviation of an electron is described by:

$$\varepsilon = A_0 \, e^{\,j\Omega(t - t_0)} \tag{82}$$

When a quantum of energy u is emitted, the energy of the electron is suddenly decreased by $-u$. After t_i, the energy goes to:

$$\varepsilon = A_0 \, e^{\,j\Omega(t - t_0)} - u \, e^{\,j\Omega(t - t_i)} \tag{83}$$

The new oscillation is written as:

$$\varepsilon = A_1 \, e^{\,j\Omega(t - t_i)} \tag{84}$$

The relation between the initial amplitude of the motion A_0 and the new amplitude A_1 is given by:

$$A_1^{\,2} = A_0^{\,2} + u^2 - 2 A_0 \, u \cos \Omega (t_i - t_0) \tag{85}$$

In general the phase $t_i - t_0$ is random and the expectation value of the cosine is zero. Then the probable amplitude change due to the quantum event is:

$$< dA^2 > = < A_1^{\,2} - A_0^{\,2} > = u^2 \tag{86}$$

The amplitude variation per unit time, integrated over the photon spectrum is:

$$\frac{d < A^2 >}{dt} = \int_0^{\infty} u^2 \, n(u) \, du = \mathcal{N} < u^2 > \tag{87}$$

$\mathcal{N} < u^2 >$ may vary around the ring. However, since both the period of oscillation ($\sim 1/\Omega$) and the damping time τ_e are much longer than the revolution time T_0, we may replace the rapidly varying quantity $\mathcal{N} < u^2 >$ by its average over one turn:

$$\frac{d < A^2 >}{dt} = Q_\varepsilon = \frac{1}{2 \pi R} \int \mathcal{N} < u^2 > ds \tag{88}$$

The quantum excitation produces an amplitude growth which is balanced by the radiation damping which contributes to the rate of change of $< A^2 >$ by:

$$\frac{d < A^2 >}{dt} = -2 \frac{< A^2 >}{\tau_\varepsilon} \tag{89}$$

When both quantum excitation and damping are present, the two rates of change of $<A^2>$ must sum to zero. The probable value of A^2 is given by:

$$< A^2 > = \frac{\tau_\varepsilon Q_\varepsilon}{2} \tag{90}$$

which corresponds to the mean value of the energy spread:

$$\sigma^2{}_\varepsilon = \frac{\tau_\varepsilon Q_\varepsilon}{4} \tag{91}$$

The relative energy spread is $\frac{\sigma_\varepsilon}{E_0}$. For an isomagnetic guide field, it can be expressed as:

$$\left(\frac{\sigma_\varepsilon}{E_0}\right)^2 = \frac{C_q \gamma^2}{J_\varepsilon \rho} \quad \text{with} \quad C_q = \frac{55}{32\sqrt{3}} \frac{h}{2\pi m c} = 3.84\ 10^{-13}\ \text{m} \tag{92}$$

The fluctuating energy oscillations are accompanied by associated fluctuations in the time displacement. The standard deviation of these fluctuations is given by:

$$\sigma_\tau = \frac{\alpha}{\Omega E_0} \sigma_e \tag{93}$$

Table 2 gives typical values of the energy spread for a few machines.

Table 2
Typical values of the energy spread for different machines

	ACO	SPEAR	ESRF	PEP	LEP
E_0 (GeV)	0.54	4.5	6.0	18	100
σ_E / E_0	$4.4\ 10^{-4}$	$1.1\ 10^{-3}$	$1.1\ 10^{-3}$	$1.2\ 10^{-3}$	$1.1\ 10^{-3}$
σ_E (MeV)	0.24	4.9	6.6	21.6	110

Bunch lengths typically stand in the cm or few cm range (5 mm for the ESRF). These values which are usually computed for zero beam current can be affected by collective effects generated by the interaction of the beam with the environment. This is the so-called "bunch lengthening" mechanism with increasing beam current.

5.4 Beam width and beam height

The emission of discrete quanta in the synchrotron radiation will also excite random betatron oscillations. These quantum-induced oscillations are responsible for the radial extent of the beam. The calculations are based on the same derivations as for the energy oscillations. We simply give here the results for an isomagnetic guide field:

$$\sigma_{x_\beta}^2 = \frac{C_q \gamma^2}{\rho} \left(\frac{< \mathcal{H} >_{mag}}{J_x} + \frac{\eta^2}{J_\varepsilon} \right) \tag{94}$$

where $< \mathcal{H} >_{mag}$ is the average of the quantity $\mathcal{H}$ taken over the magnets. That is:

$$< \mathcal{H} >_{mag} = \frac{1}{2\pi\rho} \int_{mag} \frac{1}{\beta} \left(\eta^2 + (\beta \eta' - \frac{1}{2} \beta' \eta)^2 \right) ds \tag{95}$$

For the vertical betatron motion, there is no first order effect from quantum emission on the motion if the orbit lies in the plane $z = 0$. However, the unavoidable imperfections of the guide field will yield a coupling of the horizontal and vertical betatron oscillations. This coupling is generally of a few %. In addition, there is a small contribution coming from the fact that the emisssion of a quantum slightly changes the direction of motion of the electron (this had been neglected in the previous derivations). The change in angle of the electron when a photon of energy u is emitted at the angle θ_z is given by:

$$\delta z' = \frac{u}{E_0} \theta_z \tag{96}$$

Since the radiation is emitted in a cone of half-angle $1/\gamma$, we can assume that θ_z is of the order of $1/\gamma$, thus making the change in the vertical oscillation amplitude very small and justifying the fact that this contribution is usually neglected.

Small equilibrium beam sizes are a specific requirement of low emittance lattices for synchrotron radiation sources. As shown in Eq. (95), this can be achieved in the horizontal plane by a proper minimization of the dispersion invariant function and special lattice designs. The reduction of vertical beam sizes is obtained by correction of the coupling by means of skew quadrupoles. Typical beam sizes in the new generation of synchrotron light sources are a few hundred of microns horizontally and a few tens of microns vertically.

* * *

LIST OF REFERENCES

[1] J. D. Jackson, Classical Electrodynamics, Wiley, New York (1962)

[2] A. A. Sokolov and I. M. Ternov, Synchrotron Radiation, Pergamon Press, Berlin (1968)

[3] J. Schwinger, Phys. Rev., 5, 1912 (1948)

[4] H. Bruck, Accélérateurs Circulaires de Particules, PUF, Paris (1966)

[5] M. Sands, The Physics of Electron Storage Rings, SLAC-121 (1970)

[6] K. W. Robinson, Phys. Rev. 111, 373 (1958)

[7] A. Hofmann, Proceedings CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, CERN 90-03, April 1990

[8] K. Hübner, Proceedings CERN Accelerator School, General Accelerator Physics, CERN 85-19, Nov 1985

WIGGLERS

R.P. Walker
Sincrotrone Trieste, Italy

Abstract
The main effects of wigglers on the operation of electron storage rings are discussed, including the closed orbit, linear focusing effects and matching conditions, non-linear beam dynamics, and effects on beam properties arising from the emission of synchrotron radiation.

1 INTRODUCTION

A wiggler is a special magnetic device consisting of a sequence of dipoles with alternating polarity, arranged so that there is no net deflection or displacement of the beam. Such a device in general does not have a major influence on the ring optics and so can be added later and operated almost independently of the rest of the ring in order to enhance particular features of the operation of the ring.

There are several reasons for installing wigglers in electron rings. Firstly, wigglers can be useful as a means of changing various electron beam properties. They contribute to both radiation damping and quantum excitation processes, leading to a new equilibrium with modified bunch length, energy spread, and emittance (larger or smaller) [1]. Several different types of wiggler have been proposed with different field distributions. The basic dipole wiggler, also called a "damping wiggler", gives rise principally to changes in equilibrium emittance, energy spread and damping times [2]. Use of a gradient field leads mainly to a change of damping partition, and was first proposed by Robinson as a means of counteracting radial antidamping in combined function machines [3]. A device of this kind was first used in the CEA in order to allow operation as a storage ring [4]. Finally, non-linear wigglers have been proposed with higher order fields such as sextupole and octupole in order to modify the bunch shape and energy distribution [5-7].

A second use of wigglers is to control electron spin polarization [2]. Due to the emission of Synchrotron Radiation electrons tend to become polarized with their spins aligned in the vertical direction (Sokolov-Ternov effect) [8]. Wigglers can be used to increase the level of polarization, which is of interest for particular high energy physics experiments.

The major use of wigglers however is as specialized sources of synchrotron radiation, in which context they are generally referred to as "insertion devices" (IDs). At present more than fifty IDs are in operation in synchrotron radiation sources world-wide and many more are being built for the most recent "third generation" rings, the largest of which will eventually contain between 30 and 40 IDs. Compared to the radiation emitted by the more conventional bending magnet sources, insertion devices have the following advantages :

i) Higher photon energies. The critical energy of the radiation emitted by the wiggler depends on the magnetic field strength, which can be greater than that of the ring bending magnets. The first device to be proposed for this application consisted of a single high field strength pole with half-strength poles of opposite polarity on either side in order to produce a single bump of the electron trajectory. It was termed a "wavelength shifter" because of its harder spectrum compared to the bending magnet radiation [9]. Since then in order to reach the highest possible magnetic field superconducting devices have been developed. For example a 3.5 T device was installed on VEPP3 in 1979 [10], and a 5 T wiggler on the SRS in 1982 [11]. Superconducting wigglers are currently operating in several synchrotron radiation facilities :

SRS (England), UVSOR and Photon Factory (Japan), NSLS X-ray Ring (U.S.A.) and DCI (France).

ii) Increased photon flux. Using many emitting poles the total photon flux increases proportionally. The first "multipole wiggler" devices were electromagnetic. For example, a 1.8 T, 5-pole device was installed in SPEAR (U.S.A.) in 1979 [12]. Since then permanent magnet devices have been developed with many more poles, for example the 55-pole device installed in SPEAR in 1983 [13].

iii) Increased brightness (photon flux per unit area and solid angle) and quasi-monochromatic spectrum. Under certain circumstances, particularly when the wiggler has a small field strength and period length, interference effects can become important which produces a radiation spectrum consisting of a series of lines, rather than the smooth spectrum characteristic of bending magnet sources. The radiation also has a higher degree of collimation and hence "brightness". The name "undulator" is often used for devices which make use of this interference effect, although there is no essential difference to a multipole wiggler. An undulator was first proposed [14], and later used, by Motz to generate millimetre and visible radiation with electron energies of 5 MeV and 100 MeV respectively [15]. It also formed the basis of the subsequent development of coherent radiation sources, such as the ubitron [16] and free-electron laser [17]. The first use of an undulator in a storage ring for generation of synchrotron radiation was at SSRL in 1980 [18]. A parallel development took place using the VEPP-3 storage ring at Novosibirsk [19]. A device of this kind has also been used in the SPS to generate visible light as a diagnostic tool for the proton and anti-proton beams [20].

iv) Different polarization characteristics. There is increasing interest in using synchrotron radiation with circular or elliptical polarization, which can be generated by wigglers of the "asymmetric" type, or devices with helical or elliptical electron trajectories [21].

In this lecture we consider the main effects of standard wigglers, i.e. "plane" wigglers with midplane symmetry, on the operation of electron storage rings. We examine firstly the electron trajectory in the device and the effects on the closed orbit, then the effects on linear and non-linear beam dynamics, and finally the effects on beam properties arising from the emission of synchrotron radiation.

2 TRAJECTORY AND CLOSED ORBIT

The equation of motion in the standard fixed co-ordinate system {z, x, s} [22] in the case of small angular deflection ($x' << 1$, $z' << 1$) is given by :

$$\begin{aligned} x'' &= \frac{e}{\gamma mc}(B_z - z' B_s) \\ z'' &= \frac{e}{\gamma mc}(x' B_s - B_x) \end{aligned} \tag{1}$$

In the median plane only the vertical field component, B_z, exists and so the horizontal motion can therefore be obtained by simple integration :

$$\begin{aligned} x'(s) &= \frac{e}{\gamma mc}\int_{-\infty}^{s} ds'\ B_z(s') \\ x(s) &= \frac{e}{\gamma mc}\int_{-\infty}^{s} ds' \int_{-\infty}^{s'} ds''\ B_z(s'') \end{aligned} \tag{2}$$

In the periodic part of a wiggler the field is in many cases well described by a sinusoidal distribution: $B_z = B_o \cos(k_o s)$, where $k_o = 2\pi/\lambda_o$ and λ_O is the wiggler period length. The trajectory in the periodic part then becomes:

$$x'(s) = \frac{\sin(k_o s)}{\rho_o k_o} \qquad x(s) = -\frac{\cos(k_o s)}{\rho_o k_o^2} \tag{3}$$

where $1/\rho_o = eB_o/\gamma mc$.

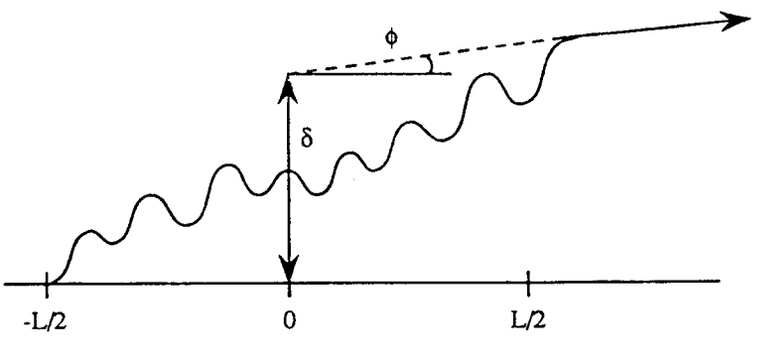


Fig. 1 Projection of the change in beam position and angle introduced by a wiggler to the centre of the device.

In considering the net change in angle (ϕ) and position (δ) introduced by the device it is convenient to calculate these quantities at the centre of the device ($s = 0$) rather than at the end (see Fig. 1) i.e. $\phi = x'(L/2)$ and $\delta = x(L/2) - x'(L/2)L/2$. We have therefore, after an integration by parts for $x(s)$:

$$\phi = \frac{e}{\gamma mc}\int_{-\infty}^{+\infty} B_z\, ds \qquad \delta = -\frac{e}{\gamma mc}\int_{-\infty}^{+\infty} sB_z\, ds \tag{4}$$

The usual requirement is for zero net change in angle or position of the beam and therefore $\phi = \delta = 0$. The most common arrangement employs a field distribution that is symmetric about the centre of the device i.e. an odd number of poles. In the ideal case, neglecting field errors, it follows directly from Eq. (4) that $\delta = 0$. Zero angular deflection is obtained by a suitable arrangement of end-poles, for example using half-strength poles at the ends. Figure 2(a) illustrates the sequence of pole strengths (1, -2, 2, .. 2, -2, 1) and the resulting trajectory for poles with a rectangular shaped field distribution; this and the following results are however valid also for the more common case of a half-sinusoid field shape. A more complicated sequence (1, -3, 4, .. 4, -3, 1) that eliminates the offset of the oscillation axis is illustrated in Fig. 2(b). Anti-symmetric arrangements with an even number of poles are also possible. In this case by symmetry $\phi = 0$ and zero change in position is obtained by a suitable end pole arrangement. One general sequence is that of Fig. 2(c), (1, -3, 4, .. -4, 3, -1) which is similar to that of Fig. 2(b).

So far we have referred only to the changes in angle and position of an electron trajectory, which enters with zero position and angle. When placed in a ring however residual errors in either plane will give rise to a distortion of the closed orbit. In the case of a localized angular deflection the change to the closed orbit is well known :

$$u(s) = \frac{\phi}{2}\sqrt{\beta(s)\beta_w}\,\frac{\cos(\pi Q - \mu(s))}{\sin \pi Q}$$
$$u'(s) = \frac{\phi}{2}\sqrt{\frac{\beta_w}{\beta(s)}}\,\frac{\sin(\pi Q - \mu(s)) - \alpha\cos(\pi Q - \mu(s))}{\sin \pi Q} \qquad (5)$$

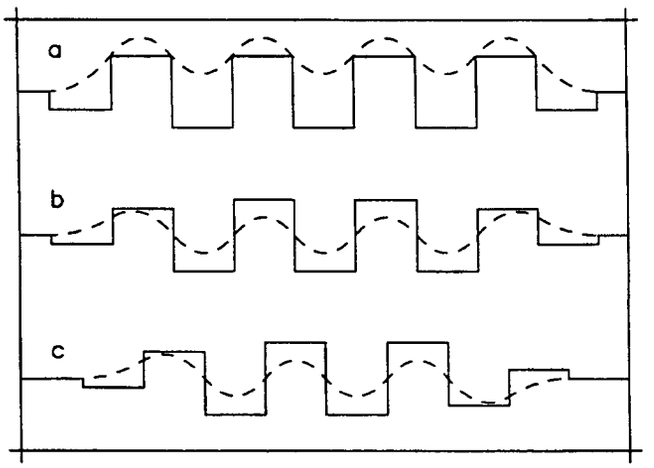


Fig. 2 Various magnetic field distributions (solid lines) resulting in no net change in position or angle of the beam trajectory (dashed lines).

In the above W refers to the wiggler location, and u refers to either x or z. The invariant for such a closed orbit, $\varepsilon_{c.o.} = \gamma u^2 + 2\alpha u u' + \beta u'^2$, is as follows:

$$\varepsilon_{c.o.} = \beta_w\left(\frac{\phi}{2\sin \pi Q}\right)^2 \qquad (6)$$

Similarly for a localized displacement, δ, the closed orbit is given by:

$$u(s) = \frac{\delta}{2}\sqrt{\frac{\beta(s)}{\beta_w}}\,\frac{\sin(\pi Q - \mu(s)) + \alpha\cos(\pi Q - \mu(s))}{\sin \pi Q}$$
$$u'(s) = -\frac{\delta}{2\sqrt{\beta_w\beta(s)}}\,\frac{[1 + \alpha_w\alpha(s)]\cos(\pi Q - \mu(s)) - [\alpha_w - \alpha(s)]\sin(\pi Q - \mu(s))}{\sin \pi Q} \qquad (7)$$

and hence the invariant is:

$$\varepsilon_{c.o.} = \gamma_w\left(\frac{\delta}{2\sin \pi Q}\right)^2 \qquad (8)$$

Thus, while the sensitivity to angle errors depends on the β function at the wiggler, the effect of displacement errors depends on the γ function and may therefore be an important effect for wigglers located at positions with a small beta function. In most cases however $\beta > 1$ and so an anti-symmetric structure with an even number of poles is sometimes used to reduce the effect on the closed orbit arising from non-perfect field integral compensation.

In synchrotron radiation sources it is usually required that the beam position and angle is stable to some fraction of its natural rms size and divergence in order to maintain the high

brightness of the synchrotron radiation beams. In other words it is required that $\varepsilon_{c.o.} < \varepsilon$, where the latter is the natural electron beam emittance in the relevant plane. This condition must be met also during switch-on or tuning of a wiggler field if other users are not to be disturbed. For example, with an emittance of $\varepsilon = 10^{-9}$ m rad, typical of a modern synchrotron radiation source, and with $Q = 0.25$, $\beta_W = 10$ m ($\alpha_W = 0$), this condition requires that the perturbations introduced by a wiggler are limited to the following values: $\phi < 14$ μrad and $\delta < 140$ μm. At 2 GeV for example this corresponds to field integrals, Eq. (4), of less than 10^{-4} Tm and 10^{-3} Tm^2 respectively. Such small field integrals are difficult to achieve in practice without active correction elements either forming part of the wiggler, or external to it. It should be noted that the limits on the field integrals decreases with energy, making wiggler operation in low energy rings even more critical.

3 LINEAR BEAM DYNAMICS

We consider firstly the strongest effect that occurs in a standard wiggler, namely the second-order focusing in the vertical plane. The analysis is extended in section 3.2 to include focusing in the horizontal plane arising from transverse field variations and in section 3.3 a more general approach is presented that is valid also in the case of more complex wigglers which lack midplane symmetry. The effects of the focusing are examined in section 3.4 and ways of dealing with them in section 3.5.

3.1 Vertical plane focusing

A simple approach to the linear dynamics of particles in wigglers is a "hard-edge model" in which each pole is treated as a region of constant magnetic field, and hence bending radius (ρ), as shown in Fig. 3 [9,23,24]. Each pole can then be modelled using the standard transport matrix for a dipole magnet, which can be multiplied together to form the matrix for a complete wiggler. In the usual case of magnet poles arranged with parallel entrance and exit faces, i.e. rectangular magnets, the beam enters and leaves each pole with a non-zero angle (ϕ) that gives rise to edge focusing. The edge effect cancels the weak focusing term in the horizontal plane leaving no net effect, but in the vertical plane there remains a net focusing which adds up from each pole edge and which is proportional to ϕ/ρ.

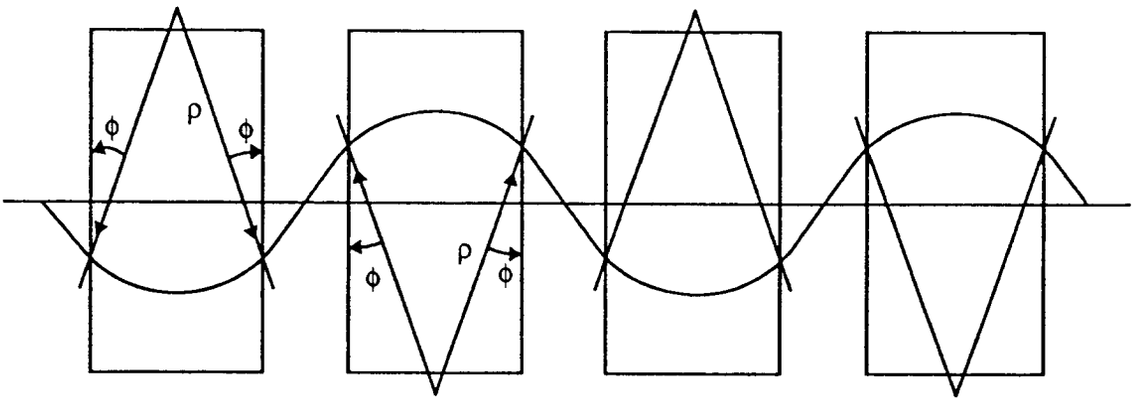


Fig. 3 Hard-edge model for a wiggler with regions of constant magnetic field

The physical origin of the vertical focusing can be understood more clearly from Fig. 4. It can be seen that off-axis vertically there is a longitudinal B_s field component that couples to the angular deflection in the horizontal plane (x') to give a force which acts always towards the beam axis. Near the median plane $B_s = (dB_s / dz)\, z = (dB_z / ds)\, z$, which inserted in Eq. (1) together with the expression for $x'(s)$, Eq. (2), gives:

$$z'' = \left(\frac{e}{\gamma mc}\right)^2 \left(\int B_z ds \, \frac{dB_z}{ds}\right) z \tag{9}$$

Averaging over the wiggler length using integration by parts, the result becomes:

$$\int z'' ds = -z \left(\frac{e}{\gamma mc}\right)^2 \int B_z^2 ds = -z \int \frac{1}{\rho^2} ds \tag{10}$$

i.e. in the vertical plane the wiggler behaves as a quadrupole whose average focusing parameter is given by $K_z = \langle 1/\rho^2 \rangle$.

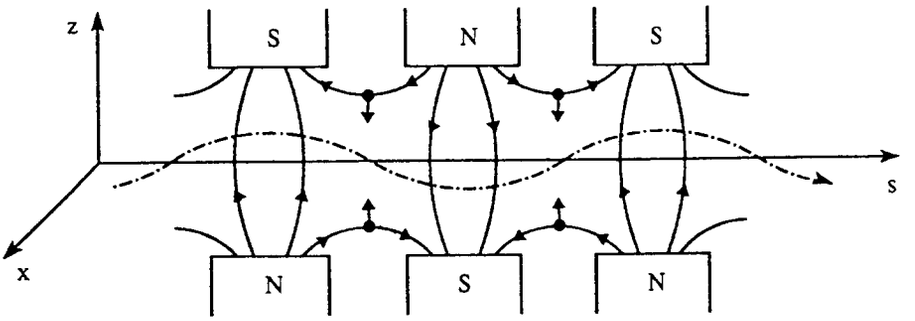


Fig. 4 Field lines and beam trajectory in a wiggler

A third approach is to consider the motion with respect to the oscillating trajectory in the device. We define the local field gradient (see Fig. 5):

$$\frac{dB_z}{d\xi} = -\frac{dB_z}{ds} x' \tag{11}$$

where x' is the angle of the reference trajectory and where we have assumed $dB_z/dx = 0$. The usual focusing parameter is then:

$$k = \frac{e}{\gamma mc} \frac{dB_z}{d\xi} = -\frac{e}{\gamma mc} \frac{dB_z}{ds} x' \tag{12}$$

Averaging over the wiggler length as above we obtain effectively the same result as before:

$$\int k \, ds = \left(\frac{e}{\gamma mc}\right)^2 \int B_z^2 \, ds = \int \frac{1}{\rho^2} \, ds \tag{13}$$

The standard linear equations of motion are therefore:

$$\begin{aligned} x'' + K_x x &= 0 \\ z'' + K_z z &= 0 \end{aligned} \tag{14}$$

with

$$\begin{aligned} K_x &= \langle (1/\rho^2 - k) \rangle = 0 \\ K_z &= \langle k \rangle = \langle 1/\rho^2 \rangle \end{aligned} \tag{15}$$

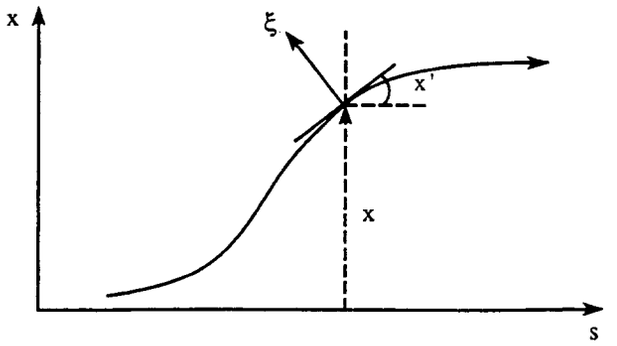


Fig. 5 Definition of the local co-ordinate perpendicular to the beam trajecory

It can be seen from the results above that the standard wiggler behaves like a drift space in the horizontal plane and as a focusing quadrupole magnet in the vertical plane. It should be noted that the focusing is a second-order effect being proportional to the *square* of the field and inversely proportional to the *square* of the energy. In the case of a sinusoidal wiggler with amplitude B_o the focusing parameter, averaged over each magnet period, becomes:

$$K_z = \left(\frac{e}{\gamma mc}\right)^2 \frac{B_o^2}{2} = \frac{1}{2\rho_o^2} \tag{16}$$

The averaging can however be carried out over the whole magnet, if the integration is carried out between points with zero field. The integrated focusing parameter is therefore $K_z L = \int 1/\rho^2 ds$ for any arbitrary field distribution $B_z(s)$. The only restrictions on the validity are that the magnet has midplane symmetry and no transverse field variation.

3.2 Horizontal plane focusing

In the previous section we have assumed that there is no transverse variation (i.e. in the x direction) of the B_z field component. However such a variation can arise due to the finite dimensions of the magnet poles, or can even be introduced deliberately by curving the pole faces. With such an alternating quadratic field variation, the beam oscillating from side to side sees an effective quadrupole field (see Fig. 6). In the case where the field amplitude increases away from the axis this gives rise to focusing in the horizontal plane and defocusing in the vertical plane.

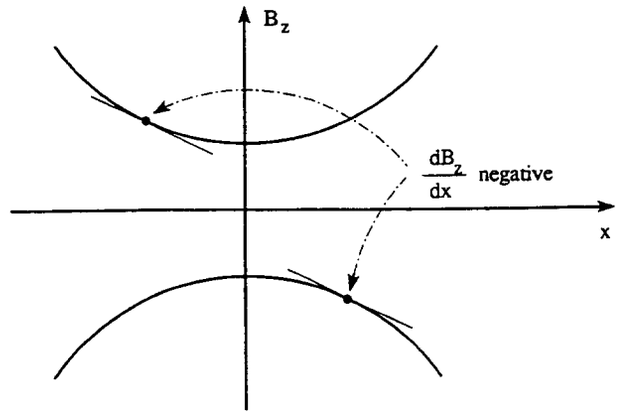


Fig. 6 Field gradient arising from a quadratic transverse field variation and the oscillating beam trajectory

Extending the analysis of section 3.1, the local gradient becomes:

$$\left.\frac{dB_z}{d\xi}\right|_x = \left.\frac{dB_z}{dx}\right|_x - \left.\frac{dB_z}{ds}\right|_x x' \tag{17}$$

In the above the subscript indicates that the derivatives are to be evaluated on the electron trajectory $x(s)$. Assuming no linear field gradient at the origin we can approximate the above as follows:

$$\left.\frac{dB_z}{d\xi}\right|_x = \left.\frac{d^2B_z}{dx^2}\right|_o x - \left.\frac{dB_z}{ds}\right|_o x' \tag{18}$$

Using Eq. (2) for the trajectory, and integrating by parts, assuming $x'(\infty) = x(\infty) = 0$, we obtain :

$$\int k\,ds = \frac{e}{\gamma mc}\int \frac{dB_z}{d\xi} = \left(\frac{e}{\gamma mc}\right)^2\left[-\frac{1}{2}\frac{d^2}{dx^2}\int(\int B_z ds)^2 + \int B_z^2 ds\right] \tag{19}$$

The integrated focusing parameters for the horizontal and vertical plane therefore become:

$$\begin{aligned}\int K_x\,ds &= \int(1/\rho^2 - k)ds = \left(\frac{e}{\gamma mc}\right)^2 \frac{1}{2}\frac{d^2}{dx^2}\int(\int B_z ds)^2 ds \\ \int K_z\,ds &= \int k\,ds = \left(\frac{e}{\gamma mc}\right)^2 \frac{1}{2}\frac{d^2}{dz^2}\int(\int B_z ds)^2 ds\end{aligned} \tag{20}$$

where use has been made of the relationship:

$$\frac{1}{2}\frac{d^2}{dx^2}\int(\int B_z ds)^2 ds + \frac{1}{2}\frac{d^2}{dz^2}\int(\int B_z ds)^2 ds = \int B_z^2 ds \tag{21}$$

In the case of a sinusoidal field distribution, the complete expression for the field components near the axis including the transverse variation is as follows :

$$\begin{aligned} B_z &= \quad\;\; B_o \cosh(k_x x)\;\cosh(k_z z)\;\cos(k_o s) \\ B_x &= -k_x/k_z\; B_o \sinh(k_x x)\;\sinh(k_z z)\;\cos(k_o s) \\ B_s &= -k_o/k_z\; B_o \cosh(k_x x)\;\sinh(k_z z)\;\sin(k_o s) \end{aligned} \tag{22}$$

where, $k_x^2 + k_z^2 = k_o^2 = (2\pi/\lambda_o)^2$. Evaluating Eq. 20 in this case, the average focusing parameters are as follows:

$$\begin{aligned}\langle K_x\rangle &= \langle 1/\rho^2 - k\rangle = \frac{k_x^2}{2\rho_o^2 k_o^2} \\ \langle K_z\rangle &= \langle k\rangle = \frac{k_z^2}{2\rho_o^2 k_o^2}\end{aligned} \tag{23}$$

With the given constraint of midplane symmetry it follows directly from Eq. (14) that the sum of the two focusing parameters depends only on the square of the vertical field component:

$$\langle K_x \rangle + \langle K_z \rangle = \langle 1/\rho^2 \rangle \tag{24}$$

In the usual case with finite pole width k_x^2 is negative; the $\cosh(k_x x)$ terms in Eq. 22 are then replaced by $\cos(k_x x)$. This gives rise to a defocusing in the horizontal plane and an increase in the vertical focusing. With pole faces curved so as to increase the field away from the axis, there is focusing in both planes. Such a situation is of particular interest in free-electron laser applications [25].

3.3 General approach to second-order effects

In the case of wigglers with an arbitrary field distribution the analysis of section 3.2 is no longer valid and a more general approach is needed [26]. Starting from Eq. (1) we can write the first-order solution using the on-axis field components directly as follows:

$$\begin{aligned} x_o'(s) &= \frac{e}{\gamma mc}\int ds\, B_z \qquad x_o(s) = \frac{e}{\gamma mc}\int ds \int B_z ds \\ z_o'(s) &= -\frac{e}{\gamma mc}\int ds\, B_x \qquad z_o(s) = -\frac{e}{\gamma mc}\int ds \int B_x ds \end{aligned} \tag{25}$$

The second-order solution can be obtained by substituting in Eq. (1) and expanding the field components about the trajectory (x_O, z_O) :

$$\begin{aligned} x'' &= \frac{e}{\gamma mc}\left(x_o \frac{dB_z}{dx} + z_o \frac{dB_z}{dz} - z_o' B_s\right) \\ z'' &= \frac{e}{\gamma mc}\left(x_o' B_s - x_o \frac{dB_x}{dx} - z_o \frac{dB_x}{dz}\right) \end{aligned} \tag{26}$$

Integration then yields the second-order angular deflections. After integration by parts and some simplification, assuming $x_O(\infty) = z_O(\infty) = 0$, we obtain:

$$\begin{aligned} \Delta x' &= -\frac{1}{2}\left(\frac{e}{\gamma mc}\right)^2 \frac{d}{dx}\int\left\{\left(\int B_x\, ds\right)^2 + \left(\int B_z\, ds\right)^2\right\} ds \\ \Delta z' &= -\frac{1}{2}\left(\frac{e}{\gamma mc}\right)^2 \frac{d}{dz}\int\left\{\left(\int B_x\, ds\right)^2 + \left(\int B_z\, ds\right)^2\right\} ds \end{aligned} \tag{27}$$

In some cases the magnet geometry may lead to second-order deflections, which can be calculated directly using the above formulae. Taking the first derivatives with respect to x and z yields the focusing parameters:

$$\begin{aligned} \int K_x\, ds &= \frac{1}{2}\left(\frac{e}{\gamma mc}\right)^2 \frac{d^2}{dx^2}\int\left\{\left(\int B_x\, ds\right)^2 + \left(\int B_z\, ds\right)^2\right\} ds \\ \int K_z\, ds &= \frac{1}{2}\left(\frac{e}{\gamma mc}\right)^2 \frac{d^2}{dz^2}\int\left\{\left(\int B_x\, ds\right)^2 + \left(\int B_z\, ds\right)^2\right\} ds \end{aligned} \tag{28}$$

Further differentiation yields the integrated non-linear terms.

3.4 Linear effects in a storage ring

A general focusing element can be represented by the following standard matrix:

$$M = \begin{pmatrix} \cos\theta & \beta\sin\theta \\ -\sin\theta/\beta & \cos\theta \end{pmatrix} \tag{29}$$

where $\theta = \sqrt{K}L$, $\beta = L/\theta = 1/\sqrt{K}$ and where K represents either K_z or K_x. Such a matrix can be used directly in any of the standard programs that calculate the linear properties of a magnet lattice. To proceed further analytically it is preferable to use the effective "thin-lens" (i.e. zero length) matrix, M_{eff}, that acts at the centre of the device which can be obtained as follows [24]:

$$M_{\text{eff}} = \begin{pmatrix} 1 & -L/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & \beta\sin\theta \\ -\sin\theta/\beta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & -L/2 \\ 0 & 1 \end{pmatrix} \tag{30}$$

After carrying out the multiplication the result can be expressed in a similar form to that of matrix M, i.e.

$$M_{\text{eff}} = \begin{pmatrix} \cos\theta^* & \beta^*\sin\theta^* \\ -\sin\theta^*/\beta^* & \cos\theta^* \end{pmatrix} \tag{31}$$

where,

$$\cos\theta^* = \cos\theta + \frac{\theta\sin\theta}{2} \tag{32}$$

and

$$\beta^* = L\left(\frac{1}{\theta^2} - \frac{\cot\theta}{\theta} - \frac{1}{4}\right)^{1/2} \tag{33}$$

Figure 7 shows the variation of β^* with bending radius for two different lengths of a standard wiggler. It can be seen that at large ρ it tends to a constant value, given by $L/\sqrt{12}$. As ρ reduces there is only a small increase until the point is reached when θ approaches a value of π. At this limit the wiggler matrix equals the negative of the identity matrix, and β^* goes to infinity.

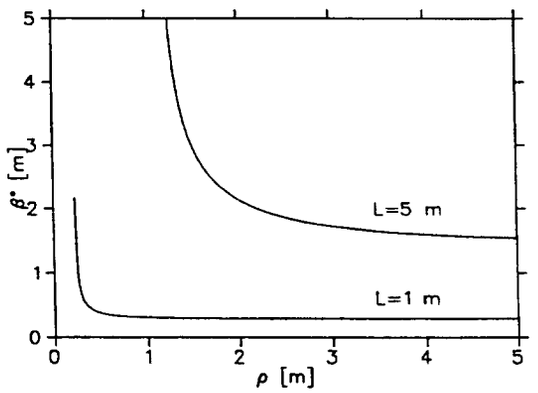


Fig. 7 Variation of β^* function with bending radius, for two wiggler lengths

The effect on the tune can be found in the standard way by calculating the modified one-turn matrix, by multiplying the unperturbed one-turn matrix by M_{eff}. The resulting expression for the modified phase advance with the wiggler activated, μ_W, becomes in the general case:

$$\cos\mu_w = \cos\mu\ \cos\theta^* - \frac{\sin\mu\ \sin\theta^*}{2}\left(\gamma\beta^* + \frac{\beta}{\beta^*}\right) \tag{34}$$

where μ is the one-turn phase advance in the absence of the wiggler. If the wiggler is located at a symmetry point the result can be written as follows:

$$\cos\mu_w = \cos(\mu + \theta^*) - \sin\mu\ \sin\theta^* \frac{(\beta^* - \beta)^2}{2\beta\beta^*} \tag{35}$$

A second effect of the wiggler is to modify the betatron function. For simplicity we consider only the case where the wiggler is located at a symmetry point and that the ring has at least two-fold symmetry. We then evaluate the modified one-turn matrix starting from the symmetry point on the opposite side of the ring:

$$\begin{pmatrix} \cos\mu/2 & \beta\sin\mu/2 \\ -\gamma\sin\mu/2 & \cos\mu/2 \end{pmatrix}\begin{pmatrix} \cos\theta^* & \beta^*\sin\theta^* \\ -\sin\theta^*/\beta^* & \cos\theta^* \end{pmatrix}\begin{pmatrix} \cos\mu/2 & \beta\sin\mu/2 \\ -\gamma\sin\mu/2 & \cos\mu/2 \end{pmatrix} \tag{36}$$

Evaluating the element m_{12} of the above matrix we obtain the result:

$$\beta_w \sin\mu_w = \beta\ \sin\mu\ \cos\theta^* + (\beta^{*2}\cos^2\mu/2 - \beta^2\sin^2\mu/2)\sin\theta^*/\beta^* \tag{37}$$

where β_W is the modified β value at the point opposite the wiggler. The beta values at other points around the ring can be found from:

$$\frac{\Delta\beta(s)}{\beta(s)} = \frac{(\beta_w - \beta)}{\beta}\left[1 - \left(\sin^2\mu(s)\left(1 + \frac{\beta}{\beta_w}\right)\right)\right] \tag{38}$$

from which it follows that the maximum relative change in betatron function at any point is given by:

$$\begin{aligned} \left(\frac{\Delta\beta(s)}{\beta(s)}\right)_{\max} &= \frac{(\beta_w - \beta)}{\beta}, \quad \text{if } \beta_w > \beta \\ &= \frac{(\beta - \beta_w)}{\beta_w}, \quad \text{if } \beta > \beta_w \end{aligned} \tag{39}$$

A third consequence of the wiggler is that a stopband is produced. The stopband is defined as the interval of μ for which $|\cos\mu_W| > 1$. The boundary points of the stopband can be calculated from the following [24]:

$$\tan\left(\frac{\mu}{2}\right) = -\tan\left(\frac{\theta^*}{2}\right)\left[h \pm \sqrt{h^2 - 1}\right] \tag{40}$$

where $h = (\beta^{*2} + \beta^2)/2\beta\beta^*$.

Provided the wiggler strength is not too large ($\theta \le 0.5$) the following approximations can be made: $\cos\theta^* = 1$, $\sin\theta^* = \theta^2/\sqrt{12}$, $\beta^* = L/\sqrt{12}$. The thin-lens wiggler matrix then becomes:

$$M_{\text{eff}} = \begin{pmatrix} 1 & \theta^2 L/12 \\ -\theta^2/L & 1 \end{pmatrix} \tag{41}$$

The tune shift ($\Delta Q = \Delta\mu/2\pi$) is given by:

$$\Delta Q = \frac{K L \beta}{4\pi}\left(1 + \frac{L^2}{12\beta^2}\right) \tag{42}$$

the change in the betatron function by:

$$\frac{\Delta\beta}{\beta} = -\frac{K L \beta}{2\sin\mu}\left(1 - \frac{L^2}{12\beta^2}\right) \tag{43}$$

and the tune stopband width is given by:

$$\Delta Q = \frac{K L \beta}{2\pi}\left(1 - \frac{L^2}{12\beta^2}\right) \tag{44}$$

Note that in the limit $L \rightarrow 0$ the standard expressions for a localized gradient error of magnitude KL are recovered. The effects in this small θ limit are thus clearly proportional to $K_{x,z}$ and hence $(B_0/E)^2$. Figure 8 shows the measured vertical tune shift introduced by the SRS 5-T wiggler, which is in very good agreement with the calculated values, and clearly demonstrates the quadratic dependence on field strength [11].

Table 1

Tune shift, stopband width and maximum change in beta function produced in the vertical plane by various wigglers at different energies, assuming a residual tune value of 0.1, and β_z = 10 m.

W1: B_0 = 0.5 T, L = 2.5 m, W2 : B_0 = 1.5 T, L = 2.5 m, W3 : B_0 = 5.0 T, L = 0.35 m

Device	E [GeV]	θ	β^*	ΔQ	$\Delta Q_{\text{stopband}}$	$(\Delta\beta/\beta)_{\text{max}}$
W1	0.8	0.33	0.73	$3.1\ 10^{-2}$	$6.7\ 10^{-2}$	$3.3\ 10^{-1}$
W1	2.0	0.13	0.72	$5.5\ 10^{-3}$	$1.1\ 10^{-2}$	$5.8\ 10^{-2}$
W1	6.0	0.04	0.72	$6.2\ 10^{-4}$	$1.2\ 10^{-3}$	$6.6\ 10^{-3}$
W1	20.0	0.01	0.72	$6.0\ 10^{-5}$	$1.0\ 10^{-4}$	$5.9\ 10^{-4}$
W2	0.8	0.99	0.82	$1.8\ 10^{-1}$	$3.3\ 10^{-1}$	2.7
W2	2.0	0.40	0.74	$4.2\ 10^{-2}$	$9.5\ 10^{-2}$	$4.7\ 10^{-1}$
W2	6.0	0.13	0.72	$5.5\ 10^{-3}$	$1.1\ 10^{-2}$	$5.8\ 10^{-2}$
W2	20.0	0.04	0.72	$5.0\ 10^{-4}$	$1.0\ 10^{-3}$	$5.3\ 10^{-3}$
W3	0.8	0.46	0.10	$3.4\ 10^{-1}$	$4.0\ 10^{-1}$	
W3	2.0	0.19	0.10	$6.3\ 10^{-2}$	$1.4\ 10^{-1}$	$7.3\ 10^{-1}$
W3	6.0	0.06	0.10	$8.4\ 10^{-3}$	$1.7\ 10^{-2}$	$9.1\ 10^{-2}$
W3	20.0	0.02	0.10	$7.8\ 10^{-4}$	$1.6\ 10^{-3}$	$8.4\ 10^{-3}$

Table 1 shows the effect of three different standard wigglers in rings of different energy, calculated using Eqs. (35, 37, 39, 40) above. The wiggler parameters are typical of those of

an undulator (W1), multipole wiggler (W2) and superconducting wavelength shifter (W3) used in synchrotron radiation sources. It is apparent that large perturbations can be introduced by the operation of high field wigglers in low energy rings. Although the tune shift may be easily overcome using the lattice quadrupoles, the beta function distortion and stopband remain and may in certain cases be large enough to prevent operation of the ring. One possibility then is to reduce the beta function at the wiggler location in order to reduce the perturbation, since all effects are roughly linear with the β value. Alternatively, the wiggler can be matched into the lattice in such a way that the effects are eliminated or at least substantially reduced, as will be discussed in the following section. The effects decrease rapidly with energy, varying as E^{-2}, so that in high energy machines no compensation is usually necessary.

A final consequence of the focusing effect is a possible change in the closed orbit. For example, if the closed orbit is displaced in the vertical plane by an amount z_o at the position of the wiggler when the wiggler is off, then switching on the wiggler will result in an angular deflection given by:

$$\phi = \frac{-K_z L z_o}{1+\beta_z K_z L \cot(\pi Q)/2} \tag{45}$$

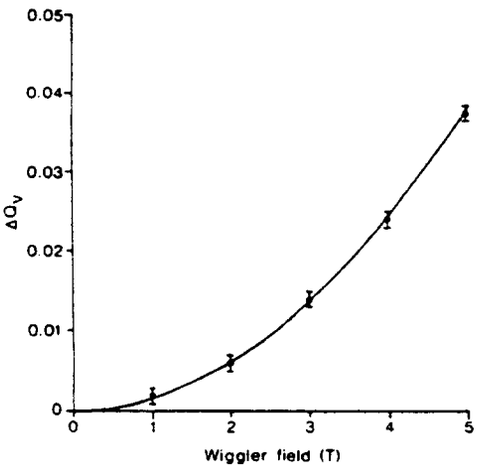


Fig. 8 Measured (points) and calculated (line) vertical tune shift due to the SRS superconducting wiggler [11].

The denominator in the above expression expresses the fact that the angular kick is reduced due to the change in closed orbit at the wiggler location. The deflection can be quite a significant effect, for example about 63 μrad for a 1 mm offset in a wiggler of type W2 (see Table 1) at 2 GeV. The closed orbit must therefore be well corrected at the position of the wiggler, and the wiggler well aligned to the nominal beam axis, to minimize unwanted closed orbit changes when the wiggler is operated.

3.5 Compensation of linear optics effects

There are several reasons why it may be desirable to compensate for the linear optics effects introduced by wigglers:

- the tune shift could bring operation near a resonance
- the vertical beta distortion causes vertical beam size changes, which could affect the beam lifetime, and affect users of the synchrotron radiation
- the change of betatron phase advance between the lattice sextupoles could adversely affect the non-linear beam dynamics (dynamic aperture).

There exits one particular matching condition such that the lattice functions outside the wiggler remain unchanged when the wiggler is energised [24]. The condition can be deduced from Eqs. (35) and (37) and is simply that the betatron function at the wiggler location (with the wiggler off) must equal the wiggler "beta function" i.e. $\beta = \beta^*$. In this case with the wiggler on we have simply: $\mu_W = \mu + \theta^*$ and $\beta_W = \beta$, i.e. there is a tune change but no lattice function change. It can be seen from Eq. 40 that the stopband width is also zero. The matching condition is also evident in Eqs. (43) and (44), since in the limit of small θ, there is zero beta function change and stopband width when $\beta = \beta^* = L/\sqrt{12}$. The matching condition may also be deduced by examining the lattice function changes introduced by the effective lens acting at the wiggler centre, Eq. (31) [27]. Since the matching condition depends on β^*, a perfect match is obtained at only one field value. As an example, consider the wiggler magnet that was installed in ADONE [24]. With a relatively high operating field of 1.8 T and low electron energy (0.54 GeV) the wiggler had a large effect on the linear optics. In order to reduce the effect, a new operating point was found which gave a betatron function closer to the matched value, $\beta^* = 0.96$ m, which substantially reduced the betatron function variation and stopband width, as shown in Table 2.

Table 2
Parameters of the ADONE wiggler at two different working points [24]

Q_z	β	ΔQ_z	$\Delta\beta/\beta$	Stopband (ΔQ)
3.1	3.1	0.20	0.64	0.27
5.2	0.92	0.15	0.05	0.01

Figure 9 illustrates the fact that in the case of a perfect match for the above example, the beta function changes only in the region of the wiggler and not outside it.

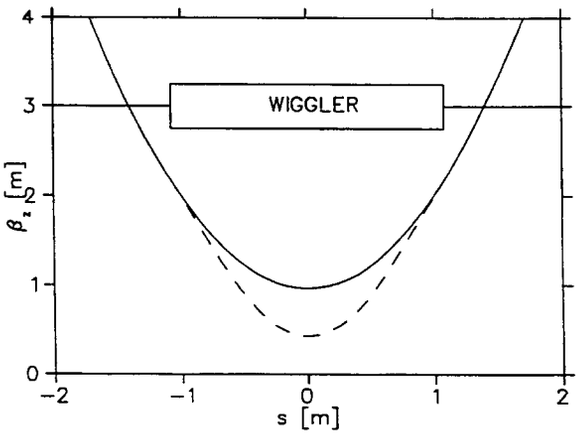


Fig. 9 Vertical beta function variation in the matched case with the wiggler off (solid line) and on (dashed line).

The above matching scheme places a severe constraint on the lattice optics, requiring in general very low vertical beta functions, and is therefore not suitable (or necessary) in the majority of cases. A more flexible approach is usually required, suitable for overcoming the effects of several different devices, using a series of quadrupoles located in the same straight section as the wiggler, or nearby. To compensate fully for the focusing effects in both planes would require in the general case adjustment of six independent parameters – α_x, α_z, β_x, β_z, μ_x, μ_z – requiring therefore six independent quadrupoles. If the wiggler is located at a symmetry point, this could be reduced to four independent quadrupole pairs, however in practise rings have at most two or three independent quadrupole pairs for each insertion device

straight section. Since all parameters cannot then be corrected, priority is usually given to the tune and secondly to the β function variation. In the following we consider the various schemes that have been proposed.

The simplest scheme to overcome the vertical focusing effects of a standard wiggler is to adjust a single defocusing quadrupole located near the wiggler in order to correct the vertical tune and at the same time reduce the vertical beta-function variation [28]. A beta variation is introduced in the horizontal plane, but since at the quadrupole $\beta_z > \beta_x$, this is generally small. A small radial tune change is also introduced that can be compensated globally. Such a scheme is used in the SRS in order to overcome the effects of the superconducting wiggler [29]; Fig. 10 shows that the change in beta function introduced by the wiggler is much reduced.

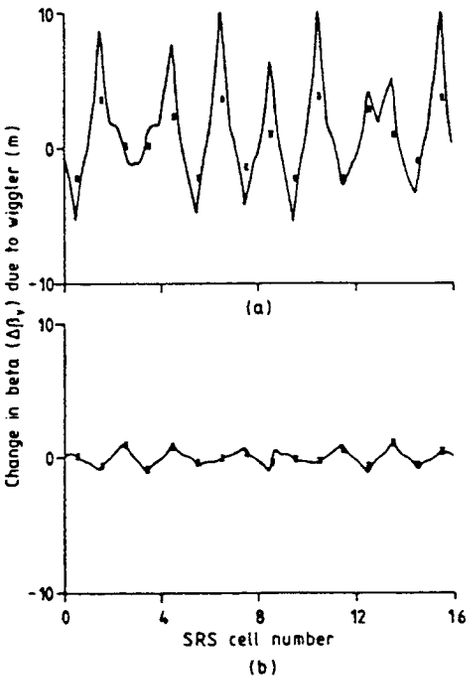


Fig. 10 Vertical beta function modulation due to the SRS wiggler with global (a) and local (b) tune correction. Computed curve and experimental points [29].

More complex lattice designs allow greater scope for overcoming the effects of wigglers. For example, with the wiggler located at a symmetry point two local quadrupole families can be used to correct the tunes in both planes and also reduce the beta-function variation, as used for example at BESSY [30]. Alternatively, the two quadrupole pairs can be used to eliminate the beta modulation in both planes, using a scheme called "alpha matching", see Fig. 11. In this technique the quadrupole strengths are modified to achieve $\alpha = 0$ in both planes at the centre of the wiggler starting from the un-modified lattice functions outside the quadrupoles. In this way the lattice functions remain unaltered outside the region of the quadrupole pairs. The remaining tune shift can then be corrected in the usual way using global quadrupole families. With three local quadrupole families there is further flexibility to either correct both tunes and minimize beta variations, correct the vertical tune and both beta variations, or make a compromise between tune and beta correction.

A different approach was taken in overcoming the significant effects of the vertical superconducting wiggler in the Photon Factory [31]. Before correction the wiggler introduces tune shifts of $\Delta Q_x = 0.081$, $\Delta Q_z = 0.023$ and beta modulations of $\Delta\beta_x/\beta_x = 70\%$, $\Delta\beta_z/\beta_z = 15\%$. A solution was sought using an arbitrary number and distribution of quadrupoles to correct both tunes and minimize the beta function variation. After correction

using three quadrupoles the tune shifts were corrected to zero and the beta modulation to less than 10%, as shown in Fig. 12.

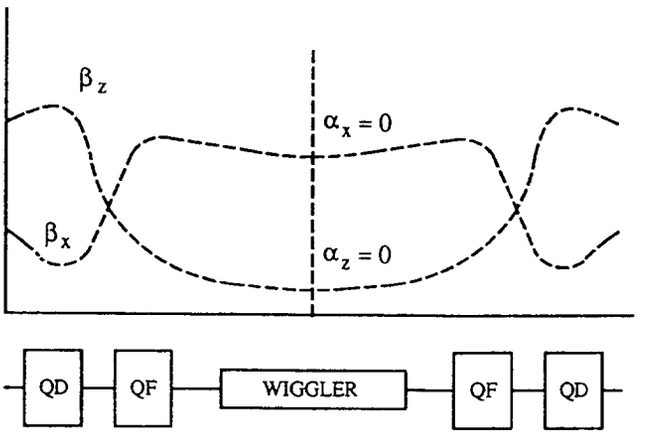


Fig. 11 "Alpha matching" scheme using two pairs of quadrupoles to correct the beta function variation in both planes due to the wiggler.

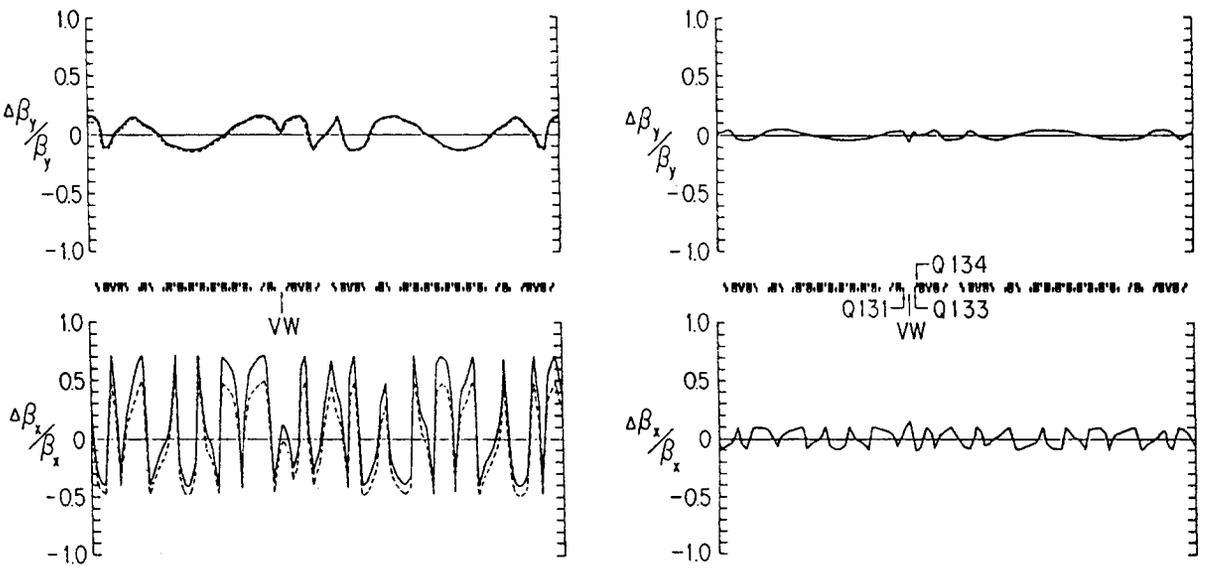


Fig. 12 Beta function modulation due to the PF vertical superconducting wiggler before (left) and after (right) correction [31]

Finally, another special matching condition is worth mentioning that can be of use in particular situations. It follows from the equations that describe the transport of the beam envelope or Twiss parameters through an element that in the case of a focusing element when $\alpha = 0$ and $\beta = 1/\sqrt{K}$ at the entrance, the Twiss parameters remain constant through the device. Such a condition is of special interest for free-electron lasers since it results in an optimum match of the electron beam to the wiggler to minimize the effects of beam emittance on spectral broadening and gain [27]. In the case of a plane magnet the matching condition is therefore $\beta_z = \sqrt{2}\rho_o$. The condition can be applied also in the horizontal plane given suitably shaped magnet pole faces, as mentioned in Section 3.2 above.

4 NON-LINEAR BEAM DYNAMICS

4.1 Non-linear motion in the vertical plane

We consider firstly the equation of motion in the vertical plane, assuming the sinusoidal field model, Eq. (22), with $B_X = 0$. Equation (1) with

$$x'(s) = \frac{e}{\gamma mc}\frac{B_o}{k_o}\cosh(k_o z)\ \sin(k_o s) \tag{46}$$

becomes:

$$z''(s) = -\left(\frac{e}{\gamma mc}\right)^2 \frac{B_o^2}{2}\frac{\sinh(k_o z)\ \cosh(k_o z)}{k_o} \tag{47}$$

Expanding the sinh and cosh terms gives:

$$z''(s) = -\frac{1}{2\rho^2}\left(z + \frac{2}{3}k_o^2 z^3\right) \tag{48}$$

The first term is the linear focusing discussed previously. The second term is a non-linear "octupole-like" term, whose magnitude is proportional to $B_O{}^2/E^2\lambda_O{}^2$.

The equivalent octupole field strength that gives the same cubic term as the wiggler ($B_x = B_3 z^3$) is given by:

$$B_3 = \frac{e}{\gamma mc}\frac{B_o^2}{3}\left(\frac{2\pi}{\lambda_o}\right)^2 \tag{49}$$

For example, the BESSY undulator with λ_O = 70 mm, B_O = 0.5 T, E = 800 MeV has an equivalent octupole field at z = 25 mm, of only 3.9 mT. Nevertheless this is sufficient to lead to observable effects on the beam such as a tune shift with vertical amplitude and excitation of octupole resonances [32].

4.2 Equations of motion

A more accurate analytic treatment of the effects of wigglers on beam dynamics has been made in Ref. [33] using the Hamiltonian formalism. Here we summarize the method and its results. A Hamiltonian for the motion can be written as follows:

$$H = \frac{1}{2}\left\{p_s^2 + \left[p_x - A_x\right]^2 + \left[p_z - A_z\right]^2\right\} \tag{50}$$

where the vector potential for the field distribution, Eq. (22), is given by:

$$\begin{aligned} A_x &= \frac{1}{k_o\rho_o}\cosh(k_x x)\cosh(k_y y)\sin(k_o z) \\ A_z &= \frac{k_x}{k_y k_o \rho_o}\sinh(k_x x)\sinh(k_y y)\sin(k_o z) \end{aligned} \tag{51}$$

Making a canonical transformation to variables which are locally perpendicular to the trajectory, averaging over the wiggler period, and expanding to fourth order in x and z results in the following:

$$H=\frac{1}{2}(p_x^2+p_z^2)+\frac{1}{4k_o^2\rho_o^2}(k_x^2x^2+k_z^2z^2)+\frac{1}{12k_o^2\rho_o^2}(k_x^4x^4+k_z^4z^4+3k_x^2k_o^2x^2z^2)$$
$$-\frac{\sin(k_oz)}{2k_o\rho_o}\left[p_x(k_x^2x^2+k_z^2z^2)-2k_x^2p_zxz\right] \tag{52}$$

The equations of motion can then be derived as follows:

$$\begin{aligned}
x''=&-\frac{k_x^2}{2k_o^2\rho_o^2}\left[x+\frac{1}{6}k_x^2x^3+\frac{1}{2}k_z^2xz^2\right]\\
&+\frac{\cos(k_oz)}{\rho_o}\left[\frac{1}{2}(k_x^2x^2+k_z^2z^2)+\frac{1}{4}k_x^2k_z^2x^2z^2+\frac{1}{24}(k_x^4x^4+k_y^4z^4)\right]\\
&+\frac{\sin(k_oz)}{\rho_o}k_ozz'\left[1+\frac{1}{2}k_x^2x^2+\frac{1}{6}k_z^2z^2\right]\\
z''=&-\frac{k_z^2}{2k_o^2\rho_o^2}\left[z+\frac{1}{6}k_z^2z^3+\frac{1}{2}k_x^2x^2z\right]\\
&-\frac{\cos(k_oz)}{\rho_o}\left[k_x^2xz+\frac{1}{6}k_x^2k_z^2xz^3+\frac{1}{6}k_x^4x^3z\right]\\
&-\frac{\sin(k_oz)}{\rho_o}k_ozx'\left[1+\frac{1}{2}k_x^2x^2+\frac{1}{6}k_z^2z^2\right]
\end{aligned} \tag{53}$$

The same result can be obtained by substituting directly in Eq. (1) a solution of the form:

$$x=\delta_x-\frac{\cos(k_os)}{\rho_ok_o^2} \tag{54}$$

in order to calculate the motion with respect to the reference trajectory.

To average the motion over a magnet period, it is not sufficient to set the terms in $\sin(kz)$ and $\cos(kz)$ to zero. The correct expression is as follows:

$$\begin{aligned}
x''=&-\frac{k_x^2}{2k_o^2\rho_o^2}\left[x+\frac{2}{3}k_x^2x^3\right]-\frac{k_x^2xz^2}{2\rho_o^2}\\
z''=&-\frac{k_z^2}{2k_o^2\rho_o^2}\left[z+\frac{2}{3}k_y^2z^3\right]-\frac{k_x^2x^2z}{2\rho_o^2}
\end{aligned} \tag{55}$$

i.e. linear focusing plus cubic "octupole-like" terms. The same result can be obtained by differentiating Eq. (27). In the case $k_x = 0$ we recover the result of section 4.1.

4.3 Effects on non-linear dynamics

The effects on the non-linear beam dynamics of inserting wigglers in a storage ring have been studied at many synchrotron radiation laboratories, see for example Refs. [34-37]. As a result of these studies it is clear that the effects on the dynamic aperture arise from two distinct processes:

i) The linear focusing of the device destroys the lattice symmetry, changing the phase advances between the chromaticity correction sextupoles and so exciting third order resonances. In the usual case where k_x is small, these effects are proportional to $1/\rho^2$ i.e. B_o^2/E^2.

ii) The non-linear terms in the equation of motion derived in Section 4.2 introduce amplitude dependent tune shifts and excitation of resonances. In particular the systematic octupole terms can excite fourth-order resonances, while the oscillating sextupole-like terms can also excite third-order resonances. In this case the effects are proportional to k^2/ρ^2 i.e. $B_o^2/\lambda_o^2E^2$.

The relative effect of these two processes has been demonstrated by carrying out dynamic aperture calculations with the complete equation of motion in the wiggler, and with the wiggler replaced simply by a vertically focusing element in the vertical plane [38]. The effect of non-linearities for a given wiggler has been shown to be important by repeating calculations with different period lengths while keeping the field strength and total length, and hence focusing effect, constant [35].

The beam dynamics effects were first studied by discretizing the equation of motion and inserting in a tracking program, for example BETA and RACETRACK. Problems arose however with the amount of computer time required, since a large number of steps were needed for each magnet period in order to obtain consistent results, and the fact that it was not rigorously symplectic. To overcome these difficulties an improved scheme based on a canonical integration method was later developed and included in RACETRACK [39]. More recently further improvements have been made using an algebraic mapping routine [40].

The results of various tracking studies indicate that:

i) the effect of introducing wigglers into a storage ring is in general similar in magnitude to the effect of random multipole errors in the main lattice magnets and is therefore not a limiting factor for the operation of the ring;
ii) significant reduction in dynamic aperture can be caused even by a single device, however the dynamic aperture does not continue to deteriorate linearly with an increasing number of devices;
iii) the effects are stronger in low energy rings;
iv) compensation of the tune shift introduced by the wiggler is important, since the dynamic aperture is sensitive to the tune values, but the results are not very sensitive to which scheme is used for the correction or whether the lattice functions are also corrected;
v) both linear and non-linear effects are in general important factors in dynamic aperture reduction.

Some examples of the effect of wigglers on dynamic aperture are shown in Fig. 13, showing in particular a much larger effect in the lower energy ring.

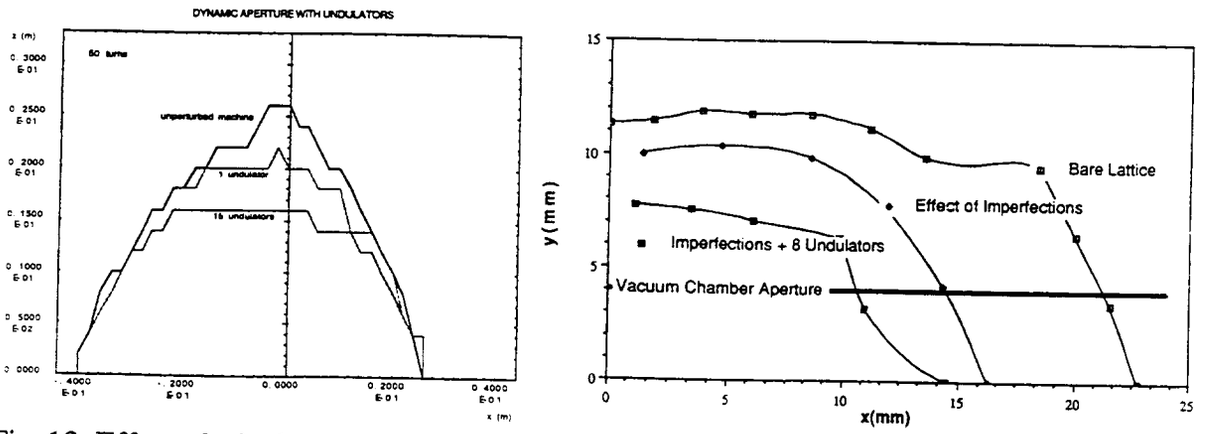


Fig. 13 Effect of wigglers on the dynamic aperture in the ESRF (left) and ALS (right) [34,36]

4.4 Effects of field errors

So far we have considered only the intrinsic non-linear effects of wigglers with an ideal periodic field distribution. In practise however a wiggler is likely to contain multipole field errors that can give rise to further beam dynamics effects. Earlier wiggler magnets have contained a number of field imperfections:

i) Skew-quadrupole field: the BESSY undulator gives rise to a change of vertical beam size with wiggler strength and an increase in the stopband width of the sum resonance $Q_x + Q_z =$ integer [32].

ii) Sextupole field: two undulators installed in SUPERACO contain sextupole errors which lead to changes in the chromaticity, since they are located in straight sections with finite dispersion [41].

iii) Skew-sextupole field : measurements of the effect of a 15-period, 1.45 T wiggler in SPEAR indicate the presence of such a field error [42].

iv) Skew-octupole field: resonances of the type $3Q_x + Q_z$ = integer and $3Q_z + Q_z$ = integer have been observed both at SPEAR, with the first permanent magnet undulator [43], and with two separate undulators on SUPERACO [41].

It should be noted that the most disturbing effects due to wiggler field errors have arisen in low energy rings, such as BESSY and SUPERACO which both have an energy of 800 MeV. Effects on the chromaticity and vertical beam size due to wiggler operation have also been seen at ALADDIN, at 0.8 and 1 GeV [44].

Progress with methods of wiggler construction, including better magnetic measurement and field correction techniques, should eliminate such errors from the newer generation of devices.

4.5 Measurement of non-linear effects

A very useful technique for determining the non-linear effects of a wiggler on the beam is to measure the tune shift as function of the horizontal and vertical position of the beam with respect to the wiggler axis. The presence of a sextupole field error is indicated by a linear tune shift with horizontal position (x); a skew-sextupole would give a tune shift with vertical position (z). Similarly, an octupole field would give a dependence of the tune shift with x^2 or z^2, while a skew-octupole would give a dependence on xz. This technique was used at SPEAR, where a skew-sextupole error was discovered [42], and at BESSY, in order to confirm the magnitude of the "pseudo-octupole" field [45], see Fig. 14.

The effect of real field errors can be distinguished from second-order effects such as these, by determining the scaling with energy (E) - the latter scales as $1/E^2$, whereas the effects of field errors scale as $1/E$.

Another method consists of performing a "resonance scan" i.e. moving the tune values so as to cross particular resonance lines, while measuring the effect on the lifetime of the stored beam. At BESSY this technique was used to determine the effect of the skew-quadrupole field errors on the $Q_x + Q_z$ = integer resonance, and of the pseudo-octupole field on the $4\,Q_z =$ integer, see Fig. 15 [32]. Similar scans at SUPERACO identified both octupole and skew-octupole driven resonances [41].

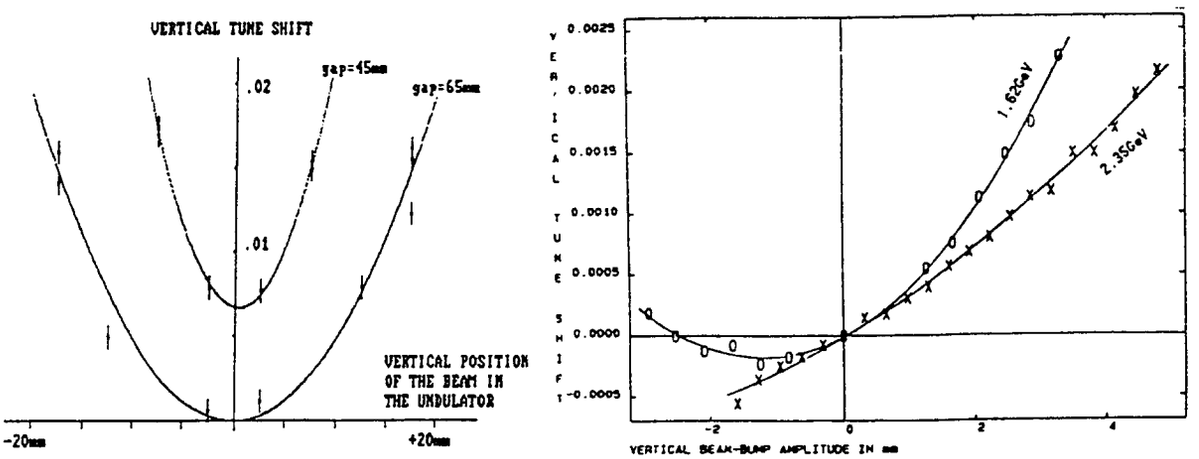


Fig. 14 Variation of vertical tune with vertical position in the wigglers at BESSY (left) and SPEAR (right) [45,42].

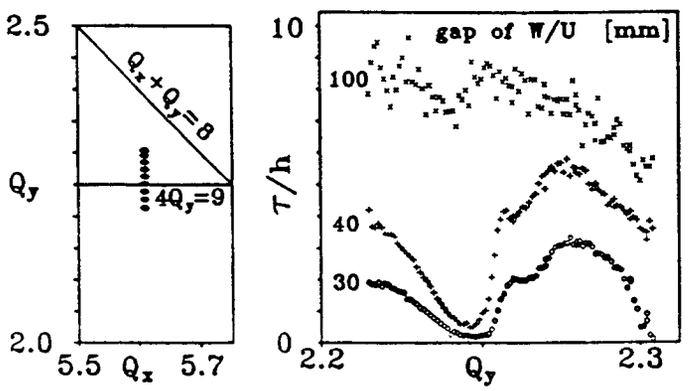


Fig. 15 Measurements of beam lifetime as function of vertical tune in BESSY with various wiggler strengths [32].

5 EFFECTS DUE TO THE EMISSION OF SYNCHROTRON RADIATION

5.1 Standard wigglers

For simplicity, we will consider the case of a wiggler with a sinusoidal field, with $k_x = 0$ and will neglect the field distribution at the end of the magnet. Then:

$$\frac{1}{\rho} = \frac{1}{\rho_o}\cos\,(k_o s) \qquad k = \frac{1}{\rho_o^2}\sin^2(k_o s) \tag{56}$$

The dispersion function is the sum of the dispersion that exists in the straight section without the wiggler (D_O, D'_O) together with a "self-dispersion" generated by the trajectory in the wiggler. Since the displacement, x, is inversely proportional to energy, E, we have $dx = -x\,(dE/E)$, and hence $D = -x$, and similarly $D' = -x'$. Thus the dispersion is given in general by:

$$\begin{aligned} D(s) &= D_o + D_o' s + \frac{\cos(k_o s)}{\rho_o k_o^2} \\ D'(s) &= D_o' - \frac{\sin(k_o s)}{\rho_o k_o} \end{aligned} \tag{57}$$

The contributions to the five Synchrotron Radiation Integrals [46] for a wiggler of length L can then be derived as follows:

$$\begin{aligned}
\Delta I_1 &= \oint \frac{D}{\rho} = \int \frac{\cos^2(k_o s)}{\rho_o^2 k_o^2} = \frac{L}{2\rho_o^2 k_o^2} \\
\Delta I_2 &= \oint \frac{1}{\rho^2} = \int \frac{\cos^2(k_o s)}{\rho_o^2} = \frac{L}{2\rho_o^2} \\
\Delta I_3 &= \oint \frac{1}{|\rho^3|} = \int \frac{|\cos^3(k_o s)|}{\rho_o^3} = \frac{4L}{3\pi \rho_o^3} \\
\Delta I_4 &= \oint \frac{D}{\rho^3} - \frac{2kD}{\rho} = -\int \frac{\cos^4(k_o s)}{\rho_o^4 k_o^2} + 2\int \frac{\sin^2(k_o s)\cos^2(k_o s)}{\rho_o^4 k_o^2} = -\frac{L}{8\rho_o^4 k_o^2} \\
\Delta I_5 &= \oint \frac{H}{|\rho^3|} \quad \text{where } H(s) = \gamma D^2 + 2\alpha D D' + \beta D'^2
\end{aligned} \tag{58}$$

To evaluate ΔI_5 a numerical calculation is generally needed, however two extreme cases can be considered:

i) large natural dispersion. In this case the self-dispersion can be neglected and so after averaging $1/\rho^3$ over the wiggler length one obtains:

$$\Delta I_5 \approx \frac{4}{3\pi} \frac{\langle H \rangle L}{\rho_o^3} \tag{59}$$

ii) zero natural dispersion. The dominant term in the expression for H is $\beta D'^2$ and so:

$$\Delta I_5 \approx \int \frac{\beta D'^2}{|\rho^3|} \approx \langle \beta \rangle \int \frac{\sin^2(k_o s)|\cos^3(k_o s)|}{\rho_o^5 k_o^2} = \frac{4}{15\pi} \frac{\langle \beta \rangle L}{\rho_o^5 k_o^2} \tag{60}$$

Making use of the above formulae we can now deduce the effects on the beam due to changes in the SR integrals:

i) Momentum compaction factor: $\alpha_p = I_1 / L_{tot}$, where L_{tot} is the ring circumference.
The change introduced is therefore:

$$\Delta \alpha_p = \frac{\Delta I_1}{L_{tot}} = \frac{L}{L_{tot}} \frac{1}{2\rho_o^2 k_o^2} \tag{61}$$

The effect in most cases is negligible, for example with $L_W = 5$ m, $L_{tot} = 100$ m, $B_O = 1.5$ T, $E = 1.0$ GeV, $\lambda_O = 0.1$ m we have $\Delta\alpha = 1.3\ 10^{-6}$.

ii) Damping partition numbers : $J_x = 1 - (I_4/I_2)$, $J_\varepsilon = 2 + (I_4/I_2)$
The change to the J_x damping partition number can be written as follows:

$$\frac{J_x'}{J_x} = \frac{1 + \dfrac{\Delta I_2 - \Delta I_4}{I_2 - I_4}}{1 + \dfrac{\Delta I_2}{I_2}} \tag{62}$$

From the expressions above it can be seen that $\Delta I_4/\Delta I_2 = 1/4\rho_o^2 k_o^2$ and hence ΔI_4 can be neglected compared to ΔI_2. Also, in most rings I_4 is very small, and hence J_x is close to unity. In most cases therefore J_x, and hence J_z and J_ε, are unchanged by the action of the wiggler.

iii) Damping times: $\tau_i = 3T_o/r_o\gamma^3 J_i I_2$, where $i = x$, z or ε.
Since the effect of the wiggler is to increase I_2, all damping times are thereby reduced.

iv) Energy spread: $\left(\frac{\sigma_E}{E}\right)^2 = C_q\gamma^2 \frac{I_3}{2I_2 + I_4}$, where $C_q = \frac{55}{32\sqrt{3}}\frac{\hbar}{mc} = 3.84\ 10^{-13}$ m

The change in energy spread can therefore be written as follows:

$$\left(\frac{\sigma_E'}{\sigma_E}\right)^2 = \frac{1+\frac{\Delta I_3}{I_3}}{1+\frac{2\Delta I_2 + \Delta I_4}{2I_2 + I_4}} \tag{63}$$

Neglecting as before ΔI_4 and I_4, then in the case of a sinusoidal wiggler:

$$\left(\frac{\sigma_E'}{\sigma_E}\right)^2 = \frac{1+\frac{4}{3\pi}\frac{L}{2\pi\rho}\frac{\rho^3}{\rho_W^3}}{1+\frac{1}{2}\frac{L}{2\pi\rho}\frac{\rho^2}{\rho_W^2}} \tag{64}$$

Thus, the energy spread is increased if $\rho/\rho_W > 3\pi/8$, i.e. roughly when the wiggler field exceeds that of the bending magnets. Assuming constant synchrotron oscillation frequency, the bunch length is changed by the same factor as the energy spread. Figure 16 shows the variation of (σ_E'/σ_E) as a function of magnetic field strength for wigglers of two different lengths in 1.5 and 6 GeV rings. The result is independent of the wiggler period length. It can be seen that the change in energy spread increases with increasing wiggler field strength and length, and reduces with higher electron beam energies.

v) Emittance: $$\varepsilon_x = C_q\gamma^2 \frac{I_5}{I_2 - I_4}$$

The change in emittance can therefore be written as follows:

$$\frac{\varepsilon_x'}{\varepsilon_x} = \frac{1+\frac{\Delta I_5}{I_5}}{1+\frac{\Delta I_2 - \Delta I_4}{I_2 - I_4}} \tag{65}$$

Neglecting I_4 and ΔI_4 it can be seen that the emittance is increased if $\Delta I_5/I_5 > \Delta I_2/I_2$.

Figure 17 shows the emittance ratio $(\varepsilon_x'/\varepsilon_x)$ as a function of field strength for wigglers of different period length and total length in rings of 1.5 and 6 GeV, assuming zero natural dispersion at the wiggler location. At low field strength the emittance reduces, due to increase of the I_2 term. The reduction therefore depends only on the total length, not the period length. At higher fields the I_5 term which causes emittance growth becomes increasingly important. This is particularly true for longer period lengths (curves a and c), in agreement with Eq. (60).

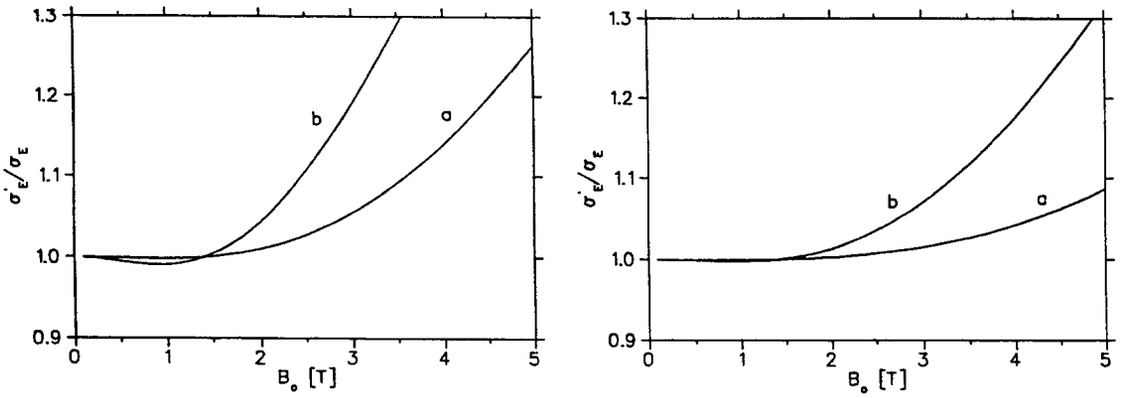


Fig. 16 Variation of energy spread with wiggler field strength in 1.5 GeV (left) and 6 GeV(right) rings. Ring parameters: $B = 1.2$ T, $J_x = 1.0$. Wiggler parameters: (a) $L = 1$ m, (b) $L = 5$ m.

The condition for emittance increase is given by:

$$\frac{4}{3\pi}\frac{\langle H\rangle_W L}{\rho_W^3}\frac{\rho^3}{2\pi\rho} > \frac{L}{2\rho_W^2}\frac{\rho^2}{2\pi\rho} \tag{66}$$

which can be expressed as follows:

$$\lambda_o^2 B_o^3 > 5.87\ 10^9\ \frac{\varepsilon_x E_{[\mathrm{GeV}]}}{\langle\beta_x\rangle} \tag{67}$$

and which therefore depends on the period length, independent of the wiggler length. At higher energy the effect on the emittance is reduced, and the point at which the emittance starts to increase is shifted to higher field strengths, in accordance with Eq. (67).

It follows from the expression above that in most practical situations the emittance reduces, except for very high field wigglers in low energy, low emittance rings. Wigglers have therefore been proposed as a means of reducing the emittance, for example in PEP, for operation as a low emittance synchrotron radiation source [47]. Wigglers are also proposed in many new damping ring designs, to reduce damping times and emittance.

Figure 18 shows the effect of non-zero dispersion (D_o) in the wiggler straight and it can be seen that with sufficiently large dispersion the emittance increases. The condition for the natural dispersion to have a larger effect than that of the wiggler self-dispersion is given by: $\langle H\rangle_W > \frac{\beta}{5k_o^2\rho_W^2}$, or in terms of the dispersion: $D_o > \frac{\beta}{\sqrt{5}k_o\rho_W}$ or $D_o' > \frac{1}{\sqrt{5}k_o\rho_W}$, assuming $\alpha = 0$. Since $1/k\rho_W << 1$ this condition is easily met. Thus even if the wiggler straight has a nominal zero-dispersion, the residual dispersion arising from errors can be the dominant effect. In this case the condition for emittance increase is given by:

$$\frac{\langle H\rangle_W}{\langle H\rangle}\frac{\rho}{\rho_W} > \frac{3\pi}{8} \tag{68}$$

which can be written as follows:

$$\langle H \rangle_W B_o > 2.68\ 10^6 \frac{\varepsilon_x}{E_{[\mathrm{GeV}]}} \tag{69}$$

The residual dispersion must therefore be carefully controlled in low emittance rings with high field wigglers to avoid any unwanted emittance increase. For example, for a 1.5-T wiggler in the 1.5-GeV ring of the above example, the dispersion must be kept below the following limits: $D_O < 0.29$ m, $D'_O < 0.029$ m.

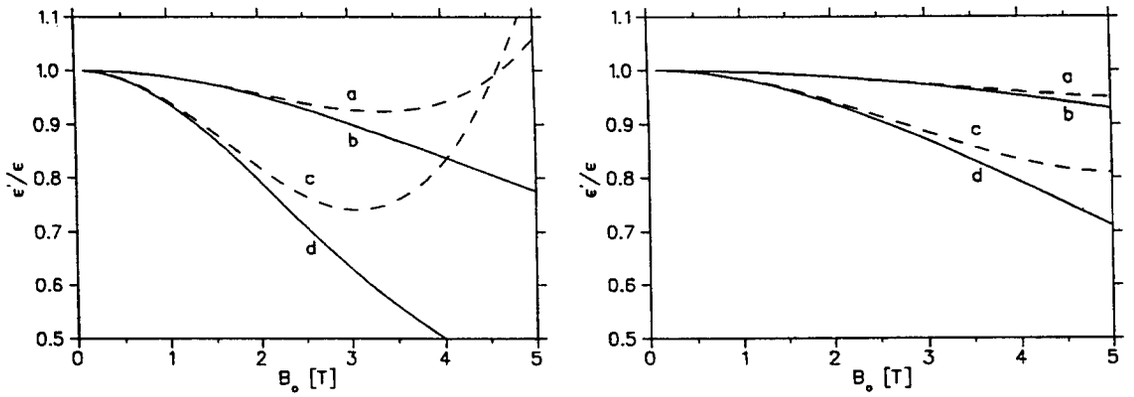


Fig. 17 Variation of emittance with wiggler field strength in 1.5-GeV (left) and 6-GeV (right) rings. Ring parameters: $B = 1.2$ T, $J_x = 1.0$, $\varepsilon_x = 7\ 10^{-9}$ m rad, $\beta_x = 10$ m. Wiggler parameters: (a) $L = 1$ m, $\lambda_o = 0.25$ m, (b) $L = 1$ m, $\lambda_o = 0.05$ m, (c) $L = 5$ m, $\lambda_o = 0.25$ m, (d) $L = 5$ m, $\lambda_o = 0.05$ m.

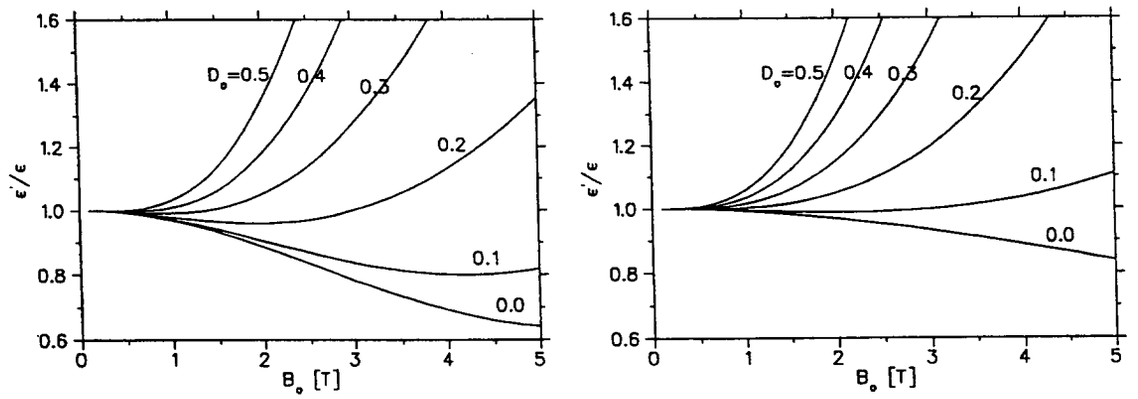


Fig.18 Variation of emittance with wiggler field strength and natural dispersion at the centre of the wiggler straight section (D_O) in 1.5-GeV (left) and 6-GeV (right) rings. Ring parameters as Fig. 17. Wiggler parameters: $L = 2.5$ m, $\lambda_o = 0.10$ m.

A practical example of the use of wigglers to change beam properties is in LEP [48]. A set of "damping wigglers", located in dispersion free regions, are used to increase the energy spread and bunch length by a factor 5-6 at injection energy to increase beam stability, while a series of "emittance wigglers", located in regions with non-zero dispersion, are used to increase the beam size at high energy in order to optimize the luminosity [49].

An emittance increase due to the operation of two superconducting wigglers is unavoidable in the SRS, since no zero-dispersion straights are available. The emittance

increase due to the latest device was kept as low as possible (50 %) by minimizing ΔI_5 and therefore $\int 1/\rho^3$ ds in the magnet design [50].

5.2 Gradient and non-linear wigglers

In combined function strong focusing lattices $(I_4/I_2) \sim 2$, i.e. $J_x \sim -1$, and so the radial motion is anti-damped [1]. Gradient wigglers, or dipole-quadrupole wigglers, were proposed by Robinson as a means of overcoming this difficulty [3] and they were subsequently employed at CEA to permit operation as a storage ring [4], and also at the PS for operation with electrons [51]. It follows from Eq. (62) that a large negative change to the I_4 integral is required, which can be obtained in a wiggler with an alternating field and field gradient such that k/ρ is positive, if located at a position of large positive dispersion. Gradient wigglers can also be used to reduce the emittance in a storage ring. For a wiggler loacted at a position of large dispersion and beta function the dominant effect in Eq. (65) is the reduction in I_4, or in other words the change in damping partition, J_x. Since the emittance can be expressed as $\varepsilon_x = C_q \gamma^2 I_5 / J_x I_2$ and since the damping partition can be changed from 1 (for a conventional separated function lattice) to 2 without loss of damping in the longitudinal plane, it follows that a reduction in emittance of a factor of 2 is achievable by this means.

Non-linear wigglers have been proposed as a means of producing a dependence of the damping partition (J_ε) with energy deviation, ε [5,6]. A dipole-sextupole wiggler can give a linear dependence, $J_\varepsilon \sim \varepsilon$, which can be used to increase the damping aperture by making $dJ_\varepsilon/d\varepsilon = 0$. A dipole-octupole wiggler gives a quadratic dependence, $J_\varepsilon \sim \varepsilon^2$, which alters the equilibrium energy and charge distribution. A quadrupole-sextupole wiggler has also been proposed [7].

To gain some insight into the effect of such devices on the beam, we consider a series of constant field regions with alternating polarity, but constant gradient, sextupole and octupole terms, located at a position with non-zero dispersion:

$$B_z(x) = B + B'x + \frac{B''x^2}{2} + \frac{B'''x^3}{6} \tag{70}$$

where $x = D\varepsilon$ and ε is the relative energy deviation. Multiplying by $e/\gamma mc$ we have:

$$-\frac{1}{\rho}(x) = -\frac{1}{\rho_W} + kD\varepsilon + \frac{r}{2}D^2\varepsilon^2 + \frac{q}{6}D^3\varepsilon^3 \tag{71}$$

where

$$\frac{1}{\rho_W} = -\frac{e}{\gamma mc}B \quad k = \frac{e}{\gamma mc}B' \quad r = \frac{e}{\gamma mc}\frac{B''}{2} \quad q = \frac{e}{\gamma mc}\frac{B'''}{6} \tag{72}$$

The damping partition number is defined by:

$$J_\varepsilon = \frac{dU}{dE}\frac{E_o}{U} = \frac{dU}{d\varepsilon}\frac{1}{U} \tag{73}$$

The energy loss per turn (U) depends on $\int 1/\rho^2$ and hence the total energy loss including the wiggler (U_W) is given by:

$$U_W = U_o + U_o \frac{\rho}{2\pi}\int\left(-\frac{1}{\rho_W} + kD\varepsilon + \frac{rD^2\varepsilon^2}{2} + \frac{qD^3\varepsilon^3}{6}\right)^2 ds \tag{74}$$

We have then:

$$J_\varepsilon = J_{\varepsilon o} + \frac{U_o}{U_W}\frac{\rho}{2\pi}\int \frac{d}{d\varepsilon}\left(-\frac{1}{\rho_W} + kD\varepsilon + \frac{rD^2\varepsilon^2}{2} + \frac{qD^3\varepsilon^3}{6}\right)^2 ds \tag{75}$$

We can now examine the effect of the various proposed combinations of field components:

i) Dipole-quadrupole wiggler:

$$J_\varepsilon = J_{\varepsilon o} - \frac{\rho L}{2\pi}\left\{\frac{2kD}{\rho_W} + 2k^2D^2\varepsilon\right\} \tag{76}$$

ii) Dipole-sextupole wiggler:

$$J_\varepsilon = J_{\varepsilon o} - \frac{\rho L}{2\pi}\frac{2rD^2\varepsilon}{\rho_W} \tag{77}$$

iii) Dipole-octupole wiggler:

$$J_\varepsilon = J_{\varepsilon o} - \frac{\rho L}{2\pi}\frac{qD^3\varepsilon^2}{\rho_W} \tag{78}$$

iv) Quadrupole-sextupole wiggler:

$$J_\varepsilon = J_{\varepsilon o} - \frac{\rho L}{2\pi}\{2k^2D^2\varepsilon + 3krD^3\varepsilon^2 + 6k^2D^2\varepsilon^2\} \tag{79}$$

In the above $J_{\varepsilon o}$ represents the energy oscillation damping partition number in the absence of the wiggler, which is itself a function of energy deviation. A dipole-sextupole wiggler can thus be used to vary the linear dependence of J_ε with ε, and make it zero if desired. A combination of dipole-quadrupole and dipole-octupole can result in a damping partition number variation of the form:

$$J_\varepsilon = -a + b\varepsilon^2 \tag{80}$$

i.e. anti-damped for particles of the nominal energy, but damped for particles of large energy deviation. As a result it is possible to modify the energy and longitudinal charge distributions, for example to reduce the peak charge density and so reduce the effects of instabilities, as shown in Fig. 19 [52].

5.3 Spin polarization

The polarization rate is [53]:

$$\frac{1}{\tau_p} = \frac{5\sqrt{3}\hbar r_o}{8m}\gamma^5\frac{I_3}{L_{tot}} \tag{81}$$

Since I_3 is always increased by the effect of wigglers it follows that this can be used to increase the polarization rate. However, the asymptotic polarization rate is affected also. For an ideal machine this is given by:

$$P_{\infty} = \frac{8}{5\sqrt{3}} \frac{I_{3a}}{I_3} \tag{82}$$

where

$$I_{3a} = \int \frac{1}{\rho^3} ds \tag{83}$$

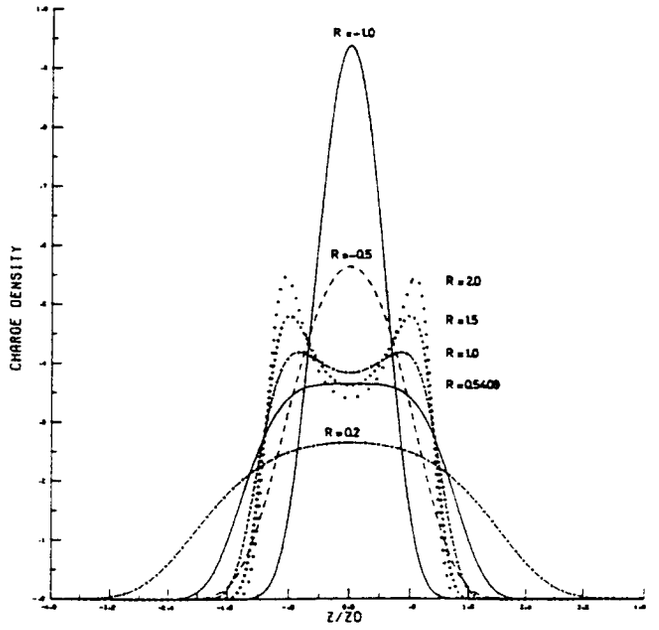


Fig. 19 Calculated longitudinal bunch profiles in LEP due to the action of a dipole-octupole wiggler [52].

In a conventional wiggler the positive and negative poles have equal strength and hence I_{3a} is approximately zero, which leads to a reduction of the asymptotic polarization rate. In order to increase the polarization rate while at the same time maintaining a high polarization level requires an "asymmetric" wiggler with different positive (B+) and negative (B-) field strengths. In order that the field integral for each period is zero, and so produces no net deflection, the pole lengths must also vary so as to satisfy: B+ L+ = B- L-. The ratio of field strengths is called the "asymmetry parameter", r = B+ / B-. Thus if r is large the changes to I_3 and I_{3a} are dominated by the positive poles and so are similar in magnitude, and hence the asymptotic polarization level is not affected very much.

For example, LEP has 12 "polarization wigglers" with an asymmetry parameter r = 8.0 [54]. Calculations show that they should reduce the polarization time from 310 min to 36 min at 46 GeV with 88% asymptotic degree of polarization [53].

* * *

REFERENCES

[1] R. P. Walker, CERN Accelerator School 5th General Accelerator Physics Course, CERN 94-01, p. 481.
[2] J.M. Paterson, J.R. Rees and H. Wiedemann, SPEAR-186, PEP-125, July 1975.
[3] K.W. Robinon, Phys. Rev. 111 (1958) 373.

[4] A. Hofmann et al., Proc. 6th Int. Conf. High Energy Accelerators, Cambridge, 1967, p. 123.
[5] H.G. Hereward, CERN AR/Int. SR/61-15 (1961).
[6] A. Hofmann, LEP Note 192 (1979).
[7] J.M. Jowett, Proc. 12th Int. Conf. High Energy Accelerators, Fermilab, August 1983, p. 300.
[8] A.A. Sokolov and I.M. Ternov, *Synchrotron Radiation*, (Pergamon Press, 1968).
[9] W.S. Trzeciak, IEEE Trans Nucl. Sci. NS-18 (1971) 213.
[10] A.S. Artamonov, Nucl. Instr. Meth. 177 (1980) 247.
[11] V.P. Suller et al., IEEE Trans. Nucl. Sci. NS-30 (1983) 3127.
[12] M. Berndt et al., IEEE Trans. Nucl. Sci. NS-26 (1979) 3812.
[13] E. Hoyer et al., IEEE Trans. Nucl. Sci. NS-30 (1983) 3118.
[14] H. Motz, J. Appl. Phys. 22 (1951) 527.
[15] H. Motz et al., J. Appl. Phys. 24 (1953) 826.
[16] R.M. Phillips, Nucl. Instr. Meth. A272 (1988) 1.
[17] J.M.J. Madey, J. Appl. Phys. 42 (1971) 1906.
[18] K. Halbach et al., IEEE Trans Nucl. Sci. NS-28 (1981) 3136.
[19] N.A. Vinokurov, Nucl. Instr. Meth. A246 (1986) 105.
[20] J. Bosser et. al., J. Physique Lett. 45 (1984) L343.
[21] P. Elleaume, Proc. 1991 US Particle Accelerator Conference, IEEE 91CH3038-7, p. 1083.
[22] J. Rossbach and P. Schmüser, CERN Accelerator School 5th General Accelerator Physics Course, CERN 94-01, p. 17.
[23] V.P. Suller, SRS/NS/75/67 (1975).
[24] M. Bassetti et. al., LNF-77/26 (R) (1977).
[25] E.T. Scharlemann, J. Appl. Phys. 58 (1985) 2154.
[26] P. Elleaume, Proc. 3rd European Accelerator Conference, Berlin, March 1992, p. 661.
[27] R.P. Walker, Nucl. Instr. Meth. 214 (1983) 497.
[28] R.P. Walker, SRS/MI/86/33, July 1986.
[29] M.W. Poole et al., Proc. 1989 US Particle Accelerator Conference, p. 208.
[30] P. Kuske et al., Proc. 1st European Particle Accelerator Conference, Rome, June1988, p. 1214.
[31] M. Katoh, KEK Report 86-12, March 1987
[32] P. Kuske and J. Bahrdt, Proc. 2nd European Particle Accelerator Conference, Nice, June 1990, p. 1417.
[33] L. Smith, ESG Tech. Note 24 (1986).
[34] ESRF Foundation Phase Report, Feb. 1987, p. 82.
[35] L. Tosi, Proc. 1st European Particle Accelerator Conference, Rome, June1988, p. 884.
[36] A. Jackson, Proc. 1989 US Particle Accelerator Conference, p. 1752.
[37] Y. Chae and E.A. Crosbie, Proc. 1993 US Particle Accelerator Conference, p. 282.
[38] A. Wrulich, ST/M-87/18, Oct. 1987.
[39] R. Nagaoka and L. Tosi, ST/M-90/6, Mar. 1990.
[40] G. Wüstefeld and J. Bahrdt, Proc. 1991 US Particle Accelerator Conference, p. 266.
[41] P. Brunelle, Particle Accelerators 39 (1992) 89.
[42] J. Safranek, Proc. 1989 US Particle Accelerator Conference, p. 1307.
[43] P. Morton, SPEAR-223 (1981).
[44] W.S. Trzeciak et al., Nucl. Instr. Meth. Phys. Res. A291 (1990) 413.
[45] P. Kuske, Proc. Wiggler-Undulator Workshop, Berlin, Oct. 1989.
[46] R.H. Helm et al., IEEE Trans Nucl. Sci. NS-22 (1975) 1765.
[47] H. Wiedemann, Nucl. Instr. Meth. A266 (1988) 24.
[48] J.M. Jowett and T.M. Taylor, IEEE Trans. Nucl. Sci. NS-30 (1983) 2581.
[49] A. Hofmann, private communication.
[50] M.W. Poole et al., Proc. 1989 US Particle Accelerator Conference, p. 1250.
[51] LEP Design Report, Vol. 1, CERN-LEP/TH/83-29, June 1983.
[52] A. Hofmann and J.M. Jowett, CERN/ISR-TH/81-23 (1981).
[53] A. Blondel and J.M. Jowett, LEP Note 606 (1988).
[54] D. Brandt et. al., Proc 3rd European Accelerator Conference, Berlin, 1992, p. 649.

FREE ELECTRON LASER

A. Gaupp
BESSY, Berlin, Germany

Abstract
A free electron laser (FEL, see Fig. 1) requires an electron beam that is made to 'wiggle' in a transverse direction usually by a static magnetic field periodic in space (wiggler or undulator; in what follows we shall use both terms synonymously, although the term 'undulator' would be most appropriate). The transverse acceleration of the charge causes the emission of radiation known as undulator radiation among the synchrotron radiation community. This emission process is enhanced (i.e. stimulated) by an 'external' optical field. If the radiation is confined within an optical cavity of sufficiently small losses the optical power may build up, i.e. the system is lasing. At increasing power levels the gain decreases until it balances the losses and the laser saturates. In this contribution we discuss the FEL from a more qualitative point of view in order to provide physical insight. More extensive discussions can be found in the literature and in the CERN accelerator school proceedings [1,2].

1. HISTORY

In the early fifties H. Motz at Oxford, England, investigated what we call today undulator radiation. Already in 1955 he proposed to use the device as an amplifier. Around 1960 R. Philips at General Electric developed the Ubitron (undulating beam interaction) [3]. It generated MW power levels at millimeter wavelength by coupling a fast electro magnetic wave inside a wave guiding structure to an undulating electron beam. At that time he developed a major part of FEL theory. However, the US Air Force funding his work limited the maximum voltage to 100 kV, and thus the wavelength shortening effect could not be exploited.

The work was classified and forgotten to allow for reinvention by J. Madey at Stanford, USA in 1970. Madey not only proposed and built a working device [4] based on what he initially called 'stimulated bremsstrahlung', but also found a powerful name: free electron laser (FEL). Since then the FEL has been discussed on many occasions, and a growing number of laboratories are doing experimental work.

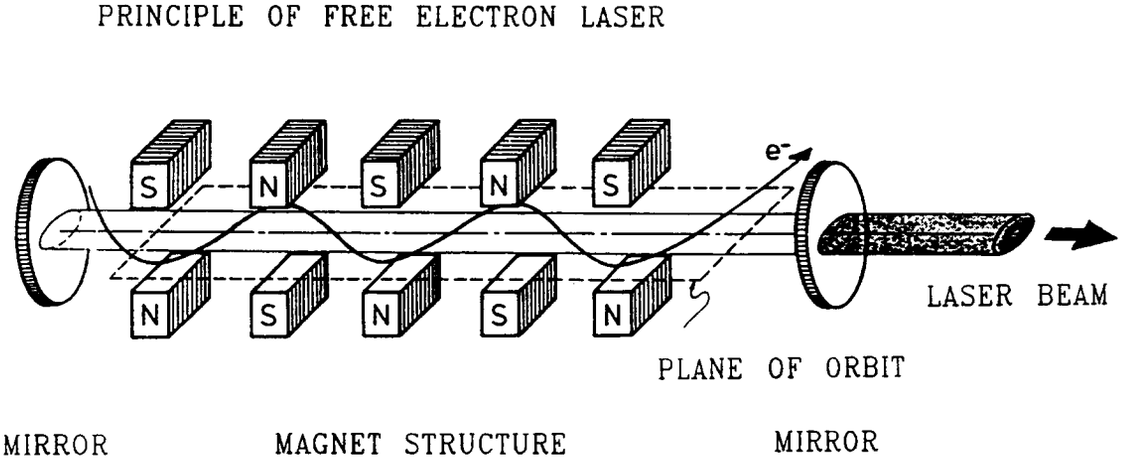


Fig. 1 Main components of a free electron laser

2. INTRODUCTION

There are a number of reasons why the FEL has attracted scientists (and non-scientists as well). Among them are:

1) The FEL is basically easy to understand. It can be described by classical electrodynamics, and performs in a predictable manner.

2) The FEL can be scaled to high power levels since no material is involved in the light generating process (except for the mirrors).

3) The FEL can be tailored to fairly high efficiency if the installation is large enough. A wall-plug efficiency of the order of 1% has been predicted.

4) The FEL may operate at any desired wavelength since no atomic or molecular levels are involved; any limitations are of a technical nature. In particular, it produces radiation in the near and far infra-red where few other lasers are available. More spectacular is the prospect to create a VUV laser or even an x-ray laser.

5) The FEL wavelength may be tuned continuously.

The price to be paid is the high quality electron beam needed to drive an FEL. This usually leads to large and expensive systems. Efforts are under way to reduce the size (and cost!) of components.

In this contribution we first discuss the FEL mechanism. Then we will mention some of the experimental work done to date. This will be an opportunity to introduce variations of the simple FEL scheme. Although the use of FEL radiation is evolving this topic is beyond the scope of this contribution and we shall conclude only with some general remarks.

3. HOW TO BUILD ANY LASER

Before proceeding let us look at the task of developing any new type of laser. Five main steps are required (cf. Fig. 2):

a) Search for an active medium that converts energy into radiation by spontaneous emission. Depending on the active medium there are different ways to supply the energy, e.g. via flash lamps in a YAG-Laser, via a discharge in an eximer laser, an electric current in a semiconductor laser, another laser in a typical cw dye laser, or a relativistic electron beam in an FEL, etc. The energy source is called the pump.

b) Amplify an external optical signal injected into the pumped active medium using stimulated emission. This requires an inversion of the transition involved, i.e. the pumping process must be such that the higher (atomic/molecular) level has a higher population than the lower one. Proper selection of the transition and sufficient pumping power is necessary.

c) Build an optical resonator around the active medium. Depending on the wavelength involved this may be a microwave cavity (maser); in the visible domain a resonator is formed, e.g. from two concave mirrors. The spontaneous radiation trapped in the resonator repetitively interacts with the active medium and is amplified. If the average gain exceeds the average losses the power builds up: the laser is above the (laser instability) threshold. In other words: the system starts lasing.

d) While the power inside the cavity builds up, the gain decreases due to some power dependent mechanism until, in a steady state, the gain equals the losses (saturation).

There may be many different mechanisms for saturation, each having a different power dependence. The power level at saturation determines the available output power and thus the usefulness of a given system. There is a high premium for improving the saturation behaviour.

e) Search for scaling laws, i.e. how the performance depends on the different variables. The usefulness of a system depends mainly on the scaling of cost with output power.

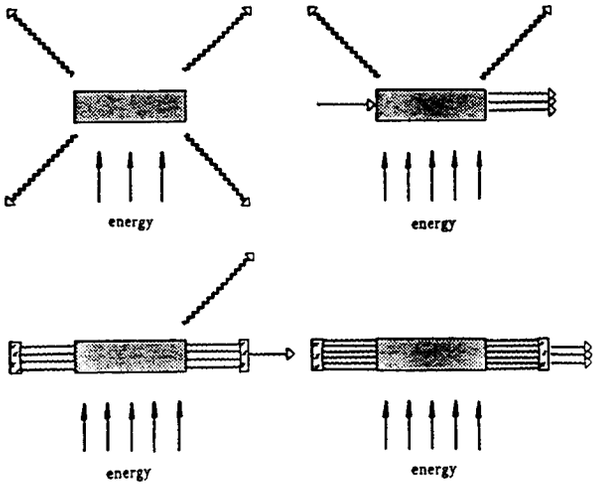


Fig. 2 Laser development, simplified. a) An active medium emits spontaneous radiation when pumped. b) By stronger pumping inversion is achieved and an external photon is amplified by stimulated emission. c) If the gain exceeds the losses of an optical resonator the laser threshold is reached. d) Due to some saturation mechanism the gain drops for increasing power density until, in a steady state, the gain balances the losses including the useful one through the output coupling mirror.

In the following discussion of the FEL interaction we shall mention spontaneous emission and spend some more time on the stimulated emission process and on the saturation behaviour. There are two models describing the FEL interaction correctly. As we shall see, they are complementary. Chronologically the 'scattering model' came first [4] but for practical work the 'classical model' is more important.

3.1 The first FEL model: stimulated scattering of radiation

This model treats the wiggler and optical fields on an equal footing by considering a pseudo rest frame (see Fig. 3) [5]. Within this frame of reference the electron executes 'mostly' transverse oscillations, i.e. in the laboratory this frame moves along the axis approximately with the averaged speed of the electrons. Since this is a relativistic velocity the wiggler period as seen from the pseudo rest frame is contracted and the optical wavelength is extended. More precisely, the pseudo rest frame is defined by the requirement that the wiggler period and wavelength of the optical field are equal. To this velocity there corresponds an electron energy $E_r = \gamma_r mc^2$ called the resonance energy. Electrons may or may not have this energy, i.e. they can also move longitudinally in the pseudo rest frame. However, the electron energy $E = \gamma mc^2$ is close to the resonance, and the motion with respect to the pseudo rest frame is non relativistic.

For $\gamma_r >> 1$, a good approximation is to replace the wiggler field in the pseudo rest frame by a free optical wave. This approximation is named after Weizacker and Williams and is widely used in plasma physics.

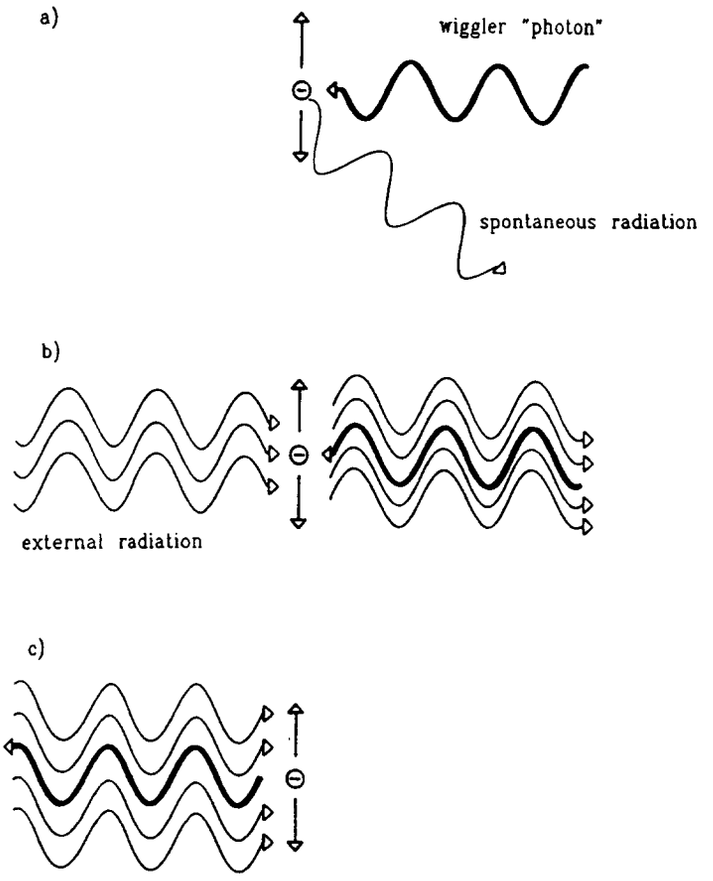


Fig. 3 FEL interaction by scattering model.
a) Spontaneous emission by scattering of wiggler 'photons' into real photons.
b) Stimulated scattering of a wiggler photon into the 'forward' direction in the presence of 'external' photons.
c) Stimulated scattering of an external real photon into the 'backward' direction in the presence of wiggler 'photons'.

The model uses quantum mechanical ideas. Spontaneous undulator radiation is described as a scattering of a wiggler photon off an electron at rest (on-resonance situation assumed). Photons scattered backward (which in the lab frame is more naturally labelled the forward direction!) have a wavelength which is shortened by a further factor γ_r. Taking into account the reduction of the longitudinal velocity (in the lab frame) due to the transverse wiggling, the resonance condition becomes

$$\lambda_l = \frac{\lambda_0}{2\gamma_r^2}\left[1+\frac{1}{2}K^2\right] \tag{1}$$

where λ_l is the optical wavelength in the lab frame, λ_0 is the period of a static magnetic wiggler in the lab frame. K is a dimensionless parameter of the order of unity describing the strength of the wiggler field and is given by

$$K = \frac{e\lambda_0 B_0}{2\pi mc^2} = 93\cdot(\lambda_0 / m)\cdot(B_0 / \text{Tesla}) \; . \tag{2}$$

B_0 is the amplitude of the wiggler field and e, m, and c the electronic charge, the electron mass, and the speed of light, respectively. The line shape is the Fourier transform of the wiggler field as shown in Figs. 4 and 5a.

We note the freedom in choosing the wavelength. For example, a 3.2 cm wiggler period and a 42 MeV electron beam emit, for $K = 0.7$, at an infra-red wavelength of $\lambda_l = 3\mu$ (these are the parameters of the first FEL built by Madey and co-workers in Stanford, Cal., in 1976 [6]).

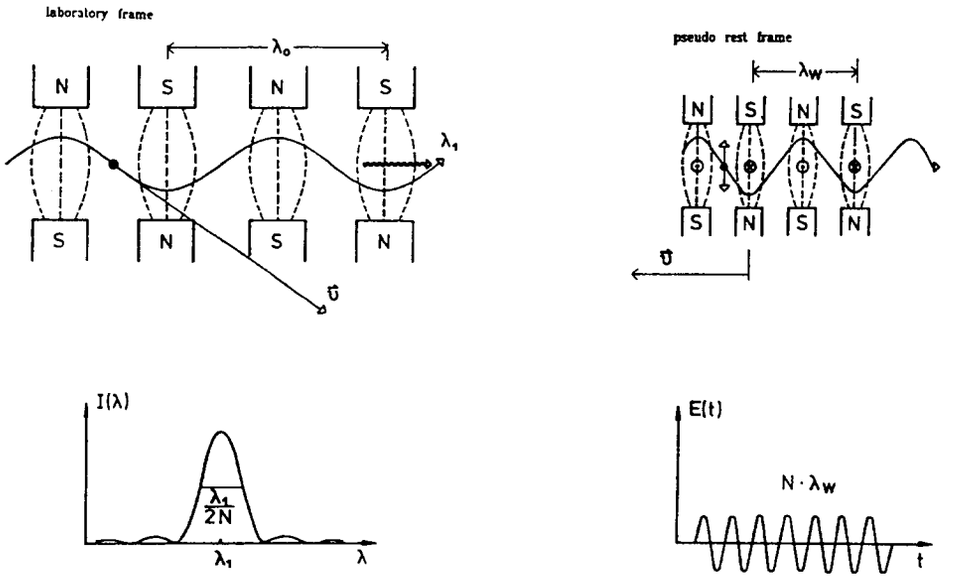


Fig. 4 The magnetic wiggler field as seen in the pseudo rest frame acquires an electric field component which at the location of the electron has a time variation $E(t)$ as shown. The Fourier transform $I(\lambda)$ is the spectrum observable in the lab frame.

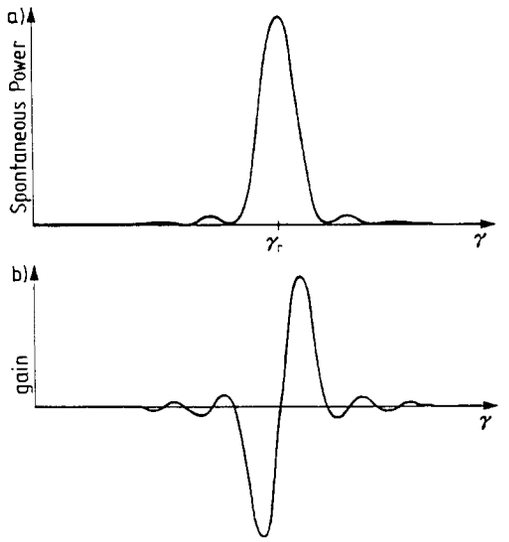


Fig. 5 a) Spontaneous spectrum as a function of electron energy; such a signal could be obtained experimentally by recording the transmission of a monochromator while the electron beam energy is swept through resonance. The fractional width of this curve is approximately $1/2N$. b) Gain of a fixed wavelength as a function of electron energy. Gain is proportional to the derivative of the spontaneous spectrum.

An FEL may be tuned by varying the electron energy or by varying the magnetic field amplitude. The latter procedure is usually preferred.

The scattering can be stimulated by applying an external driving field. The stimulating field should enhance the back scattering, i.e. it must co-propagate with the electron in the

laboratory frame and must have a frequency as given by Eq. (1) above and must have the same polarization as the spontaneous radiation. The stimulated scattering rate is proportional to the number of incident photons which in turn is proportional to the number of 'wiggler photons' N_{wigg}, and to the number of driving photons N_{ext}. (Remember the Weizäcker-Williams approximation.) This rate is very high.

However, the process discussed so far is only half the story. There is also the scattering of an external photon back into the wiggler field. This process is stimulated by the wiggler field. Thus this scattering rate is also proportional to $N_{ext} \cdot N_{wigg}$. Indeed it can be shown that for electrons exactly on resonance, the two rates are equal and the net effect is zero! Thus an FEL provides no gain on resonance!

A positive gain, i.e. a finite net scattering rate from the wiggler field into the external field, requires injection of electrons above resonance. It turns out that in the description of the total process the recoil of the electron due to scattering has to be taken into account. This is a small effect, as can be seen by comparing the photon energy in the pseudo rest frame (3.2 meV in the above example) to the rest energy of the electron of 511 keV, the ratio being less than 10^{-8}.

However, the scattering rate is high enough that the small recoils add up to a significant change of the electron energy. Since 'forward' and 'backward' scattering rates are about equal, the FEL interaction increases the energy spread of the electron beam (beam heating). When the electron energy gets far enough from resonance the resonance condition is violated and the interaction ceases. This is basically the saturation mechanism within this model.

A distribution of electrons injected off resonance experience an asymmetric energy spread and the mean kinetic energy of the electron beam E_{kin} changes. By energy conservation, the optical energy E_{opt} then changes correspondingly:

$$\Delta E_{opt} = -\Delta E_{kin} \quad . \tag{3}$$

The gain profile is shown in Fig. 5b. The gain is proportional to the derivative of the spontaneous spectrum (gain-spread-relation or Madey theorem, cf. below).

We will not pursue this model since the alternative description provides further insight into the FEL mechanism. However, we make the following statements about the scattering model:

- It correctly describes the FEL mechanism.
- It is the natural starting point for a quantum mechanical treatment.
- The final results are independent of Planck's constant h suggesting a classical model.

3.2 The second FEL model: classical description

Classical description of the FEL mechanism were made by different authors at about the time of the first experiment. Here we follow the description by W.B. Colson [7] which is not only well suited for didactic purposes but is also widely used today. We are dealing with a classical model, i.e. phase and amplitude of the optical field are simultaneously well defined. Similarly the electrons follow well defined trajectories in space and time. The frame of reference is the laboratory frame.

An external optical field with an electric field vector $\vec{E}(\vec{r},t)$ co-propagating with a single electron couples to this electron due to the transverse velocity $v_\perp$ of the electron in the wiggler field. The question to be asked is: What is the change of the particle's kinetic energy? Since the magnetic wiggler field cannot participate in the energy exchange all changes of electron kinetic energy ΔE_{kin} leads to a corresponding change of energy of the optical field ΔE_{opt} due to energy conservation:

$$\Delta E_{\text{kin}} = -\Delta E_{\text{opt}} \ . \tag{4}$$

For simplicity we assume a plane-wave optical field. The rate at which energy is exchanged is proportional to the product $E(r,t)\cdot v_{\perp}$. Both components oscillate rapidly in time, and in general the time integral over the interaction region will involve much cancellation with a correspondingly small net effect. However, when the electric field amplitude at the position of the electron remains in constant phase relationship to the transverse wiggling motion of the electron, the cancellation is diminished and the amount of energy ΔE_{kin} exchanged can be substantial (see Fig. 6):

$$\Delta E_{\text{kin}} = \int_0^T dt \ \ e\cdot \vec{E}(\vec{r},t)\cdot \vec{v}(\vec{r},t) \ . \tag{5}$$

This occurs on resonances when the electron slips behind the optical field by one optical wavelength λ_l, while the electron transverses one wiggler period of length λ_o

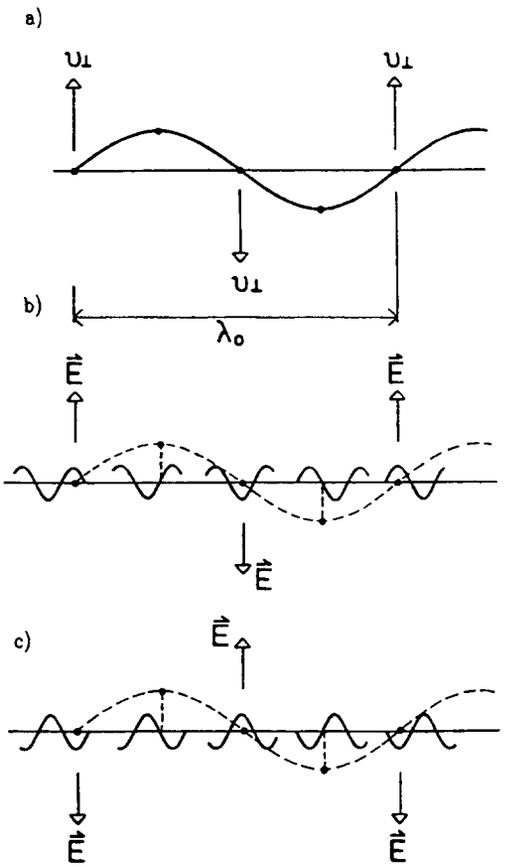


Fig. 6 FEL interaction by classical model, gain:
a) An electron is shown at five different times while it traverses a wiggler period of length λ_0.
b) The phase velocity of an external optical field is the speed of light and thus exceeds the electron velocity. On resonance the electric field at the location of the electron maintains a fixed relative phase. The interaction $\int dt\cdot E\cdot v_{\perp}$ is large. Initial phase is such that energy is transferred to the electron (inverse FEL 'linac').
c) Initial phase such that electron loses energy (FEL).

$$\frac{\lambda_1}{c - v_{\|}} = \frac{\lambda_0}{v_{\|}} \tag{6}$$

where $v_{\|}$ is the average longitudinal electron velocity within the wiggler. In the limit $\gamma >> 1$, Eqs. (6) and (1) are equivalent.

While meeting the resonance condition ensures a constant relative phase of $\bar{E}(\bar{r},t)$ and $\bar{v}_{\perp}(\bar{r},t)$ the actual value of the integral Eq. (5) depends on the initial phase, i.e. at which time the electron enters the interaction region. An electron injected half an optical period later (or earlier) experiences an energy change of equal magnitude, but of opposite sign. It is rather difficult to generate an electron beam bunched at the scale of an optical period. Therefore all possible initial phases are equally populated by electrons. Averaging over all particle gives a vanishing net energy change; on resonance the optical gain vanishes!

However, the fractional energy change of an individual electron can be rather large, and is of the order of $1/2N$ in the saturated regime (N is the number of periods in the wiggler). The FEL interaction increases the electron energy spread ('beam heating'). The longitudinal motion due to the energy modulation causes the electron to bunch with a period of an optical wavelength. On resonance the bunching occurs at a phase of the optical field where there is no net exchange of energy.

Let us spend some more time on this point: the transverse motion of the electron couples to the transverse optical field causing an energy change. This in turn gives rise to a longitudinal 'slow' motion superimposed on the rapid motion through the wiggler. It is this slow superimposed motion which contains 'all the FEL physics'. It can be described as if the electron sees a sinusoidal potential with the period of a wavelength copropagating along the axis at a speed corresponding to resonance energy. It is called the ponderomotive potential. This is only a means to describe the particle motion, there is no physical meaning associated with the ponderomotive potential except that it describes the longitudinal effects due to the transverse coupling. The motion in the potential is described by the well known pendulum equation. The motion is slow in the sense that particles execute less than one oscillation in the ponderomotive potential during one pass through the wiggler while performing hundreds of transverse oscillations. There are, however, situations where more than one oscillation occurs.

This pendulum equation is formally equivalent to a linac, and the ideas from linac theory may be transferred to FEL (and vice versa). Indeed, the longitudinal motion is referred to as synchrotron motion. Particles may be trapped in the ponderomotive potential, and in what follows we shall use the term 'optical trap' for the potential well with an extension of wavelength (in contrast to rf-buckets generated by the rf acceleration system of accelerators). Phase space trajectories are shown in Fig. 7. They resemble ordinary synchrotron oscillations except for the longitudinal scale being in the accelerator case of the order of tens of cm, and in the FEL case of the order of an optical wavelength.

When particles are injected slightly above resonance energy they move forward with respect to the trap and will push against the ponderomotive potential so losing energy. Bunching occurs at a phase where work can be done on the optical field, see Fig. 8. The existing optical field gains energy due to the interaction. This is the FEL mechanism. Another way to look at it is that the bunched beam containing an AC component radiates at this frequency. For completeness we quote here the maximum small-signal-gain formula for an ideal FEL as given by W.B. Colson [7]

$$g = 0.135 \frac{8N(\pi eKL)^2}{\gamma^3 mc^2} \rho(z) \tag{7}$$

where N = number of periods in wiggler
K = wiggler strength (formula requires $K \leq 1$)
L = wiggler length
$\rho(z)$ = electron density.

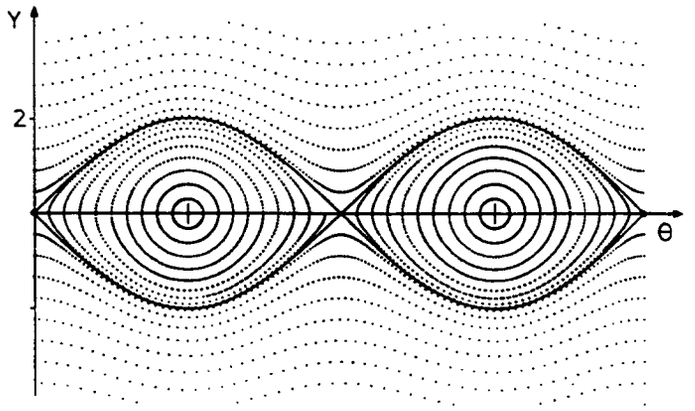


Fig. 7 Trajectories in longitudinal phase space. The horizontal co-ordinate is the phase difference θ between the electric field vector and the transverse motion of the electron. A range of 4π is shown corresponding to two optical wavelengths. The vertical co-ordinate Y is the normalized energy deviation from resonance which is the canonical conjugated variable to θ.

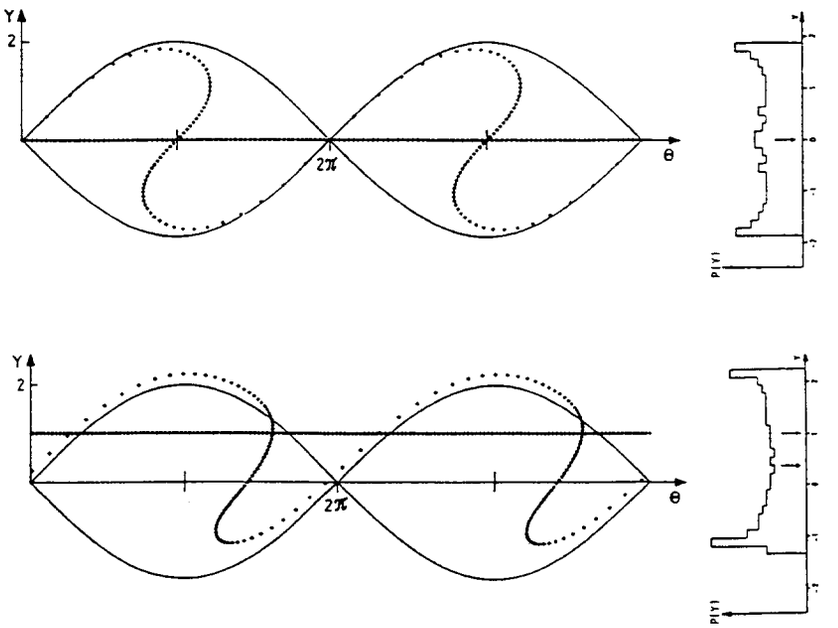


Fig. 8 Electron distributions in longitudinal phase space before and after the FEL interaction. a) Injection on resonance $Y = 0$. The final S-shape distribution bunches particles at $\theta = \pi$. b) Injection above resonance $Y = 1$. The final distribution bunches at $\theta > \pi$. The final energy distribution $P(Y)$ shown to the right indicates both beam heating and energy loss.

Let us assume radiation is stored in an optical cavity enclosing the FEL interaction region. That is the light bounces back and forth between two concave mirrors focused by their spherical surface. With the FEL interaction turned on, the radiation is amplified on each round trip. If the gain is bigger than the optical losses the field amplitude grows and the system is above the

lasing threshold. The intracavity power grows until the gain decreases due to some non-linearity. This is saturation.

In the classical model of the FEL, the saturation process is identified as 'overbunching', see Fig. 9. The longitudinal motion responsible for the build-up of the bunching occurs at a pace depending on power density. At higher power levels the bunching is achieved before the end of the interaction. The particles continue to move in the ponderomotive potential destroying the bunching, and the gain decreases.

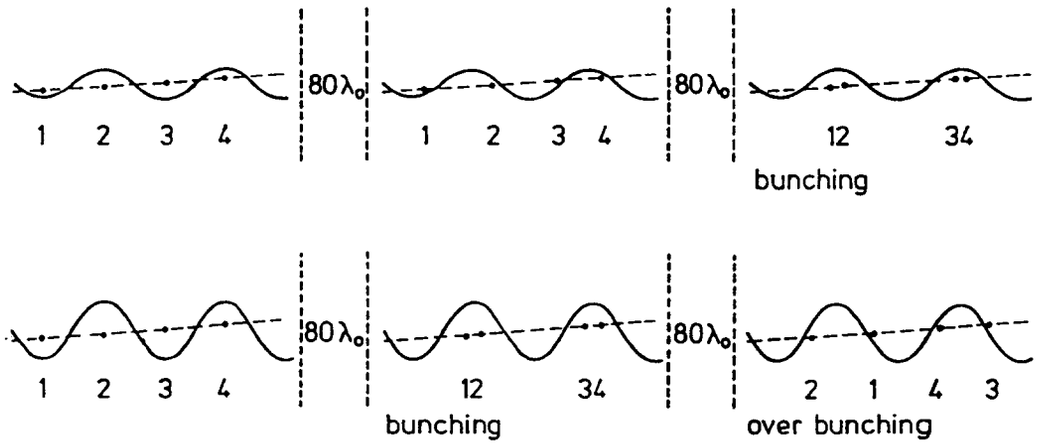


Fig. 9 a) At low optical power density maximum electron bunching occurs towards the end of the interaction (the first FEL had 160 periods). b) At high optical power density maximum bunching occurs early during the interaction. The electrons continue to move longitudinally so that bunching and hence gain is reduced (saturation). For clarity the electrons are numbered.

The maximum fraction of electron kinetic energy converted into radiation is $1/2N$, i.e. equal to the line width of the spontaneous spectrum of an ideal FEL. The kinetic energy has to be supplied by the accelerator, usually as rf energy. One can define an academic efficiency of an FEL as the fraction of kinetic energy converted into radiative power. Although often used, such a definition does not include the efficiency of rf generators and the power consumption of auxiliary equipment such as magnets, diagnostic equipment, pumps, etc. For ordinary FEL wiggler magnets the academic efficiency is limited by $1/2N$ and is typically below 1%.

Saturation as well as gain requires longitudinal motion of electrons. Usually high energy electrons are used and, due to relativity, relative motions are slow. This implies small gain in an FEL, but also slow saturation, i.e. the gain decreases only slowly as a function of power. An FEL by its very physics is well suited to high power performance! Also the radiation is generated in vacuum, and no matter can become non-linear (this statement does not apply to the mirrors of the optical cavity).

4. THE FIRST PRACTICAL FEL

With the precaution of the historical remarks, the first FEL was built by J. Madey and co-workers at Stanford University in 1976 [3]. The 43 MeV electron beam from a linac using superconducting accelerating structures reached a peak current of more than 1.5 A at an emittance of 6×10^{-8} m rad and a relative energy spread of 5×10^{-4}. The wiggler had $N = 160$ periods of length $\lambda_0 = 3.2$ cm. It was formed from a superconducting double helix generating a field amplitude of $B_0 = 2.4$ kGauss. The emitted wavelength was $\lambda_1 = 3.4\ \mu$. At a small-

signal gain of several percent the system saturated at an output power of 7 kW which is 10^8 times the spontaneous power. Good agreement was obtained between the experiment and theoretical predictions so that the performance of other experiments can be forecast with some confidence.

This agreement includes the gain-spread relation also known as Madey's theorem [8]. It relates the spectrum of the spontaneous emission, the gain, and the energy loss of the electrons. It holds whenever a classical description is adequate, and an average over initial phases is needed to describe the experiment, i.e. it is valid under a very broad range of situations. It allows one to calculate the gain from the more easily obtainable spontaneous spectrum. It also states that there is no gain without spread, i.e. the FEL interaction heats up the electron beam (an undesirable situation, e.g. in a storage ring).

The Madey theorem is given by the following two formulae:

$$\left\langle E_f - E_i \right\rangle = \frac{1}{2}\frac{\partial}{\partial E_i}\left\langle \left(E_f - E_i\right)^2 \right\rangle \tag{8}$$

and

$$\left\langle \left(E_f - E_i\right)^2 \right\rangle \sim \frac{d\Gamma}{d^3k} \tag{9}$$

where E_i and E_f are the initial and final electron energy, respectively. The right hand side of (9) is the spontaneous spectrum, and $\langle ... \rangle$ denotes the average over all beam particles.

Successful operation of a FEL also demonstrated its major drawback: an electron beam of high quality with respect to current, cross section, divergence, and energy spread is required. A large fraction of experimental FEL work is indeed devoted to the accelerator.

A given wiggler requires a certain electron beam quality to work properly. If the accelerator does not match the specifications the laser gain profile is inhomogeneously broadened beyond the homogeneously broadened line width of $1/2N$, with the consequence that not all electrons contribute to the lasing process. Since this is so important we shall discuss this point in more detail.

4.1 Inhomogeneous line broadening

Since gain is notoriously low and scales as N^3, a large number of periods must be used (typically $N \sim 100$). An ideal FEL has a gain profile shown schematically in Fig. 5 which has a fractional line width of the order of $1/2N$. This is the homogeneous line width. The profile can be inhomogeneously broadened by any of the following effects (the list is not complete):

Energy spread. Due to the peculiar gain profile an electron beam energy spread $\Delta E/E$ has to be small compared to $1/2N$.

Beam cross section. A periodic transverse magnetic field increases in strength along the field lines (taken to be the y axis) proportional to cosh $(2\pi y/\lambda_0)$. Particles off axis have thus a different resonance condition. This must be within the homogeneous line width in order to add to the gain at the centre wavelength. This effectively limits the useful electron beam cross section.

Beam divergence. As discussed above the γ^2 dependence of λ_1 can be understood to be due to a twofold (Lorentz) transformation between the lab frame and the pseudo rest

frame. Divergence introduces an angle θ between the direction of the electron beam and the optical axis, and the resonance condition becomes

$$\lambda_1 = \frac{\lambda_0}{2\gamma^2}\left[1+\frac{1}{2}K^2\right]\left(1+\gamma^2\theta^2\right) . \qquad (10)$$

There is thus a stringent limit on the divergence of the electron beam.

Wiggler field. For long wigglers (e.g. $N > 100$) the fluctuation of the magnetic field amplitudes becomes an issue requiring close attention in design and manufacturing of the periodic magnetic structure. Any deviation of the field from the ideal one causes a change of the particle's trajectory. As a consequence the (local) longitudinal velocity $v_{\|}$ deviates from the nominal value. As can be seen from Eq. (6) the particle may thus get out of phase with the optical field causing a reduction in performance.

These and other considerations (cost!) have prevented a wide spread use of FEL. However, there are a large number of applications whose needs can presently only be met by FEL, thus justifying further work. Different guide lines have been followed, some of which will be discussed below.

5. A LOW-TECHNOLOGY EXPERIMENT

The Mark III FEL [9] uses one section of the 20-year old Stanford linac (the same rf structures were used later at SLAC). A newly developed microwave electron gun injects into the linac providing a peak current of 20 A at a normalized emittance of 7 x 10^{-6} π m rad. The wiggler is built from permanent magnets as proposed by K. Halbach in 1979 [10]. Since this is an important aspect of FEL (and undulators and wigglers for synchrotron radiation) we shall devote an extra section to permanent magnets later. The laser wavelength can be tuned between 2.6 μ and 3.1 μ by varying the magnetic field amplitude. This system is used regularly not only to study FEL physics issues, but also in practical applications. At the time of the 1987 FEL conference more than 1000 h of user operation had been achieved. The system also emits at higher harmonics of the fundamental.

Among the FEL physics studied was the optical guiding: the FEL interaction not only drives the electric field amplitude, but also the phase. Both effects tend to focus the optical beam towards the e-beam axis. Experiments performed at the Mark III FEL verified theoretical predictions quantitatively for the first time. Optical guiding allows in principle for a wiggler length not limited by diffraction of the optical mode.

Another linac based FEL has operated in France [11]. The system named CLIO uses an electron beam between 32 MeV and 50 MeV to tune the fundamental between 2.5 μm and 15 μm. At a net gain of 140% the cavity length detuning curve is 40 μm wide and a peak power of 2.5 MW at a 0.4% efficiency is reported. Also, lasing on the 3rd harmonic was possible.

The user-oriented facility FELIX in the Netherlands [12] provides optical beams in the far infrared part of the spectrum. Using a 24 MeV electron beam a relative bandwidth of 10^{-6} was measured at 24.7 μm.

In the past electron sources have been developed using laser driven photoemission from cathodes placed inside a rf cavity. This allows for high electric field strength of the order of 1 MV/cm near the cathode causing rapid acceleration to relativistic energies. Thus emittance blow-up due to space-charge effects can be limited. Such sources feeding linacs deliver high-brilliance beams capable of driving single-pass FEL in the VUV.

5.1 Permanent magnets

A wiggler should have a fairly small period (several cm), a high field amplitude on axis (a few kGauss), and provide sufficient aperture (a few cm) for the e-beam and the optical mode. These objectives are essentially impossible to achieve using a room temperature electromagnet. The current density would be well above 1 kA/cm^2, which is about the limit imposed by cooling requirements. Consequently the first FEL wiggler was superconducting.

With the advent of permanent magnets based on rare-earth cobalt alloys (REC) the scene changed. The materials have both a high remanent field B_r and high coercive forces $_BH_C$. The best material to date (October 1987) has B_r = 12.5 kGauss and $_BH_C$ = 11.5 kOersted. Important in practice is the linear behaviour of the material in the $B(H)$ diagram, not only in the second, but also a significant way into the third quadrant. The permeability is close to unity. Thus the material can be thought of as surface current sheets with current densities of 1 kA/mm! Thus the abreviation CSEM = current sheet equivalent material.

The exceptional properties are due to the strong coupling of magnetic moments to the non-isotropic lattice. This coupling reaches an equivalent field of 500 kGauss. However, lattice imperfections tend to reduce the effective coupling and only much smaller coercive forces are achieved in practice.

The material is formed in several steps. First the components (e.g. $SmCo_5$, Sm_2Co_{17}, $Nd_2Fe_{14}B$) are mixed in the proper ratio and melted. After cooling, the alloy is ground to a particle size of a few microns. The grain size should correspond to the size of an individual magnetic domain, i.e. all magnetic moments within a grain should be aligned along the same axis. Then the powder is placed in a magnetic field and treated (e.g. by ultra-sonic means) such that grains align along the field. This process defines the easy axis of the finished blocks. The powder is pressed and sintered, the temperature being controlled so as not to distort the grain alignment. Before the final magnetization along the easy axis the blocks may be shaped, cut, etc. The material, in particular the Sm alloy, is hard and brittle. Dust exposed to air may spontaneously oxidise (explosive!).

The alloy $Nd_2Fe_{14}B$ has only recently been discovered (partially by chance!). Compared to $SmCo_5$ it provides for up to 30% higher B_r. However, the Curie temperature is only around 320 °C as compared to ~ 750 °C for $SmCo_5$. Thus the long term stability, temperature sensitivity, and radiation resistance is worse for NdFe than for SmCo. The prices are presently similar at around 2 US$/cm^3.

K. Halbach at Berkeley was the first one to recognize the potential of wigglers built from permanent magnets. Among the obvious advantages are the facts that no power or cooling is necessary. However, the material is difficult to handle (and can loose its magnetization). A variable on-axis field amplitude requires a mechanical motion of the magnets under heavy load (up to several tons!) and with low mechanical tolerance (~ 20 μ).

Halbach proposed two different schemes for constructing a wiggler structure from permanent magnets. The first one, called 'pure REC design', generates the field using only permanent magnets. In the so-called 'hybrid design' pole-pieces made from steel or vanadium Permendur are used to channel field lines to the beam axis. In either case the scatter in properties between individual magnet blocks as delivered must be taken into account when assembling the structure in order to avoid inhomogeneous line broadening and/or e-beam steering due to field errors. A hybrid is usually more forgiving with respect to magnetisation errors. At large gap-to-period ratio, i.e. for short periods, the REC design leads to higher on-axis fields since the field generating material is closer to the axis on average (cf. Fig. 10).

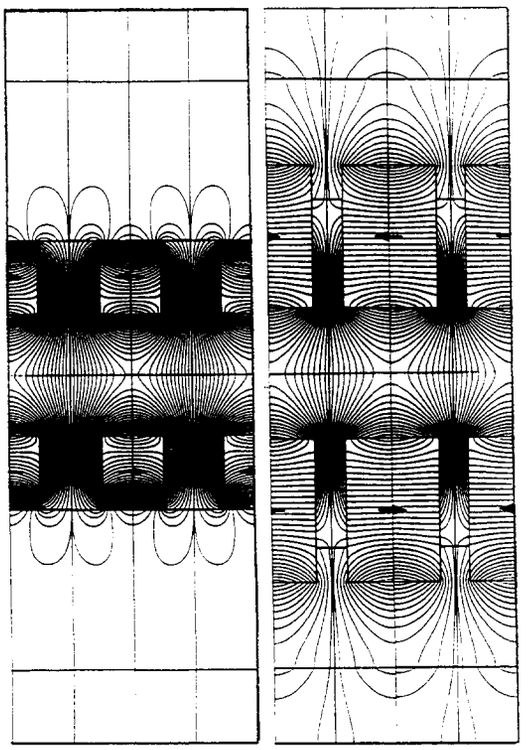


Fig. 10 Calculation of magnetic field lines in a pure REC structure (left side) and in a hybrid (right side). The directions of magnetisation of individual blocks are indicated by arrows. The hybrid contains pole-pieces made of steel or vanadium Permendur. The on-axis field amplitude is the same in both calculations; it can be seen that a hybrid requires more magnet material.

6. AN ELECTROSTATIC FEL DRIVER

The system built by L. Elias and co-workers at the University of California, Santa Barbara (UCSB) is based on an electrostatic 3 MV van der Graaf accelerator [13] (Fig. 11). The electrons once they have passed through the FEL are decelerated in another column and collected at the high voltage terminal of the accelerator. Thus a large current of up to 1.2 A can be driven through the FEL at a high voltage power supply rating of only 100 μA. A voltage drop of 0.6 kV/μs at 90% beam recovery has to be accepted, causing a corresponding change in laser wavelength. It is shown, however, that at any given time only a single longitudinal mode with a fractional line width of 10^{-8} is excited. The laser operates in the far infra-red region between 100 and 400 μ wavelength. This region is presently not well covered by more conventional lasers although a large number of important excitations and level splittings have energies in this range. These include, but are not limited to, vibrational modes of large molecules, band gaps in superconductors, phonons in solids, etc.

The UCSB laser operates regularly for far infra-red spectroscopy, and first results have been published. Based on operational experience it is proposed for future systems of this kind to place the FEL inside the high voltage terminal while the more critical components, namely electron gun and collector are at or near ground potential [14] (Fig. 12).

7. STORAGE RING FEL

The fraction of electron kinetic energy converted into radiation is small. It is tempting to recycle the electrons, i.e. to install the FEL into a storage ring. In doing so the repetitive FEL

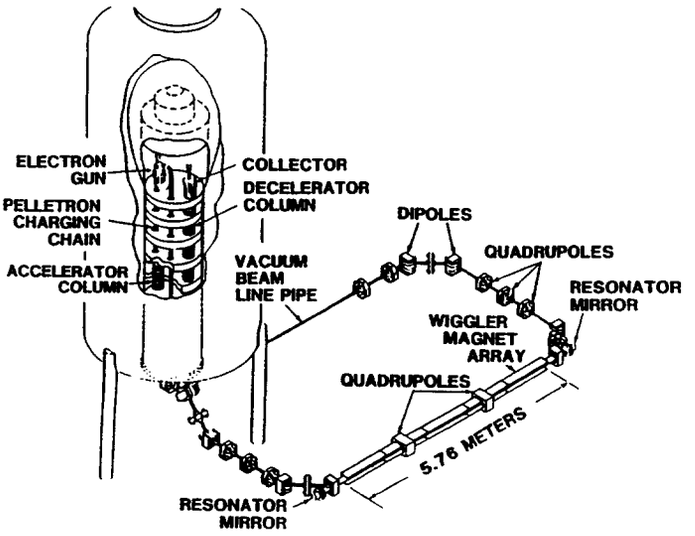


Fig. 11 The far infra-red free electron laser at the University of California, Santa Barbara. The electrons are recovered after the FEL interaction and decelerated in a second column of the electrostatic accelerator.

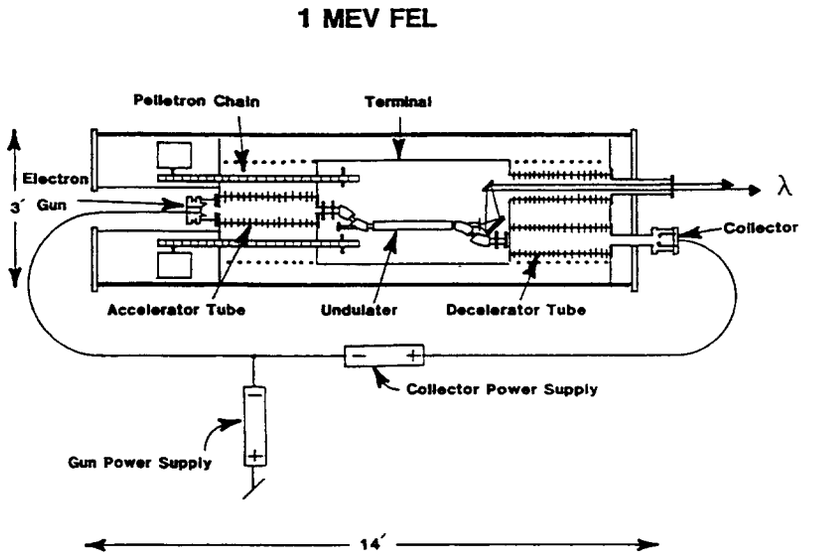


Fig. 12 Proposed compact far-infra-red FEL with the wiggler at high potential and electron gun and collector near ground potential

interaction increases the energy spread of the beam until the spread becomes comparable to the homogeneous laser line width, i.e. $1/2N$ (assuming an energy acceptance of the ring larger than $1/2N$). Then the gain decreases and the system saturates. Note that the saturation mechanism differs from the normal single-pass devices. The system is stabilized by the damping due to synchrotron radiation, and it can be shown that the laser power P_{FEL} is related to the synchrotron radiation power P_{syn} emitted in the ring by

$$P_{\mathrm{FEL}} \leq \frac{1}{2N} P_{\mathrm{syn}} \ . \tag{11}$$

Since the energy loss by synchrotron radiation is balanced by an accelerating rf cavity, the efficiency of a storage ring FEL is the same as that of a single pass FEL. This is known as the Renieri limit [15]. However, storage rings provide bright beams at high energies (~ 1 GeV) making them attractive as short-wavelength FEL drivers. Analysis indicates that dedicated rings are required for laser wavelengths below 1000 Å (corresponding to 10 eV photon energy). This is dictated by the lack of suitable mirrors. At Duke University, NC, and at Dortmund, Germany, dedicated 1 GeV low-emittance rings are under construction [16].

The lack of short wavelength mirrors has triggered the idea to amplify the spontaneous emission on a single pass up to saturation, i.e. to operate the FEL in a super-radiant high gain mode [17]. This typically requires wigglers with 1000 periods. Due to the scaling of the FEL gain low electron energies (< 1 GeV) and high wiggler strength ($K \geq 2$) are required. The cosh-dependence of field strength on wiggler gap reduces the transverse acceptance of the FEL section for electrons to a value below the feasible limit for storage ring operation. Thus the FEL is located in a by-pass into which the e beam is kicked at a low rate (~ 10 Hz). Design studies have been made [18] predicting MW coherent power levels at tens of eV photon energy.

However, the first storage ring FEL at ACO in Orsay, France [19] has much lower performance levels. It operates throughout the visible part of the spectrum at an electron energy as low as 160 MeV (injection energy of ACO is 250 MeV). The major difficulty of the experiment is the low gain of below 10^{-3} requiring very-low-loss mirrors for the optical cavity.

To arrive at the quoted gain, optimum use of the available space of 1.3 m had to be made by converting the wiggler into an optical klystron [20]. As discussed above, the FEL interaction first generates an energy modulation which leads, via the associated velocity modulation, to bunching by 'waiting'; it is the bunched beam that radiates. The optical klystron leads to a faster bunching by introducing a dispersive insert in the middle of the undulator providing negative momentum compaction. This improves the small signal gain at the expense of a lower saturation power level (see Fig. 13). The ACO storage ring FEL has been used for a number of investigations. In particular, the heating of the stored beam due to the FEL was studied, as well as the mirror degradation due to high energy photons [21].

The VEPP-3 storage ring at Novosibirsk, Siberia, provides for over 8 m free space in a bypass to the normal straight section. An electromagnetic (!) optical klystron with a period of 20 cm allows FEL operation in the UV below 350 nm [22].

Other storage-ring FELs have operated as well. The NIJI-IV ring operating at 240 MeV supported an FEL at 590 nm [23]. The higher electron energy of 500 MeV made the UVSOR FEL tunable between 430 nm and 480 nm [24]. At SUPER-ACO in France dynamical instabilities near saturation are studied. This FEL tunes through the major part of the visible spectrum well into the UV [25].

In order to avoid the bad effects caused by FEL beam heating it is proposed to install the wiggler in a section of the ring with dispersion $\eta(s)$ and to provide a transverse gradient $B_0 = B_0(x)$ to the wiggler field such that the resonance condition is met for all electrons irrespective of the energy [26] (in linear approximation)

$$\begin{aligned} d\lambda_1 &= \frac{\partial \lambda_1}{\partial \gamma} d\gamma + \frac{\partial \lambda_1}{\partial B_0} dB_0 \\ &= \frac{\partial \lambda_1}{\partial \gamma} \frac{\partial \gamma}{\partial x} dx + \frac{\partial \lambda_1}{\partial B_0} \frac{\partial B_0}{\partial x} dx = 0 \ . \end{aligned} \tag{12}$$

By this trick the gain curve as a function of electron energy becomes wider, hence the name gain expander.

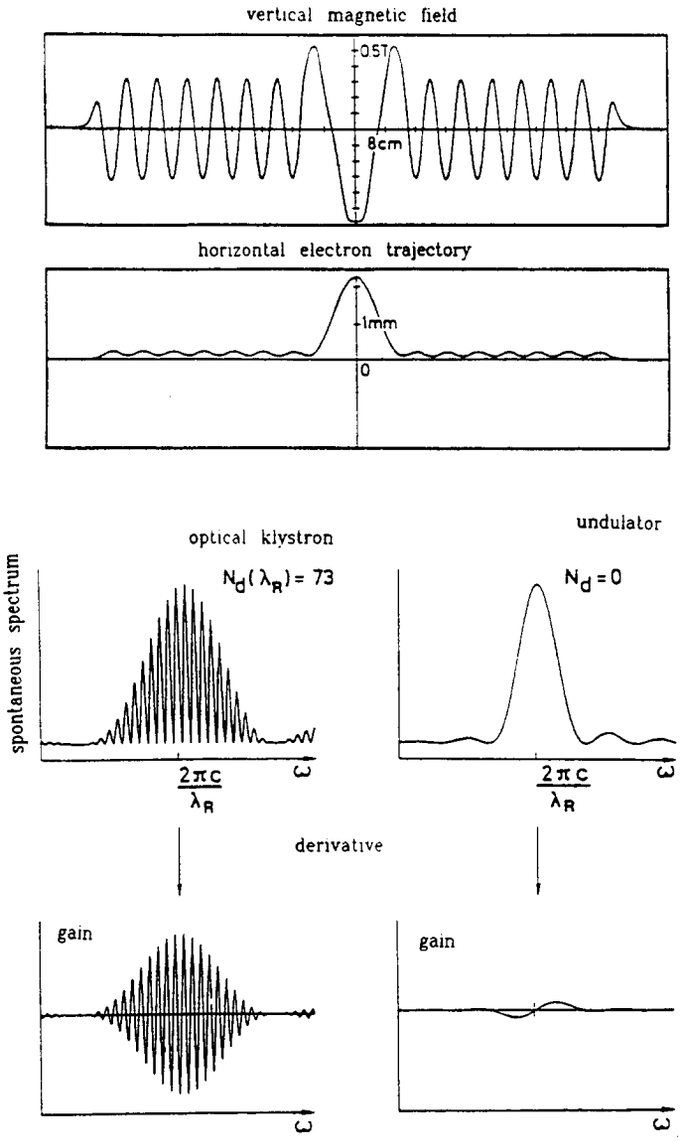


Fig. 13 a) The optical klystron is an ordinary wiggler with a dispersive insert, which converts the energy modulation into an intensity modulation. b) The spontaneous spectrum is a coherent superposition of the emission in the two sections and exhibits interference structures. c) The gain being proportional to the derivative of the spontaneous spectrum is strongly enhanced. The optical klystron is more sensitive to inhomogeneous line broadening caused by a finite energy distribution width.

It turns out that a generalized gain spread-relation holds relating gain and transverse excitation. The latter spoils the correlation of energy and transverse position. A benefit, however, is that saturation power depends on the energy acceptance of the ring and not on the number of wiggler periods leading to higher efficiency [27].

Another proposal is the isochronous operation of the storage ring requiring the spread of round-trip times of different electrons to be small compared to an optical period [28]. Then the initial FEL phase is correlated from round-trip to round-trip which corresponds to an 'infinitely' long wiggler interspaced by the guide field of the ring. Particles can be trapped in the

ponderomotive potential, and the distribution is stabilized by radiation damping. Extremely high efficiencies (approaching 100% for large systems and high power levels) are predicted.

Isochrony requires not only a small coupling between horizontal transverse and longitudinal motion (see Fig. 14), but also operation close to transition energy, where momentum compaction and phase focusing vanishes, i.e. close to an inherently instable situation.

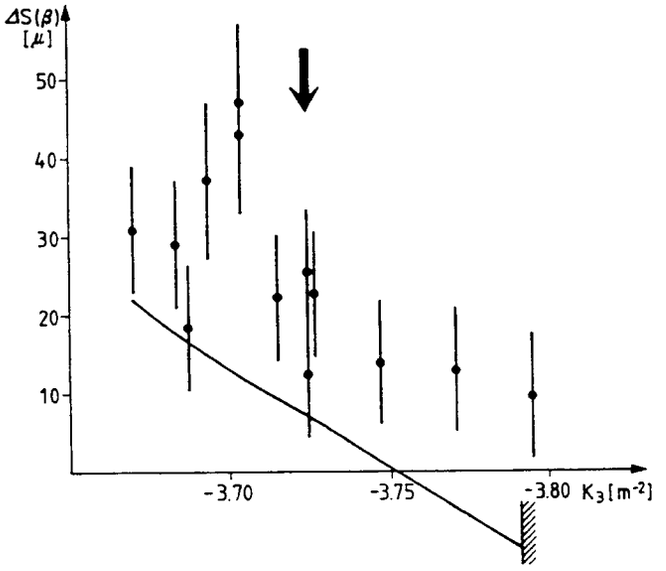


Fig. 14 The longitudinal transverse coupling in an electron storage ring leads to an amplitude dependent longitudinal displacement per turn with mean = 0 and variance $\Delta s(\beta)$. A change of the focusing strength K_3 of the third quadrupole family of the BESSY I storage ring leads to the data as shown. The solid line is the prediction from the computer model of the ring. The arrow indicates the normal working point. The measurement is performed by exciting the beam at the tune and by comparing AM and FM of the signal of an electrostatic pick-up.

8. ENERGY RECOVERY

In the search for high optical power (many MW to TW) efficiency becomes more and more an issue. Since recycling the electron in a storage ring does not improve efficiency, studies have been made on recycling the electrons' energy. In a linac-based single-pass system the spent electrons can be turned around and decelerated in some rf structure. A straightforward solution is to use the same accelerating structure by injecting the used electrons 180° out of phase. However, instabilities may arise due to the simultaneous presence of electron bunches with different energy. The Los Alamos group therefore adapted a separate deceleration rf structure and coupled the rf power to the accelerator (see Fig. 15).

9. TAPERED WIGGLER

Instead of and/or in addition to recycling the electrons or the energy, better use of the electron energy can be made by tapering the wiggler. Usually this means a reduction of wiggler field amplitude along the length of the wiggler such that particles stay in resonance although they constantly lose energy to the optical field. In practice only some fraction of all electrons participate in that process.

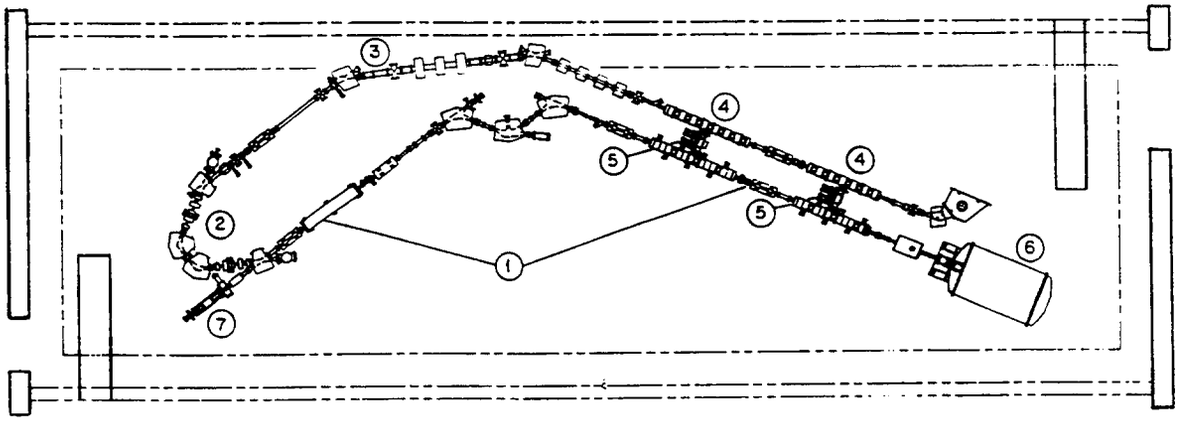


Fig. 15 Layout of the 10 μ Los Alamos FEL with energy recovery. 1) Original Linac-FEL with achromatic bend, 2) and 3) magnets to bend and guide the hot electron beam, 4) decelerating structures, 5) bridge couplers, 6) electron gun and injector, 7) optical diagnostics.

The Lawrence Livermore Laboratory [30] operates a single pass FEL amplifier for mm wavelengths. An electron beam of 3.5 MeV and 1000 A from an induction linac is reported to deliver up to 34% of its power to the radiation field using a 40% taper of the magnetic field amplitude. Such a large extraction requires a high optical field amplitude so that the height of the ponderomotive potential exceeds the energy spread in the beam. Then, not only almost all electrons are trapped, but also they remain trapped while the potential slows down due to the field taper.

Tapered wigglers provide larger energy extraction at the expense of a reduction in small-signal gain.

10. CERENKOV LASER

We mention this laser here although the gain mechanism is rather different from an ordinary FEL. An electron beam is passed parallel and close to the surface of a dielectric medium with a high refractive index [31] (e.g. the plastic TPX has an index $n = 2.12$ at far infra-red wavelength). Already for low energy electrons ($2 < \gamma < 3$) their velocity exceeds the speed of light in the medium so generating Cerenkov radiation. The particle-wave interaction leads to bunching of the electron beam thus enhancing emission. It is demonstrated that a very compact high power (10 ... 200 kW) far infra-red (400 μm to 950 μm) continuously tuneable source can be realized.

11. MOTIVATION FOR FEL

The FEL is a unique source of electromagnetic radiation. The ability to operate at any wavelength (only constrained by technical limitations) and to generate power levels well above 1 MW makes it an obvious tool for a number of applications. However, their high cost (depending on wavelength, but a minimum of the order of 10 million US $) requires some justification.

The FEL has been considered as a directed energy weapon by the Strategic Defense Initiative. Consequently a number of the FEL projects in the US rely on military funding. It is beyond the scope of this contribution to discuss the feasibility of FEL for this purpose. FEL's have been, and will be, used for all kinds of spectroscopic purposes. These applications presently are

concentrating on the infra-red around 3 μ (Mark III; Rocketdyn, Stanford; TRW, Stanford), and in the far infra-red around 400 μ (Santa Barbara). These applications will certainly continue and expand. Projects to meet the anticipated demand have already been started.

The choice of wavelength is dictated by technical issues as well as by the existence of alternative sources. A system operating in the gap between 10 μ and 100 μ can be as successful as the near infra-red FEL mentioned above. User oriented facilities in the Netherlands (FELIX) and in France (CLIO) concentrate on this spectral range.

The prospect to extend FEL operation to shorter wavelengths, into the UV ($\lambda \geq 200$ nm) and even into the VUV, is more spectacular. The applications are numerous (for example extending high resolution laser spectroscopy to non-valence electron shells) and are usually not questioned. The major obstacle in view of the gain scaling is the lack of suitable mirrors to form optical cavities. Using multilayer coatings on metal substrates 17% reflectivity at normal incidence and a photon energy of 100 eV (wavelength ~ 10 nm) has already (October 1987) been measured. More recently 50% reflectivity at normal incidence near 4.5 nm has been reported. Further progress is to be expected, opening exciting research opportunities.

CONCLUSION

The FEL mechanism can be understood by using classical electrodynamics (including special relativity) without the need to consider quantum mechanics. The promise of high power (scalable in principle to TW) at any wavelength has initiated a number of projects, some but not all of which are mentioned in this chapter. The more recent projects aim at specific applications, and use modifications of a simple FEL set-up. The goals are mainly high power (e.g. for military applications), and tuneable radiation in regions of the spectrum inaccessible to conventional lasers. Applications are predominantly of spectroscopic nature, but material processing (ceramics, biological tissue) is also considered. The author's belief is that these latter motivations will drive a continuous FEL programme throughout the world even though the militarily motivated projects may not be continued for technical and/or political reasons.

REFERENCES

[1] Nucl. Instrum. Methods A331, Proceedings of the 1992 FEL conference and references therein.

[2] G. Dattoli, A. Renieri and A. Torre, Proceedings of the 1985 CERN Accelerator School on Advanced Accelerator Physics, CERN 87-03, p. 792, see also Ch. Braun, Laser Focus, May 1981, and Ph. Sprangle, T. Coffey, Physics Today, March 1984.

[3] R.M. Phillips, US patent 3,129,356 granted April 14, 1964.

[4] J.M.J. Madey, J. Appl. Phys, 42 (1971), 1906.

[5] J.C. Gallardo et al., Nucl. Instrum. Methods A259, 245 (1987) and references therein.

[6] D.A.G. Deacon et al., PRL 38 (1977) 892.

[7] W.B. Colson and J. Blau, Nucl. Instrum. Methods Phys. Res. A259 (1987) 198 and references therein.

[8] W. Becker and J.K. McIver, to be published in Appl. Phys.

[9] S.V. Benson, J.M.J. Madey, J. Schultz, M. Marc, W. Wadensweiler and G.A. Westenskow, Nucl. Instrum. Methods Phys. Res. A250 (1986) 39.

[10] K. Halbach, J. Physique 44 (1983) C1-211. See also Nucl. Instrum. Methods A250 (1986) 115 and references therein.

[11] R. Prazeres, J.M. Berset, F. Glotin, D. Jarszynski and J.M. Ortega, Nucl. Instrum. Methods A331, 15 (1993).

[12] D. Oepts et al., Nucl. Instrum. Methods A331, 42 (1993) and references therein.

[13] A. Amir, L.R. Elias, D.J. Gegoire, J. Kotthaus, G.J. Ramian and A. Stern, Nucl. Instrum. Methods Phys. Res. A250 (1986) 35, and references therein.

[14] G. Ramian, L. Elias and I. Kind, Nucl. Instrum. Methods Phys. Res. A250 (1986) 125.

[15] A. Renieri, Frascati Report No. 77-33 (1977) unpublished.

[16] J H. Wiedemann, Proc. Int. Free Electron Laser Conf. 1982, J. de Physique 44 (1983 C1-201; see also J.E. LaSala, D.A.G. Deacon and J.M.J. Madey, Nucl. Instrum. Methods Phys. Res. 250 (1986) 262.

[17] See e.g. K.J. Kim, Phys. Rev. Lett. 57 (1986) 1871.

[18] J M. Cornacchia et al., Nucl. Instrum. Methods Phys. Res. A250 (1986) 57; see also: 1-3 GeV Synchrotron Radiation Source (formerly called ALS) Conceptual Design Report June 1986, Lawrence Berkeley Laboratory Pub. 5172. BESSY II, Eine optimierte Undulator/Wiggler-Speicherring-Lichtquell fur den VUV- und XUV-Spektralbereich Berliner Elektronenspeicherring-Gesellschaft fur Synchrotronstrahlung mbH, Berlin, November 1986.

[19] J M. Billardon et al., Phys. Rev. Lett. 51 (1983) 1652. P. Elleaume et al., J. Phys. 45 (1984) 989.

[20] N.A. Vinokurov and An.N. Skrinsky, Preprint INP 77-59, Novosibirsk 1977.

[21] M. Billardon et al., Nucl. Instrum. Methods Phys. Res. A259 (1987) 72, and references therein.

[22] G.N. Kulipanov et al., Nucl. Instrum. Methods A331, 98 (1993) and references therein.

[23] T. Yamazaki et al., Nucl. Instrum. Methods A331, 27 (1993).

[24] S. Takano, H. Hama and G. Isoyama, Nucl. Instrum. Methods A331, 20 (1993).

[25] M.E. Couprie et al., Nucl. Instrum. Methods A331, 37 (1993).

[26] T.I. Smith, J. Appl. Phys. 50 (1979) 4580.

[27] N.M. Kroll, Physics of Quantum Electronics, Vol. 8 edited by S.F. Jacobs et al., Addison-Wesley-Publ. 1982.

[28] D.A.G. Deacon and J.M.J. Madey, Phys. Rev. Lett. 44 (1980) 449, D.A.G. Deacon, Physics Reports 76 (1981) 349.

[29] D.W. Feldmann et al., Nucl. Instrum. Methods A259 (1987) 26.

[30] T.J. Orzechowski et al., Phys. Rev. Lett. 54 (1985) 889. T.J. Orzechowski et al., Nucl. Instrum. Methods Phyls. Res. A250 (1986) 144. T.J. Orzechowski et al., Phys. Rev. Lett. 57 (1986) 2172.

[31] E.P. Garate et al., Nucl. Instrum. Methods A259 (1987) 125.

ION TRAPPING AND CLEARING

A. Poncet
CERN, Geneva, Switzerland

Abstract
Ion trapping in the electron beams of future factories is becoming increasingly important. After a brief review of the adverse effects of neutralization in particle storage rings, the basic topics of ion production by ionization of the residual gas are recalled: ion production rate, natural clearing rates, ion kinematics, and conditions of trapping for bunched and unbunched particle beams with positive or negative space charge. Different methods of clearing are described and their performance discussed, namely, d.c. clearing electrodes, empty buckets (in electron storage rings) and beam shaking. Examples of neutralization effects and diagnostics are taken from CERN machines.

1. INTRODUCTION

In accelerators and storage rings, ions created by the circulating particles from neutral molecules of the residual gas may be trapped in the beam space-charge potential, and may generate all sorts of ill effects: reduced beam lifetime (increased pressure), emittance growth and losses through excitation of resonances, and coherent beam instabilities. Whilst they can occur in proton beams (e.g. CERN ISR trapping electrons), these neutralization phenomena mainly affect machines with negative beam space charge, such as electron storage rings, and antiproton accumulators.

Low-energy machines are more subject to ion trapping because of their small size, which leaves little space between bunches for ions to escape the beam potential, and suffer most because of their inherent high sensitivity to space charge effects. To illustrate this point, the incoherent space charge tune shift can be written as [1]:

$$\Delta Q = \frac{N}{k}\left(\eta - \frac{1}{\gamma^2}\right) \tag{1}$$

(where k relates to beam transverse dimensions and the bunching factor, γ is the relativistic factor, η the neutralization coefficient defined as the ratio of trapped charges to beam charges, and N is the number of particles in the beam).

ΔQ can be unacceptably large if γ is small (low energy), and/or η is high.

For instance: ΔQ is ~ 0·1 η in the CERN 600 MeV EPA (Electron Positron Accumulator [2]) for its nominal electron beam, where in the absence of clearing η can reach values close to one. This has to be compared with the value $\Delta Q \cong 0.1/\gamma^2$, i.e. ~ 10^{-7} for perfect clearing. A large tune shift is accompanied by a large tune spread, owing to the non-linearity of the ion focusing forces on the beam particles. This results in the excitation of a large number of resonances, as can usually be seen in the tune diagram. Figure 1 below illustrates this effect as seen from the beam intensity and stationary emittances in the EPA.

Ion trapping is one critical issue for the high-intensity electron beam of the particle factories proposals and projects [3]. The very high luminosity aimed at in e^+e^- collisions in these future machines is limited by a maximum allowable beam–beam tune shift of currently 0.06.

The effect of ions on the e^- beam is similar to the effect of one beam upon the other during collisions (beam-beam interaction). One must thus place an upper limit on the ion density in the e^- beam by demanding that the ion-induced tune shift be substantially less than the maximum beam–beam tune shift, for instance 0.01. This number puts a very severe upper limit of 10^{-4} or less on acceptable neutralization levels in these machines, difficult to achieve with present clearing means [4].

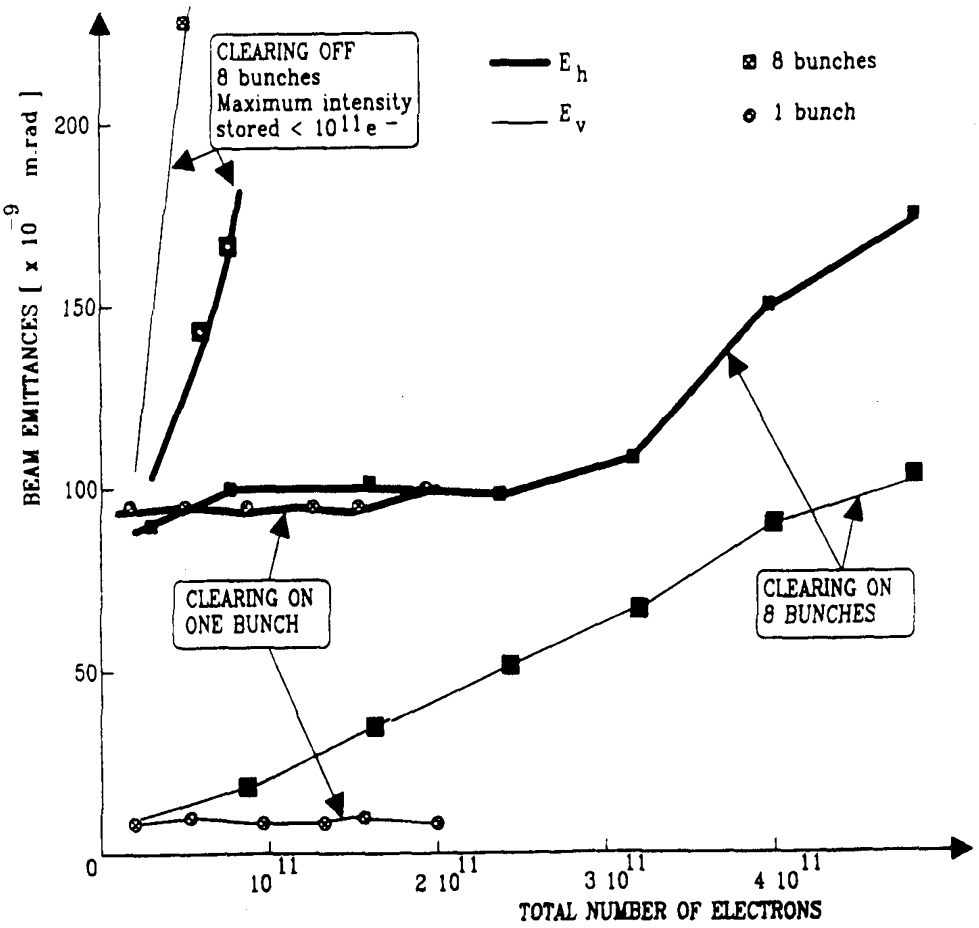


Fig. 1 EPA emittances with and without clearing

2. ION PRODUCTION

Beam particles with relativistic velocities interact with nuclei and the electrons of the gas atoms via the Coulomb forces. The energy transfer $\Delta E(b)$ in a collision, which depends on the impact parameter b (distance between the target and the projectile), may be sufficiently large to break the liaison between the nucleus and the electrons, leading to ionization:

$$\Delta E(b) > Z' \times 13.5 \text{ eV (ionization potential } E_1) \qquad (2)$$

In S.I. units [5]: $$\Delta E(b) = \frac{e^4}{8\pi^2 \varepsilon_0^2 \beta^2} \frac{Z^z Z'^2}{m_o} \frac{1}{b^2} \text{ (J)} \quad (3)$$

where Z and Z' are the atomic numbers of the projectile and the target, respectively, m_0 the mass of the target, e the elementary charge, c the speed of light, and ε_0 the vacuum permittivity. If the target is an electron: $m_0 = m_e$, $Z' = 1$; for a nucleus $m_0 \sim 2m_p Z'$.

One sees from (3) that at identical impact parameter:

$$\Delta E(b)_{electron} = \frac{2m_p}{m_e} \frac{1}{Z'} \Delta E(b)_{nucleus} >> \Delta E(b)_{nucleus} \left(\frac{m_p}{m_e} \sim 1836 \right) \quad (4)$$

This shows that collisions with orbital electrons are the main cause of energy loss, and therefore ionization rate calculations need only to consider electrons. Usually expression (3) is integrated over a range of possible impact parameters:

$$\frac{dE}{dt} = \int_{b_{\min}}^{b_{\max}} \Delta E(b) \left[2\pi b \beta c N_a Z' db \right] \quad (5)$$

where the quantity between brackets represents the number of electrons at distance b to the projectile during time dt, and N_a is the atom's density of charge Z'.

By considering the collision time with the orbital period of electrons, $b_{\max}$ is obtained from [5]:

$$b_{\max} = \frac{\gamma \beta c h}{E_1} \quad (6)$$

where h is the Plank constant = 4.14×10^{-15} eV·s, and $E_1 = Z' \times 3.5$ eV. The minimum impact parameter $b_{\min}$ (giving the maximum energy transfer for 'trappable' particles, namely the beam space charge potential U) is obtained from expression (3):

$$b_{\min} = \frac{r_0 c Z}{\beta} \left(\frac{2m_0 Z'}{U} \right)^{1/2} \quad (7)$$

since species created with $\Delta E(b) > U$ can escape to the vacuum chamber wall, and are therefore not trappable.

Table 1

Values of parameters r_0, m_0, and Z' for different beam trapping events. r_p and r_e are the classical proton and electron radii $\left(r_{p,e} = e^2 / 4\pi \varepsilon_0 m_{p,e} c^2 \right)$

Trappable particle	Electron (positive ion beam)	Proton (e^- or antiproton beam)	Ion (e^- or antiproton beam)
r_0	r_e	r_p	r_p

m_0	m_e	m_p	$2m_p$
Z'	1	1	Z'

The production rate is therefore:

$$R_p = \frac{1}{E_0}\ \frac{dE}{dt} = \frac{2\pi\, m_e\, c^3\, r_e^2}{\beta}\ \frac{N_a\, Z'}{E_0}\ \ell\mathrm{n}\left(\frac{\gamma\beta^2\, h}{E_1\, r_0\, Z}\ \frac{U}{2m_0\, Z'}\right)^{1/2} \tag{8}$$

E_0 being the average energy for the formation of an ion-electron pair (~35 eV).

A consequence of this expression is that – everything being equal (energy, beam potential) – a positive space-charge beam (e.g. protons) will trap less electrons than a negative one will trap positive ions, since the fraction of electrons produced in the ionization process with sufficient energy to escape the beam potential is larger.

This consideration, together with expression (4), illustrates the fact that positive ions are created with much less energy than electrons (in fact with quasi thermal energies of < 0.04 eV), and are therefore generally all trappable. Indeed a quick numerical application of expression (3) would show that the impact parameter has to be 4.7×10^{-14} m, i.e. a quasi head-on collision with the nucleus, and therefore highly improbable – for the H_2^+ ion to be produced with 10 eV energy (the electron would obtain 20 keV).

The neutralization coefficient of a beam is the ratio of the ion production rate R_p to the clearing rate R_c [s^{-1}]. Since the ions are virtually produced at rest, the production rate of ions can therefore be obtained from experimentally-determined ionization cross sections σ_i:

$$R_p = \sum_{i=1}^{n} \sigma_i\, N_{mi}\, \beta c \tag{9}$$

N_{m_i} being the molecular density of gas species *i*.

As a typical example, the values for the EPA (600 MeV, γ = 1200 for electrons) at $P = 10^{-9}$ mbar [6] are shown in Table 2:

Table 2
EPA ionization cross sections

Gas	σ_i (m^2)	R_p (s^{-1})	Ionisation time (s)
H_2	0.4×10^{-22}	0.4	2.5
CO	1.54×10^{-22}	1.5	0.7

Owing to a logarithmic dependence on the energy of the primary particle, these cross sections would only grow by 20% at 6 GeV/*c*.

2.1 Beam heating

Distant collisions with a large impact parameter – much more probable than close ones leading to ionization – are important, since they feed energy differentially to ions. In some

circumstances (neutralization pockets) this may be a clearing mechanism, i.e. when the trapped species get enough energy to escape the beam potential:

$$R_c = \frac{1}{eU}\ \frac{dE}{dt} = \frac{1}{eU}\ \int_{b_{\min}}^{b_{\max}}\ \Delta E\ (b)\ \left(2\pi\, b\beta\, cN_p db\right) \tag{10}$$

This represents the 'natural' clearing rate for a singly-charged species. The expression between brackets is the number of projectiles passing at distance b to the ion target during time dt. N_p is the projectile density of charge $Z = 1$.

In a good approximation, $b_{\max}$ and $b_{\min}$ can be chosen to have the same values as the ion and nucleus radii respectively, leading to [5]:

$$R_c = \frac{2\pi\, m_0\, c^3 r_0^2}{\beta}\ \ \frac{N_P\, Z'}{eU}\ \ell\mathrm{n}\left(3\times 10^4 \times Z'^{-2/3}\right) \tag{11}$$

with m_0 and r_0 being m_e, r_e, if the trapped species is an electron; m_p, r_p for a proton; and 2 m_p, r_p for an ion of charge Z'.

As an example, typical clearing times for the EPA machine with 6×10^{11} electrons (300 mA), and 1 mm beam radius, giving a beam potential of ~50 V, are shown in Table 3:

Table 3
EPA natural ion clearing rates

Gas	Clearing rate R_c (s^{-1})	Clearing time τ_c (s)
H^+	3×10^{-3}	350
H_2^+	6×10^{-3}	166
CO^+	0.04	25
CO_2^+	0.07	15

The process is thus slow compared with typical ionization rates, but may be important to explain why, in some circumstances (pockets, very low gas pressure: 5×10^{-11} Torr [7] fully ionized light ions can chase heavy ones, and accumulate up to a dangerous level.

2.2 Gas cooling

Seldom taken into account, gas cooling could perhaps be an important process for high pressures and long ion sojourn times. In addition, charge-exchange phenomena by which a positive ion captures an electron from a gas molecule may occur at ion energies of only a few eV. The new ion is created with the primary molecule's energy, while the newly created neutral species carries away the initial ion energy. Resonant capture cross sections between an ion and its own neutral molecule can be very high at low energy:

$$\sigma = 1.2\ \times\ 10^{-15}\ \mathrm{cm}^2\ \ \text{for}\ \ \mathrm{He}^+\ \ \text{in}\ \ \mathrm{He} \tag{12}$$

for an ion energy of 3 eV [8].

2.3 Limits on ion accumulation

In the vast majority of cases (electron storage rings with typical pressures of 10^{-9} mbar, and ionization times of a second or less), ionization is, however, the dominant effect in the absence of any clearing mechanism. The production rates are (assuming identical ionization cross-sections for all charge states):

– For singly-ionized species (density N_i^+):

$$\frac{d\left(N_i^+\right)}{dt} = N_m\, N_p\, \sigma_i\, c - N_i^+\, N_p\, \sigma_i\, \mathrm{c} = N_p\, \sigma_i\, \mathrm{c}\left(N_m - N_i^+\right) \tag{13}$$

– Doubly ionized:

$$\frac{d\left(N_i^{++}\right)}{dt} = N_i^+\, N_p\, \sigma_i\, c - N_i^{++}\, N_p\, \sigma_i\, c = N_p\, \sigma_i\, c\left(N_i^+ - N_i^{++}\right) \tag{14}$$

etc. until (Z' being the total number of electrons of the gas atom):

$$\frac{d\left(N_i^{Z'+}\right)}{dt} = N_i\,\left(Z'-1\right) + N_p \sigma_i\, c \tag{15}$$

In the steady state (assuming constant ionisation cross sections):

$$N_i^+ = N_i^{++} = ... = N_i^{(Z'-1)+} \;\leq\; N_m \tag{16}$$

Therefore partially ionized ions can, at most, reach the molecular density N_m. Only the fully ionized state $N_i^{Z'+}$ could get close to the particle density (usually much larger than the gas density) divided by the final charge state: $N_i^{Z'+} \leq N_p/Z'$ corresponding to full neutralization of the particle beam.

In practise the ionization cross-sections are not independant of the charge state. Detailed calculations however do not change qualitatively the picture: highly ionized states should be present, unless unstable [9].

The degree of neutralization of a particular beam can be estimated from the incoherent tune shift. Almost full neutralization has been measured in the CERN AA when all the clearing electrodes are turned off (Fig. 2).

3. ION DYNAMICS (d.c. BEAM)

Produced with near thermal velocities, ions are generally not free to simply drift in the beam. Their motion is mostly governed by the beam space-charge potential when the beam is not fully neutralized, and by external forces such as magnetic fields in dipoles and quadrupoles. The Lorentz force due to the beam magnetic field is weak, and can be neglected in practical cases.

3.1 Ion oscillatory motion and azimuthal drift in magnetic field-free regions

The transverse distribution of beam charge results in an electric field E to which the charged trapped species is sensitive. This centripetal force (directed towards the centre of the beam) provokes a radial oscillatory motion with an amplitude equal to the radius at birth. Its frequency is proportional to the square root of the local field derivative times the ion charge, divided by its mass (bounce frequency).

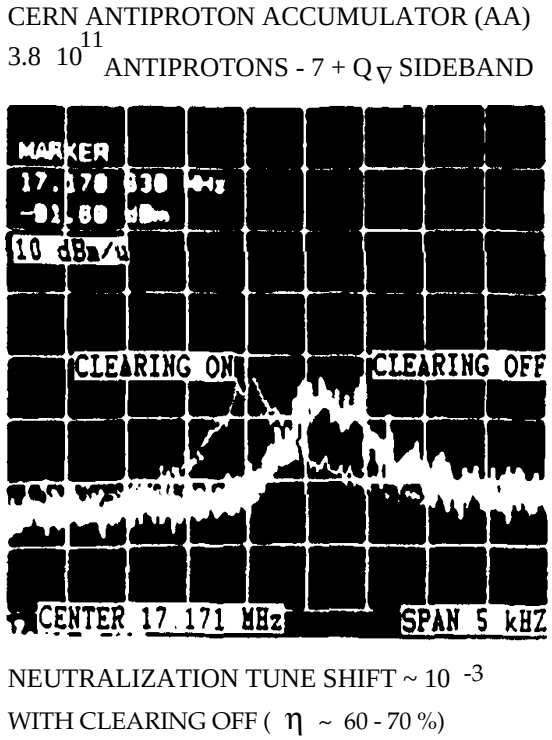


Fig. 2 AA transverse vertical Schottky scan showing the band $(7 + Q_V)f_{rev}$ with clearing electrodes turned OFF (full neutralization) and ON (little neutralization)

The longitudinal modulation due to changing beam sizes, and the varying vacuum chamber dimensions, give rise to longitudinal fields which drive the oscillatory ions around the machine. Usually negligible, the effect of the beam magnetic field may also contribute to the longitudinal motion for very high intensities.

To illustrate this with numbers, we consider the simple case of a round beam in a circular vacuum chamber, with a uniform transverse distribution of charges (Fig. 3).

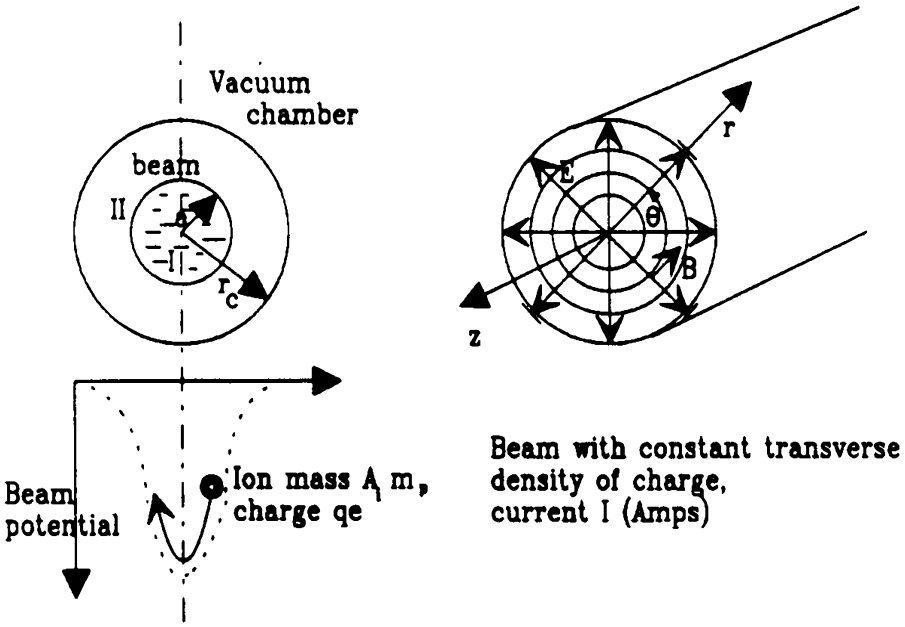


Fig. 3 Beam potential, and magnetic and electric fields of a uniform beam

The Lorentz Force acting on the ion is:

$$\vec{F} = qe\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{17}$$

with components:

$$\begin{cases} F_r = qe\left(E_r - B \cdot \dot{z}\right) \\ F_\theta = 0 \\ F_z = qe\left(B \cdot \dot{r} + E_z\right) \end{cases} \tag{18}$$

From the Gauss theorem, the electric field can be written as:

in region I:
$$E_r(r) = \frac{I}{2\pi\,\varepsilon_0\,\beta\,c}\;\frac{r}{a^2}$$

in region II:
$$E_r(r) = \frac{I}{2\pi\,\varepsilon_0\,\beta\,c}\;\frac{1}{r} \tag{19}$$

The magnetic field is simply:
$$B = \frac{\beta\,E_r}{c}$$

The beam potential at the centre is [1]:

$$U = \frac{I}{2\pi\,\varepsilon_0\,\beta\,c}\left(\ell\mathrm{n}\;\frac{r_c}{a} + \frac{1}{2}\right) \tag{20}$$

The forces on the ion reduce to:

$$F_r = A_i\, m_p\, \ddot{r} = \frac{qe\, I}{2\pi\, \varepsilon_0 \beta\, c\, a^2} \left(1 - \beta \frac{\dot{z}}{c}\right) r$$

$$F_z = A_i\, m_p\, \ddot{z} = \frac{qe\, I}{2\pi\, \varepsilon_0 \beta\, c^2}\, \dot{r}\, r + qe\, E_z \tag{21}$$

As the ions can only have a maximum potential energy equal to the beam potential times their charge, i.e. typically up to a few hundred eV, they are non-relativistic. Equations (21) can be uncoupled by neglecting $\beta\, \dot{z}/c << 1$. The ion motion is thus transversally oscillatory with a frequency ('bounce' frequency) of:

$$\omega_i = 2\pi\, f_i = \left(\frac{I}{2\pi\, \varepsilon_0\, \beta\, c} \; \frac{1}{a^2} \; \frac{q\, e}{A_i\, m_p} \right)^{1/2}$$

$$r(t) = r_m \cos\, \left(\omega_i t + \alpha\right) \tag{22}$$

The second expression in Eq. (21) can be integrated to yield the longitudinal drift velocity (neglecting some oscillatory terms) [10]:

$$\dot{z}(t) = \dot{z}(o) + \frac{1}{4c}\, \omega_i^2\, r(o)^2 + \frac{qe}{A_i\, m_p} E_z t \tag{23}$$

The second term is only important for very high bounce frequencies (50–100 MHz), i.e. for intense positive beams (I > 1 A) trapping electrons, as in the CERN ISR, and can be usually neglected. The third term is important; it dominates the longitudinal dynamics of the trapped species. As a numerical example, a 1 A negative beam 5 mm in radius in a 0.16 m diameter vacuum chamber would have a space charge field of 12 kV/m at the beam edge, resulting in an H_2^+ ion bounce frequency of 1.7 MHz. The beam potential at the centre would be ~ 200 V with typical azimuthal gradients (longitudinal field), resulting from beam and chamber size variations, of one to several V/m. With 1 V/m, 0.1 ms would be sufficient to get a drift velocity of 5000 m/s, well in excess of the thermal velocity $\dot{z}_0$. This illustrates the fact that the ion motion in field-free regions is governed by the beam space charge. In particular, so-called neutralization pockets may exist in places where the beam potential is deepest (bellows, chamber enlargments etc.): ions created there do not have enough energy to overcome the potential barriers to eventually reach the nearest clearing electrodes.

Neutralization pockets – or potential barriers – may also be created by highly insulating ceramic vacuum chambers becoming electrostatically charged [7]. Metallization of their inner surface is therefore important, as is the necessity to keep the vacuum chamber cross section as uniform as possible (shielded bellows, screens, etc.), thus, in this respect, joining the conditions imposed by impedance considerations. A precise knowledge of the potential variation around a machine is therefore important, as this also determines the strategic locations of the clearing electrodes.

Better expressions exist for more precise calculations of the beam potential of a beam at any location in a rectangular vacuum chamber. For elliptic beams with Gaussian transverse distributions, closed solutions have been given for the electric field [11] and the beam potential [12].

Using the beam envelope (Twiss) parameters, and the vacuum chamber dimensions as input to a computer program, these formulae can be used to produce a plot of the beam potential, maximum electric field, ion bounce frequencies, etc. around a machine [13], useful for locating clearing electrodes.

3.1.1 Transverse distribution of ions

In contrast to a widely used assumption, it has recently been shown that the transverse distribution of ions produced at rest is not a replica of the beam's distribution, but that instead it is composed of a central core narrower than the beam, with tails, if ion-ion and ion-neutral gas molecule forces can be neglected [14]. By explicitly solving the Liouville equation in a one-dimentional model, assuming a transverse beam Gaussian distribution, it can be shown that the transverse phase space distribution of ions produced at rest and oscillating in the beam space charge is greatly diluted at the beam edges when projected on its horizontal transverse axis (Fig. 4).

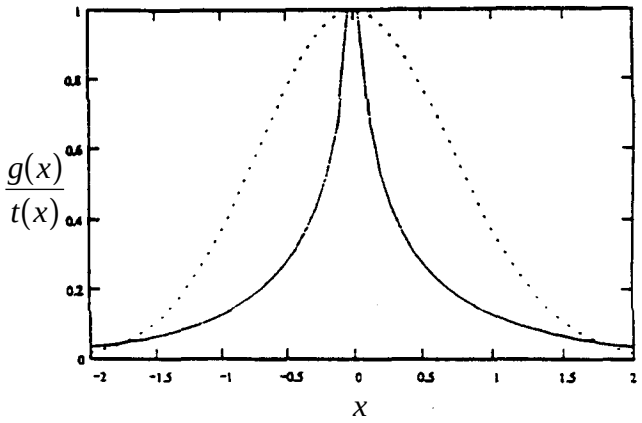


Fig. 4 Ion-cloud profile for ions created at rest (no thermal energy spread). The dashed line is the electron beam; the solid line is the ion cloud.

3.2 Drift motion in dipoles and duadrupoles, magnetic mirror (containment effect)

The motion of a non-relativistic, trapped species of charge qe and mass $A_1 m_p$ is well covered in Ref. [1] and classical electrodynamics textbooks, but will now be summarized for the more usual magnetic fields.

3.2.1 Quadrupole fields (gradient dB/dr)

The main effect is a longitudinal drift (perpendicular to the field lines) with velocity:

$$V_D = W_{kin} \frac{1}{qe\, B^2} \left(\frac{dB}{dr} \right) \tag{24}$$

neglecting the small additional term caused by the field lines curvature. W_{kin} is the ion's kinetic energy perpendicular to the beam axis. For ions produced at the centre of the beam, the drift can be very slow indeed.

Example: minimum ion energy (thermal): $W_{kin} = 4 \times 10^{-2}$ eV (ion produced at the centre of the beam).

$$B = 1 \text{ T}, \quad \frac{dB}{dr} = 10 \text{ Tm}^{-1} \tag{25}$$

yields a drift velocity of $V_D = 0.4$ m/s (for $W_{kin} = 10$ eV, $V_D = 400$ m/s).

Because of this possibility of relatively high neutralization, clearing electrodes are sometimes installed in quadrupoles, possibly combined with beam position electrostatic pick-ups. This is the case of the CERN AA, in which another effect has been experimentally seen when measuring the ion current drawn by these electrodes : namely an enhanced ionization rate possibly caused by a containment effect of electrons. As hinted in Ref. [1], electrons from the primary ions, of the same charge and thus repelled by the beam, may spiral around the quadrupole field lines towards the poles and be reflected back, in a so-called mirror effect. Without clearing, this additional effect could enhance the neutralization coefficient, through an increased ion production rate.

3.2.2 Uniform fields (B)

The cyclotron motion around the field lines has a frequency [1]:

$$\omega_{ci} = \frac{qe\,B}{A_i\,m_p} \tag{26}$$

with a radius $r_i = A_i m_p v_\perp / qeB$, $v_\perp$ being the species' velocity perpendicular to the field lines. The motion of the radius of gyration along the lines of force ($V_{//}$) is not affected; the spinning particle in dipoles, behaves vertically as in field-free regions. Owing to the combined action of the beam's space charge field $E_{(r)}$ and B, the longitudinal cross-field drift velocity is independent of the charge and mass of the ion, to a first approximation :

$$V_0 = \frac{E_{(r)}}{B} \tag{27}$$

If one takes into account the non-linearity of the electric field $E_{(r)}$, there is a slight dependence on the ion mass: heavy ions drift out more slowly than light ones [15].

Trapped species drift in opposite directions on each side of the beam, and the velocity at the centre is zero. As an example, for a 0.1 A beam with a radius of 5×10^{-3} m in a field of 1 T, the field at the beam edge is 1.2×10^3 V/m, and the drift velocity 1200 m/s. But the fact that the drift velocity falls to zero at the centre of the beam may result in a relatively high neutralization in dipoles. Experiments of beam shaking (see Section 6) in the CERN AA seem to support this fact, as hinted by the behaviour of the clearing currents drawn at the extremities of long bending magnets [16], the only ones around the ring to react to the shaking perturbation. In addition, the observation that neutralization effects can be reduced mainly with a vertical shaking of the beam near ion bounce frequencies (the CERN AA and the EPA) concords with the hypothesis that it affects ions in dipoles, where the vertical motion is the only degree of freedom that a coherent beam force can excite.

3.2.3 Dipole fringe fields

It has been shown [17] that slow ions with low kinetic energy drifting from no-field regions, can be reflected by the longitudinal gradient $\partial B/\partial z$ of the dipole field (fringe field). An ion oscillating horizontally in the beam space charge potential, and drifting towards a dipole, starts a cyclotronic motion around the vertical field lines, with a decreasing radius as it proceeds towards this magnetic field.

Motion reversal may occur for some initial conditions, such as a small azimuthal ion drift velocity. The criterion for ion reflection is [12, 18]:

$$\dot{z}_0 \leq \left| x_0 \, \frac{\omega_{ci}}{\Omega} \, \frac{\omega_{\mathrm{i}}}{2\pi\sqrt{2}} \right|$$

where:

$$\Omega^2 = \omega_{ci}^2 + \omega_i^2 .$$

Dipole fringe fields can therefore represent potential barriers for trapped species, and straight sections between dipoles must incorporate clearing electrodes. This effect, added to the field reversal, could be very detrimental in terms of ion trapping for undulators in electron storage rings.

4. ION DYNAMICS (BUNCHED BEAMS)

4.1 Field-free regions

Up to now we have only considered unbunched beams. Obviously bunching does not change the ion production rate but with our neutralization coefficient definition inside the bunch:

$$\eta_{\max}\left(N_i = N_p\right) = \frac{1}{B}$$

with the bunching factor defined as:

$$B = \frac{2\pi \, R}{n \, l_b} \; >> \; 1$$

R being the machine radius, n the number of bunches, and l_b the bunch length.

A trapped ion will therefore be submitted to the repetitive bunch space charge centripetal (focusing) force, independent of neutralization. If the ion motion is slow, and/or the ion heavy, or if the bunch spacing is uniform and small with respect to the machine circumference, the ion will mainly respond to the d.c. component of the Fourier expansion of the passing charge. In this case, on the average, the ion dynamics will be that of an unbunched beam, plus some stable oscillations. If these conditions are not met, the ions may perform resonant oscillations, thus becoming unstable and lost to the chamber wall. The ion motion in fact obeys a Hill's equation, similarly to the beam particles in a synchrotron. Analogously, the beam bunches represent thin focusing lenses for the ions. With uniform bunch repartitions, and in the linear approximation of a uniform beam, the analysis leads to the concept of a critical mass (or rather mass-to-charge ratio) above which ions perform stable oscillations and thus can be trapped [1]:

$$A_c \geq \frac{N_p\, r_p}{\beta\, n^2}\ \frac{\pi R}{b^2\left(1+\frac{a}{b}\right)} \tag{28}$$

Equation (28) ensures vertical stability of ion motion in a uniform elliptic beam of horizontal and vertical sizes b and a, respectively. For a similar horizontal stability criterium, b and a have to be interchanged in (28). Since usually $b < a$, the vertical stability criterium is the most stringent one. This expression shows that the critical mass increases (i.e. neutralization is likely to be less severe) with the number of beam particles, with reduced beam emittances, and with a smaller number of bunches. Figure 5 relevant to the CERN EPA machine, confirms this behaviour of the critical mass with varying numbers of bunches. For a given beam intensity, emittances (and neutralization) are lower for a small number of equidistant bunches, which corresponds to an increased critical mass.

On the other hand, for a given number of bunches, the calculated critical mass (circled numbers) remains more or less constant as the number of beam particles increases. This is because the increase in N_p in Eq. (28) is compensated by an increase in beam sizes a and b: neutralization therefore remains more or less constant as intensity increases, meaning more ions in absolute numbers, i.e. an increasing strength of non-linear forces on individual particles. This leads to increasing emittances as particle accumulation proceeds.

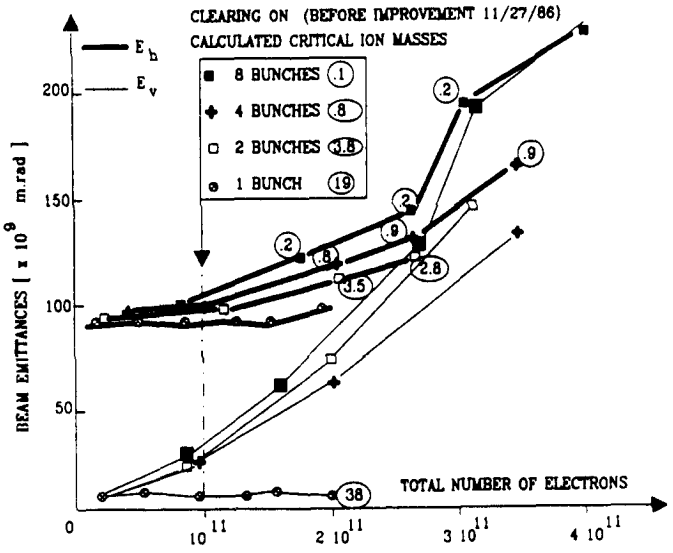


Fig. 5 Beam emittances and ion critical masses in the EPA for equidistant electron bunches

For completeness, it must be stressed that so far the criterium of ion stability has assumed linear and regular forces. Computer simulations can easily include the non-linearity of the force, important at large amplitude, and show that a large fraction of ions which ought to be stable according to the linear theory are in fact unstable if created with large initial position and velocity (r_0, $\dot{r}_0$). In addition, bunches are rarely equally populated, and this represents random gradient errors for the ion motion, in a similar way to synchrotrons, leading to 'enlarged' ion stopbands [19]. Finally, ions represent additional thick defocusing lenses for themselves, and a detailed analysis shows that this influences the limit of accumulation (although usually only weakly).

4.2 In magnets (bunched beams)

We have seen that in uniform fields (dipoles), the vertical motion of the centre of gyration is unaffected by the presence of the magnetic field B. Therefore the vertical stability criterium (Eq. (28)) holds.

In addition, in some rare cases and rather fortuitously, there may be a resonant condition between the horizontal cyclotron motion of the ions and the frequency of the passing bunches, leading to horizontal instability. In quadrupoles, the same argument holds to first order along the field lines with, in addition, an increased longitudinal drift velocity as resonant ions gain energy (see Eq. (24)).

5. CLEARING MEANS

5.1 Clearing electrodes

Clearing electrodes consisting of negatively polarized plates fitted into the vacuum chamber to provide a transverse electric field which diverts beam-channelled ions onto them so that the ions are neutralized and return into the gas phase. Figure 6 qualitatively represents the potential variation across a vacuum chamber of radius r_c, with an electrode on one side with a potential of U_{ce}.

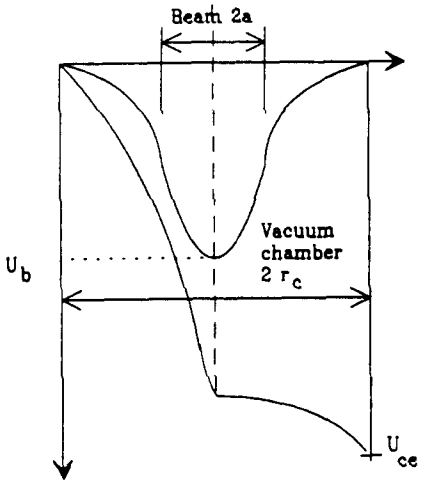


Fig. 6 Transverse potential distribution across a vacuum chamber, due to the beam space charge and a clearing electrode

A minimum condition for capture of the passing ion is that the transverse field provided by the electrode equals the maximum beam space charge field:

$$\frac{U_{ce}}{2r_c} > E(a) \sim \frac{I}{2\pi\beta c\varepsilon_0 a} \tag{29}$$

In fact, because ions may have transverse and longitudinal velocities corresponding to energies of up to a few eV, and because a clearing electrode is necessarily limited in size, the field provided by the electrode must usually be larger than the calculated beam field. For instance, in the EPA ring, where electrodes are of the button type with a diameter of 20 mm and are installed flush to the beam, transverse clearing in field-free regions is complete with the following parameters [20]:

$$U_{ce} = -6 \text{ kV} \tag{30}$$

electrode field: 30 kV/m on beam axis
beam max. field: 12 kV/m for $I = 0.3$ A
and ~10^{-8} mrad horizontal emittance with 10% coupling.

The number and the optimum location of electrodes are in principle dictated by the tolerable degree of residual neutralization. In practice, even with a large number of electrodes, uncleared pockets always remain, and contribute to typical residual neutralizations of a fraction to a few per cent.

No small electron storage ring exists which has reached a fully satisfactory ion-free situation, even with clearing electrodes. Perhaps one reason for this is that up to now clearing systems have not been very complete, partly owing to the fact that clearing electrodes complicate mechanical design of the vacuum chamber and may contribute to the machine impedance. At the CERN EPA, button-type clearing electrodes presenting negligible coupling characteristics with the beam have been designed. They are made of a ceramic body, coated with a highly-resistive glass layer (thick-film hybrid technology), and are terminated with a highly-lossy wide-band filter [21].

Nevertheless, it still remains to be demonstrated that full beam clearing can be achieved with a clearing system. The CERN AA has reached a low neutralization level for a d.c. beam machine (< 1%) with the help of an ever increasing number of electrodes, but neutralization pockets still remain.

5.2 Missing bunches (electron storage rings)

Many small electron storage rings prone to ion trapping have partly solved their problems by introducing one or several gaps in the bunch train, by not filling certain buckets at injection. To complement Ref. [1], and using the same notations, over one revolution period of a train of p consecutive bunches, the motion of a trapped ion (vertical here) is the solution of:

$$\begin{bmatrix} y \\ \dot{y} \end{bmatrix}_1 = M_{Ty} \begin{bmatrix} y \\ \dot{y} \end{bmatrix}_0 \tag{31}$$

with the transfer matrix:

$$M_{Ty} = \left(\begin{bmatrix} 1 & 0 \\ -a_y & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{f} \\ 0 & 1 \end{bmatrix} \right)^p \begin{bmatrix} 1 & (h-p)\frac{1}{f} \\ 0 & 1 \end{bmatrix} \tag{32}$$

The terms in parenthesis represent the linear kick received at the p bunch passages where [1]:

$$a_y = \frac{N}{n} \frac{r_p c}{b\left(1+\frac{a}{b}\right)} \frac{1}{A_i} \tag{33}$$

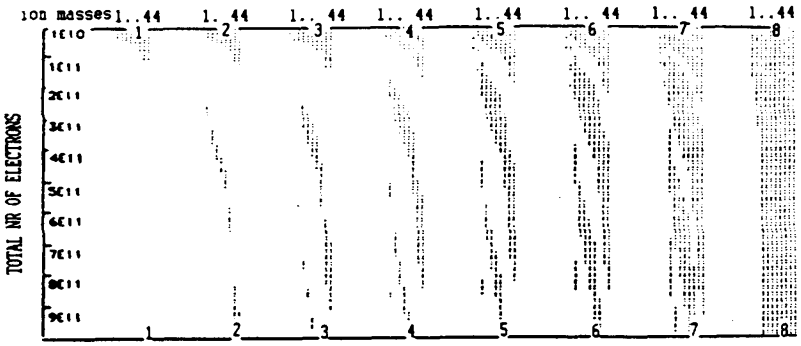
interleaved with drifts of duration $1/f$, f being the radio frequency. The period of p successive kicks plus drifts is terminated by the drift in the time interval (h - p) $1/f$, where h is the cavity

harmonic, i.e. the maximum number of bunches that the machine can handle. The Floquet's condition of stability for the ion of mass-to-charge A_i:

$$-2 < Tr\left(M_{Ty}\right) < 2 \tag{34}$$

does not lead to a simple criterion defining which A_i are stable. Rather, the trace of the transfer matrix (Eq. (32)) is of the order p in N, the total number of circulating particles. This means that there are p stable bands of ion mass-to-charge ratios for a given N, or that a given ion will be stable or unstable, depending on the number of beam particles, or on its location around the ring.

Figure 7 illustrates the conditions of linear stability for various ions in the EPA ring, as a function of the number of beam particles and consecutive bunches [15]:



LINEAR STABILITY OF IONS OF MASS = 1.2.12.16.18.20.22.28.40.44
AS FUNCTION OF THE NR OF CONSECUTIVE BUNCHES & TOTAL
NR OF ELECTRONS RMS HOR.&VERT. BEAM SIZES (mm) = .75 .25
(EPA average nominal beam dimensions with 10% coupling and 9 10^{-8}
m.rad nominal emittance (0 coupling))
(dotted vertical lines indicate stability for the given ions ...)

Fig. 7 Linear stability of ions in the EPA as a function of total beam intensity and the number of consecutive bunches for nominal beam emittances

Although not always absent, ion trapping is indeed less severe in the EPA with missing bunches [15]. This stems from the fact that since a given trapped ion can be further ionized, it then has a good chance of falling in an unstable band, and thus of being cleared.

6. RESONANT TRANSVERSE SHAKING OF THE BEAM

Quite recently, neutralization effects have been considerably reduced when exciting vertical coherent oscillations with a transverse kicker at a given frequency [22], both in electron storage rings (bunched) and Antiproton Accumulators (unbunched). This technique of 'RF knock out' has been determinant in solving ion problems in the CERN AA, where it has been studied both theoretically [23] and experimentally [24], and where it is permanently implemented, with the following parameters:

shaking: Vertical
shaking frequency: 490 kHz
sideband frequency (fractional tune q_v): 480 kHz
length of kicker electrodes: 0.6 m

kicker field: $\sim$ 20 V/cm

Although still at an early stage of both understanding and development, some beam shaking experimental observations can be summarized as follows:

1) Beam shaking works best when applied vertically: one possible reason is that neutralization is high in dipole fields (low ion drift velocity) where the motion along the lines of force is the only practical degree of freedom.

2) To work, beam shaking relies on the longitudinal motion of the ions. Owing to changing beam dimensions, the ion 'bounce frequency' spectrum is wide compared to the 'knock out' frequency : ions have to 'sweep' through this resonance. For this they must be free to move longitudinally. This is probably the reason why shaking works best in conjunction with clearing electrodes, since it ensures a low level of neutralization, permitting longitudinal field gradients which drive the ions around.

3) Beam shaking depends on the non-linearity of the beam space-charge field: this allows the 'lock-on' of the sweeping ions onto the resonance, where they keep large oscillation amplitudes, thus reducing their density in the beam centre.

4) Beam shaking is efficient even with low RF fields of only a few 10 V/cm, provided it is applied close to a beam betatron side band whose frequency lies close to the ion bounce frequency. In this case, the beam resonant response ensures sufficiently large non-linear forces on the ion. Experimentally it is found that for a weakly exciting RF field, shaking works best above a band $(n + Q)$ or below a band $(n - Q)$. This observation of assymetry of weak resonant shaking is important in that it validates the non-linear character of the ion motion and the 'lock-on' conditions.

To illustrate this in a simple way, we use a quasi-linear description of the two-body resonant conditions (from D. Möhl, see also Ref. 23]) for an unbunched beam. We consider only one ion species i, of mass-to-charge ratio A_i, with the following definitions:

Ω: circular revolution frequency of circulating beam ($\Omega = 2\pi f_i$)

$Q_i = 2\pi f_i / \Omega$ the ion bounce number in the beam potential well

$$\Omega_i \, Q_p^2 = \frac{2 N_p \, r_p \, c^2}{\pi \, b \, (a + b) \, \gamma R} \tag{35}$$

If Q_v is the beam particle unperturbed incoherent tune, and Q_p the beam particle bounce number in the ion-potential well where:

$$\Omega^2 \, Q_p^2 = \frac{2 N_i \, r_p \, c^2}{\pi \, \mathrm{b} \, (\mathrm{a} + \mathrm{b}) \, \gamma R} \tag{36}$$

and:

$$Q = \left(Q_v^2 + Q_\mathrm{p}^2 \right)^{1/2}$$

is the perturbed beam tune, then a beam particle and an ion obey the coupled set of linear differential equations:

particle:
$$\left(\frac{\partial^2}{\partial t^2} + \Omega\ \frac{\partial^2}{\partial \theta^2}\right) y_p + Q^2\ \Omega^2\, y_p - Q_p^2\, \Omega^2\, \bar{y}_i = F e^{i\omega t} \tag{37}$$

ion:
$$\left(\frac{d^2}{dt^2}\right)\ y_i + Q_i^2\ \Omega^2\ \left(y_i - \bar{y}_p\right) = 0$$

where the bar on y denotes the average vertical position of each beam, and the $Fe^{i\omega t}$ term is the harmonic of the external driving force close to beam and ion resonance:

$$\omega\ \sim (n \pm Q)\ \ \Omega \sim \mathrm{Q_i}\ \Omega$$

Assuming solutions of the form:

$$y_p = \xi_p\ e^{i\,(\mathrm{n}\theta + \omega\,\mathrm{t})}$$
$$y_i = \xi_i\ e^{i\omega\,\mathrm{t}} \tag{38}$$

the ion amplitude becomes:

$$y_i = \frac{Q_i^2\ \dfrac{F}{\Omega^2}\ e^{i\omega\, t}}{\left(x^2 - Q_i^2\right)\left((n+x)^2 - Q^2\right) - Q_p^2\, Q_i^2} \tag{39}$$

with:
$$x = \frac{\omega}{\Omega}\ \rightarrow\ (n \pm Q) \sim Q_i \tag{40}$$

Therefore, shaking works when y_i becomes large, i.e. when the denominator $\rightarrow 0$. But, as shown in Fig. 8, as the ions gain large amplitude, we have non-linear detuning such that $x^2 > Q_i^2$ (lock-on). Therefore for y_i to become large, requires that :

$$(n+x)^2 - Q^2\ \rangle\ 0$$

i.e.: $x < n - Q$ for excitation near a "slow-wave" beam frequency : $\omega\ \cong (n-Q)\ \Omega;\ n > Q$
$x > n + Q$ for a “fast-wave” frequency : $\omega\ \cong (n+Q)\ \Omega;\ n > Q$

This asymetry has been verified both in the CERN and Fermilab antiproton accumulators, and in the EPA ring [22].

To conclude on beam shaking as a means to suppress ion effects, it must be stressed that this technique is still at an early stage of development although already applied permanently to Antiproton Accumulators [24]. In the CERN EPA, shaking has made it possible to overcome a neutralization threshold which no other method (clearing electrodes, transverse kicks, etc.) could achieve.

From tune-shift measurement, clearing electrodes reduce the neutralization coefficient down to typically 2–3%. Shaking the beam reduces it further to well below 1%.

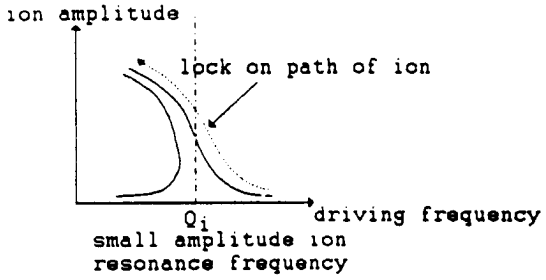


Fig. 8 Qualitative amplitude response curve of an ion versus the driving frequency, near resonance

7. CONCLUSION

Often very detrimental in their effects, ions may represent a real challenge to machine designers. We have seen that they can be partly eliminated with clearing electrodes, but this requires a careful design if one wants to avoid increasing the machine impedance.

Empty buckets on bunched machines and, more recently, beam shaking, are other means which are used to further reduce their numbers. The latter is still at the experimental stage but represents a real hope for some future machines where the very small beam sizes and high intensities may render very difficult the design of clearing electrodes.

REFERENCES

[1] Y. Baconnier, Neutralization of Accelerator Beams by Ionization of the Residual Gas, CERN/PS/PSR/84-24, CERN Accelerator School, Gif-suf-Yvette, France, 1984, CERN 85-19, pp. 267-288.

[2] S. Battisti, M. Bell, J.-P. Delahaye, A. Krusche, H. Kugler, J.H.-B. Madsen and A. Poncet, 12th. Int. Conf. on High Energy Accelerators, Fermilab, p. 2050, (1983).

[3] S. Kamada, Survey of Particle Factories Proposals and Projects, KEK Preprint 92-119, XVth Int. Conf. on High Energy Accelerators, Hamburg, Germany, (1992).

[4] D. Sagan, Ion Trapping in the CESR B-Factory, unpublished, CBN 91-2, (1991).

[5] J.-D. Jackson, in Classical Electrodynamics, John Wiley & Sons, Inc., New-York, (1962).

[6] F.-F. Rieke and W. Prepejchal, Ionisation Cross sections of Gaseous Atoms and Molecules for High-Energy Electrons and Positrons, Phys. Rev. A6 (1972).

[7] F. Pedersen and A. Poncet, Proton-Antiproton Instability in the AA, CERN PS/AA/ME Note 81, (1981).

[8] S.-C. Brown, in Basic Data of Plasma Physics, Technology Press, J. Wiley & Sons, Inc. New York, (1959).

[9] P. Tavarès, Bremsstrahlung Detection of Ions Trapped in the EPA Electron Beam, CERN/PS/93-25 (LP).

[10] B. Innert, Unpublished Note, ISR/VA/tn, (1977).

[11] M. Gygi-Hanney and B. Zotter, Field Strengh in a Bi-Gaussian Beam, CERN/LEP/Theory Note 44, (1987).

[12] R. Alves-Pires, Conformal Mapping for Two-dimensional Electrostatic Beam Potential Calculations, CERN PS/87-66 (AA), (1987).

[13] R. Alves-Pires, Beam Dimensions and Beam Potential in the CERN Antiproton Accumulator Complex, CERN PS/87-70 (AA), (1987).

[14] P. Tavarès, Transverse Distribution of Ions Trapped in an Electron Beam, CERN-PS/92-55 (LP), (1992).

[15] A. Poncet, Trapping of Ions in the EPA Electron Beam: Stabillty Conditions and Diagnosis, CERN/PS 88-14 (ML), (1988).

[16] F. Pedersen, A. Poncet and L. Søby, The CERN Antiproton Accumulator Clearing System, CERN PS/89-17 (ML), 1989 Particle Accelerator Conf., Chicago, (1989).

[17] Y. Miyamara, K. Takayama and G. Horikoshi, Dynamical Analysis on the Longitudinal Motion of Trapped ions in Electron Storage Rings, Nucl. Inst. Methods in Phys. Res., A270, p. 217, (1988).

[18] D. Sagan, Some Aspects of the Longitudinal Motion of Ions in Electron Storage Rings, CORNELL CLNS 90/1040, (1990).

[19] G. Brianti, Y. Baconnier, O. Grõbner, E. Jones, D. Potaux and H. Schönauer, Proceedings of the Workshop on $p\bar{p}$ in the SPS, SPS $p\bar{p}$.1, p. 121, (1980).

[20] J.-C. Godot, K. Hübner and A. Poncet, Comparison of the Electric Field in the Beam to the Available Clearing Fields in EPA, Note PS/LPI/87-25, (1987).

[21] F. Caspers, J.-P. Delahaye, J.-C. Godot, K. Hübner and A. Poncet, EPA Beam-Vacuum interaction and Ion Clearing System, CERN PS/88-37 (ML), 1988, Proc. EPAC, Rome, (1988).

[22] Y. Orlov, The Suppression of Transverse Instabilities in the CERN AA by Shaking the $\bar{p}$ Beam, CERN PS/89-01 (AR), 1989 Proc. Particle Accelerator Conf., Chicago, (1989).

[23] R. Alves-Pires et al., On the Theory of Coherent Instabilities due to Coupling between a Dense Cooled Beam and Charged Particles from the Residual Gas, CERN/PS/89-14 (AR), 1989 Proc. Particle Accelerator Conf., Chicago, (1989).

[24] J. Marriner, D. Möhl, Y. Orlov, S. Van der Meer and A. Poncet, Experiments and Practice in Beam Shaking, CERN/PS/89-48 (AR), 1989, Proc. 14th Int. Conf. on High Energy Accelerators, Tsukuba, Japan, (1989).

POLARIZATION OF ELECTRON AND PROTON BEAMS

J. Buon
Laboratoire de l'Accélérateur Linéaire, IN2P3-CNRS
et Université de Paris-Sud, France

J. P. Koutchouk
CERN, Geneva, Switzerland

Abstract
After defining the concept of polarization vector, its dynamics in static electro-magnetic fields is described in terms of the Thomas-BMT equation. The periodic orbital motion in circular machines is shown to excite spin resonances; they perturb the spin precession and cause depolarization. The acceleration of polarized protons illustrates the methods developed to avoid or weaken the depolarization. In electron rings, the consequence of synchrotron radiation is analysed in terms of spontaneous (Sokolov-Ternov) polarization and spin diffusion in presence of optics imperfections. It is shown how the spontaneous polarization can be used to determine with great accuracy the beam energy. Finally the local rotation of the spin in the direction of the velocity to allow colliding-beam experiments is discussed.

1. INTRODUCTION

1.1 Physics motivation

Spin is an important feature of nuclei and subnuclear particles, as well as their mass and electric charge. In general, the particle interactions depend to various degrees on their spin states. Experimental studies of these interactions with polarized beams and targets revealed important and new aspects of Nature. For instance, at SPEAR [1], the e^+e^- annihilation into multi-hadrons has been studied at 7.4 GeV center-of-mass energy. The angular distribution of the final state particles was found consistent with the quark-parton model; the data obtained with polarized beams confirmed that quarks are spin-1/2 particles. At Saturne II [2], experiments using polarized proton and deuteron beams on polarized proton targets have allowed an extensive study of the nucleon-nucleon scattering at intermediate energies.

Nowadays, the interest in polarized electron beams at high energies has come up again to test the Standard Model of the electro-weak interaction. In this model, the production of the Z vector boson with longitudinaly polarized e^+e^- beams is left-right asymmetric. An accurate measurement of the asymmetry can give information on the mass range of the top quark and Higgs boson. Another example of the relevance of polarized beams is the study of the distribution of the angular momentum inside the nucleons.

A less direct but important contribution is to make possible an extremely accurate measurement of the beam energy and hence of the mass of the particles created in the e^+e^- annihilations. This is illustrated by the figures of Table 1 taken from [3] and by the present campaign of measurement of the mass and width of the Z particle at LEP.

Table 1
Improvement of the determination of particle masses
obtained from resonant depolarization of polarized e^+e^- beams

Particle	World average value (MeV)	Experimental results (MeV)	Year publication	Accuracy improvement
$K^{\pm}$	493.84 ± 0.13	493.670 ± 0.029	1979	5
K^o	497.67 ± 0.13	497.661 ± 0.033	1987	4
ω	782.40 ± 0.20	781.780 ± 0.10	1983	2
ϕ	1019.7 ± 0.24	1019.52 ± 0.13	1975	2.5
J/ψ	3097.1 ± 0.90	3096.93 ± 0.09	1981	10
ψ'	3685.3 ± 1.20	3686.00 ± 0.10	1981	10
Υ	9456.2 ± 9.50	9460.59 ± 0.12	1986	80
Υ'	10016.0 ± 10.	10023.6 ± 0.5	1984	20
Υ''	10347.0 ± 10.	10355.3 ± 0.5	1984	20

1.2 Historical summary

In the early sixties, knowledge of beam polarization and the associated technology allowed one to contemplate the acceleration of polarized particles: Froissart and Stora [4] calculated the depolarization on spin resonances for protons while Bargmann, Michel and Telegdi [5] reformulated the Thomas equation for the spin precession in arbitrary electromagnetic fields. Ternov, Lostukov and Korovina [6] discovered theoretically the spontaneous polarization of electrons in a magnetic field, which was analysed in detail by Sokolov and Ternov [7]. In 1970, radiative polarization was indeed observed and studied at ACO (Orsay) [8,9] and at VEPP-2 (Novosibirsk) [10,11] at about 0.5 GeV. Soon after, polarized protons were accelerated up to 12 GeV in the ZGS (Argonne) [12].

The last two decades have seen the development of techniques to accelerate polarized beams at higher and higher energies. Polarized protons have been successfully accelerated in several synchrotrons. At Saturne (Saclay) [13,14] beams of 2×10^{11} protons/burst, fully polarized, are routinely accelerated at energies up to 3 GeV. The maximum energy of 22 GeV [17] has been reached at the AGS (Brookhaven) [15]. Acceleration of polarized particles in the latest projects (RHIC, SSC) has been considered. Polarized electron beams

have also been accelerated at the Bonn 2.5 GeV synchrotron [16], at the high-energy linac of SLAC to 22 GeV [17] and recently up to 46 GeV at the Stanford Linear Collider (SLC) [18]. The most outstanding observation is the in-situ polarization build-up of electron beams circulating in large storage rings (HERA [19], Tristan [20]) up to the highest energy of 46 GeV at LEP [21], in spite of the potentially very strong depolarization phenomena.

1.3 Outline of the lecture

The first section introduces the concept of polarization for a spin-1/2 particle beam; the spin motion in static fields is discussed and the depolarizing effects explained by the coupling between the spin precession and the orbital motion in the accelerator. The enhancement of the depolarization by resonant mechanisms is studied and the different kinds of spin resonances arising in circular accelerators are given.

The second section deals more specifically with the acceleration of polarized protons in synchrotrons (see also previous reviews [22,23,24]). The depolarization due to resonance crossing during acceleration and the cures applied are discussed. One of the most promising techniques to overcome depolarization, the "Siberian snake", is presented as well as its first experimental tests.

In the last section, the specific case of relativistic electron beams in storage rings is considered (see also previous reviews [25,26,27]). The in-situ polarizing mechanism due to synchrotron radiation (the Sokolov-Ternov effect) is introduced. In the presence of optics imperfections the synchrotron radiation is also responsible for a harmful spin diffusion. This depolarizing mechanism, enhanced on spin resonances, is discussed as well as ways to improve polarization. The method to accurately calibrate an electron ring in energy, by crossing an RF resonance applied on purpose, is described. Finally, the spin rotators, which allow the rotation of the polarization vector from its natural vertical direction to the direction of the particle velocity, are presented as well as the requirements they must fulfill to avoid depolarization.

The polarized sources and the polarimeters are beyond the scope of this lecture. We avoided the SU(2) algebra, not essential and not familiar in accelerator physics, and formal proofs of the basic equations. We have preferred to limit ourselves to an analysis of the physical contents of the basic equations and of their consequences, illustrated by experimental results.

2. GENERAL SPIN DYNAMICS

2.1 Polarization of a spin-1/2 particle beam

2.1.1 Spin and magnetic moment

The spin of a particle is an internal degree of freedom which behaves like angular momentum. It is represented by the quantum vectorial operator $\hat{\mathbf{S}} = \{\hat{\mathbf{S}}_x, \hat{\mathbf{S}}_y, \hat{\mathbf{S}}_z\}$. The class of spin $s = 1/2$ particles (electrons, muons, protons, neutrons, ..) corresponds to the eigenstates of $\hat{\mathbf{S}}^2$ with eigenvalue $s(s+1) = 3/4$. With respect to an arbitrary quantization axis Oz, a spin - 1/2 particle can only be found in one of the two eigenstates of $\hat{\mathbf{S}}_z$ with the eigen values $S_z = +\hbar/2$ ("up" state) or $S_z = -\hbar/2$ ("down" state). $\hbar$ is the Planck constant divided by 2π.

A magnetic moment $\hat{\mu}$ proportional to the spin $\hat{\mathbf{S}}$ is associated to a spinning charged particle:

$$\hat{\mu} = \frac{g}{2}\frac{e}{m_0}\hat{\mathbf{S}} = (1+a)\frac{e}{m_0}\hat{\mathbf{S}} \qquad (1)$$

where e and m_0 are respectively the electric charge and the rest mass of the particle. The gyromagnetic ratio g is 2 for a point-like spin-1/2 particle in the Dirac theory. For real particles its deviation from 2 is expressed by the gyromagnetic anomaly [28] $a = (g - 2)/2$ (very often designated by G in the literature):

electron	muon	proton
$a = 1.5965 \times 10^{-3}$	1.16592×10^{-3}	1.79285

In a field of magnetic induction **B**, the magnetic energy $\hat{\mathbf{W}}$ of a charged particle at rest is given by :

$$\hat{\mathbf{W}} = -\hat{\mu}\cdot\mathbf{B} \qquad (2)$$

2.1.2 Polarization of a single particle

A single spin-1/2 particle in a pure spin state along a given direction is fully polarized. Its polarization vector **P** is the unit vector along that direction. In general, a particle is in a mixed state, i.e. it can be found in any of the pure spin states with a certain probability distribution. Its polarization vector is the distribution-weighted average of the polarization vectors of the pure spin states. The degree of polarization is the modulus of **P**, in general smaller than one. A particle is unpolarized when $|\mathbf{P}| = P = 0$. It can be shown that two mixed states having the same polarization vector cannot be distinguished. A mixed state of an electron can thus be considered as the incoherent superposition of the "up" and "down" states along the direction of **P** with the probabilties $1/2\,(1 \pm P)$. In an ideally planar ring for instance, the ultimate degree of polarization is 92.36%, corresponding to probabilities of 96.18% and 3.82% respectively. In a gedanken experiment where the particle would be duplicated at infinity in the same mixed state, the degree of the polarization is the asymmetry of the populations N_+ and N_- of the "up" and "down" states:

$$P = \frac{N_+ - N_-}{N_+ + N_-} \qquad (3)$$

It can be shown that the polarization vector is the quantum average of the spin operator $\hat{\mathbf{S}}$ expressed in units of $\hbar/2$ when the particle is in a pure state.

$$\mathbf{P} = \frac{2}{\hbar}\langle \psi\,|\hat{\mathbf{S}}|\,\psi\rangle = \mathbf{S} \qquad (4)$$

In the semi-classical description of the spin dynamics, it is by convention named the spin vector noted **S.**

The same property is also valid for a mixed case. It follows from (4) that the asymmetry of a spin-dependent quantity is proportional to **P.** For instance, in an experiment only sensitive to the longitudinal spin component (parallel to the particle velocity), it would be proportional to the longitudinal component of **P**.

According to the general rules of quantum mechanics (Ehrenfest theorem), the quantum average of $\hat{\mathbf{S}}$ behaves as a classical quantity. Its evolution can be described by a classical deterministic equation which accounts for the motion in a static electromagnetic field, including quantum spin flips between the eigenstates due to interaction with a radiative field.

2.1.3 Polarization of a beam of particles

The particles in a beam are distributed in oscillation amplitudes, defining the beam emittances. The same spreading occurs for the spin vectors. For instance, in a proton beam, particles with very small betatron amplitudes may be fully polarized along the vertical direction; particles with larger amplitudes suffer spin precessions in the quadrupole fields and become unpolarized with respect to the vertical direction. The polarization vector **P** of the beam of particles is the statistical average of the individual spin vectors. Its length is the polarization degree of the beam. A beam is unpolarized ($P = 0$) if the spin vectors point to all directions isotropically. It is fully polarized ($P = \pm 1$) if all the particles are in the same pure spin state. For the same reason as above, the polarization vector of a beam is a classical quantity. The general approach to evaluate the beam polarization is to study the dynamics of the individual spin vectors and average over the six orbital coordinates of the particles to obtain the beam polarization.

This description of the beam polarization can be extended to the case of a deuteron beam. Deuterons are particles of spin 1 with three eigenstates $S_z = +\hbar, 0, -\hbar$ along any direction Oz. Their gyromagnetic anomaly [29] is $g = -0.142562$. The polarization vector **P** is again the statistical and quantum average of the spin quantum operator $\hat{\mathbf{S}}$ and is mostly sufficient to determine the polarization state. The acceleration of polarized deuterons is similar to the acceleration of protons and even easier.

2.2 Spin precession in a static electromagnetic field

2.2.1 Non-relativistic particles

For non-relativistic non-radiating particles, the simplest way to establish the motion of the spin is to apply the fundamental law of classical mechanics: the rate of change of the angular momentum is equal to the torque exerted by the external field:

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \mu \times \mathbf{B} \tag{5}$$

Expressing μ in (5) in terms of **S** according to (1) yields the Thomas equation:

$$\frac{d\mathbf{S}}{dt} = \Omega_L \times \mathbf{S} \tag{6}$$

This is the equation of a rotation (precession) of the spin vector **S** around the direction defined by the rotation vector Ω_L, colinear to the magnetic field **B** (Fig.1). The modulus of the rotation vector is the Larmor frequency.

$$\Omega_L = -\frac{g}{2}\frac{e}{m_0}\mathbf{B} \tag{7}$$

This precession is similar to the velocity rotation in a magnetic field:

$$\frac{d\mathbf{v}}{dt} = \Omega_c \times \mathbf{v} \tag{8}$$

with

$$\Omega_c = -\frac{e}{m_0}\mathbf{B} \tag{9}$$

For point-like spin-1/2 particles ($g = 2$) the spin precession frequency is equal to the cyclotron frequency. The relative deviation of the spin precession frequency with respect to the cyclotron frequency is just the gyromagnetic anomaly a :

$$\frac{\Omega_L - \Omega_c}{\Omega_c} = a \tag{10}$$

The measurement of this deviation is the principle of all the "$g-2$" experiments that aim to measure the gyromagnetic anomaly.

2.2.2 Relativistic particles

In an electromagnetic field a relativistic particle experiences an acceleration. To transform the electro-magnetic fields from the laboratory to the accelerated rest frame of the particles, a double Lorentz transform is required [30]. The spin precession retains the same law as (6), albeit with a different rotation vector. In this form, it takes its name from Bargman, Michel and Telegdi [5] (abbr. BMT equation) :

$$\frac{d\mathbf{S}}{dt} = \Omega_{BMT} \times \mathbf{S} \tag{11}$$

with :

$$\Omega_{BMT} = -\frac{e}{m_0\gamma}\left[(1+a\gamma)\mathbf{B}_\perp + (1+a)\mathbf{B}_{//} - \left(a + \frac{1}{\gamma+1}\right)\gamma\,\beta \times \frac{\mathbf{E}}{c}\right] \tag{12}$$

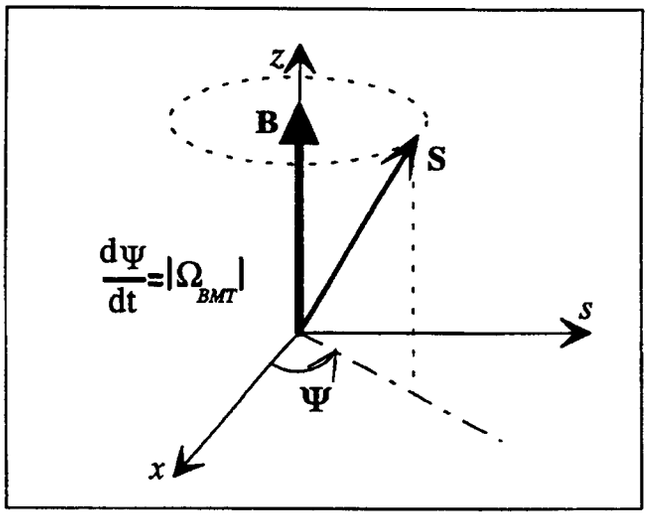


Fig 1: Precession of the spin vector **S** about the magnetic field **B.**

where $\mathbf{B}_\perp$ ($\mathbf{B}_{//}$) is the transverse (longitudinal) component of the induction **B** relative to the particle velocity ; γ is the relativistic Lorentz factor and β is the velocity in units of the light velocity c. In that equation the fields **E** and **B** and the time t are expressed in the laboratory frame (all quantities in MKS units). On the other hand, the spin vector **S** is expressed in the instantaneous rest frame of the particle and the BMT equation includes the Thomas precession [31,32] that occurs in that frame.

The spin precession has the following general properties :

- The rotation in an electric field **E** or a magnetic field $\mathbf{B} = \mathbf{E}/c$ is of the same order of magnitude: an electric field of $3\text{x}10^8$ V/m has a rotating strength comparable to that of a 1 Tesla magnetic field. Applied electric fields in accelerators are in general smaller and their effect on the spin negligible. However, it results from the Lorenz transformations that the electric and magnetic forces between two charged and ultra-relativistic particles have asymptotically the same modulus. They cancel out when the velocities are parallel, e.g. in the same beam, but add to each other when the velocities are anti-parallel, e.g. when the beams are colliding. For the spin precession, the same property results from the BMT equation.
- The spin rotating strength of a field $\mathbf{B}_{//}$ parallel to the velocity is inversely proportional to the particle momentum **p**, as the velocity rotating strength of a transverse field **B** is. The parallel-field integral needed to rotate the spin by one radian is :

$$\int \mathbf{B}_{//} ds \ \text{(Tm/rad)} = -\frac{1}{1+a}\frac{\mathbf{p}}{e} \approx -\frac{10.479}{\pi\,(1+a)}\mathbf{p} \qquad (13)$$

At 1 GeV/c, a parallel-field integral of 10.479 Tm is needed to rotate the spin by 180° for a particle with vanishing gyromagnetic anomaly.

- The difference in frequency of spin and velocity precessions about a transverse magnetic field $\mathbf{B}_\perp$

$$\Omega_a = \Omega_{BMT} - \Omega_c = \gamma a \left[\frac{e}{m_0 \gamma} \mathbf{B}_\perp \right] \tag{14}$$

is exactly independent of the particle energy, and is γa times larger than the relativistic cyclotron frequency $\Omega_c = e\mathbf{B}_\perp / m_0\gamma$.

$$\Omega_a = \gamma a = \Omega_c \tag{15}$$

The vector Ω_a is the spin rotation vector with respect to a frame attached to the particle trajectory (the so-called orbit frame); the latter rotates at the cyclotron frequency Ω_c following the velocity β.

- The transverse-field integral required to rotate the spin by one radian with respect to the orbit frame is slightly dependent on the energy through the relativistic velocity

$$\int \mathbf{B}_\perp ds(\text{Tm/rad}) \approx -\frac{5.484}{\pi}\beta \quad \text{for a protron,} \tag{16}$$

$$\int \mathbf{B}_\perp ds(\text{Tm/rad}) \approx -\frac{4.618}{\pi}\beta \quad \text{for an electron} \tag{17}$$

In a given transverse field, the spins of protons and electrons precess by comparable amounts: the larger mass of the proton is compensated by its larger gyromagnetic anomaly.

- At high energies, when $\gamma a >> 1$, the spin rotation by transverse fields becomes nearly energy-independent while the spin rotation by longitudinal fields becomes less and less effective. Transverse fields are preferred for spin manipulations at high energies. Moreover, the spin rotation requires much less bending field than the trajectory does and the spin vector precesses much faster than the particle velocity. For example, the spin precesses 103.5 times per machine revolution at the 45.6 GeV in LEP (Z production energy), i.e. in about 0.1 ms. The time scale of this process is in great contrast with that of the spontaneous polarization (minutes to hours), as we shall see later, and with the time it takes to emit a photon (typically 10^{-10} s). This makes it possible to study separately the polarizing and depolarizing phenomena.

2.3 Spin precession in a circular particle accelerator

2.3.1 Ideal accelerator

As we have seen above, the spin motion in a constant magnetic field is a simple precession. Calculating the spin motion in an ideal accelerator or storage ring is a straightforward extension but allows one to introduce the formalism necessary to treat the general case. All the elements (dipoles, quadrupoles, sextupoles) of the ideal ring are supposed to lie in the same plane, e.g. horizontal. At any energy, the closed orbit lies in the same horizontal plane. A vertical spin of a particle circulating on the closed orbit will stay vertical, parallel to the guide field. This constant spin direction is called the spin closed solution $\mathbf{n}_0$.

Any other spin direction precesses about the spin closed solution $\mathbf{n}_0$, vertical in this case. The number ν of spin precessions per orbit revolution is found by replacing Ω_c by 2π in (15):

$$\nu = a\gamma \tag{18}$$

i.e. numerically:

	electrons	protons	deuterons
$\nu =$	$\frac{E(\text{GeV})}{0.44065}$	$\frac{E(\text{GeV})}{0.52335}$	$\frac{E(\text{GeV})}{13.1565}$

as a function of the total relativistic energy E. Unlike the betatron tunes, the spin tune increases linearly with the energy.

In fact, the magnetic fields are distributed in a piece-wise way along the orbit. The spin precesses in the bending magnets and remains constant in straight-sections as long as the trajectory passes the quadrupoles on-axis. If the spin precession is Fourier-analysed with respect to the machine azimuth, the spectrum of the spin precession is no longer a single line; it shows instead satellites at $\pm k$, $k \in N$, which reflect the distribution and symmetry of the bending magnets.

Particles experiencing horizontal betatron oscillations and synchrotron oscillations are confined to the horizontal plane. The vertical direction is again the spin closed solution $\mathbf{n}_0$. The spin precession frequency about $\mathbf{n}_0$ is still $\nu = a\gamma$ on average. The spectral analysis shows however satellites at $\pm k Q_x$ and $\pm k Q_s$, $k \in N$ where Q_x and Q_s are the betatron and synchrotron tunes. These satellites are due to the extra spin rotations caused by the non-vanishing amplitudes of the beam oscillations in the quadrupoles and sextupoles. For all these horizontal trajectories, the vertical spin component S_z is conserved; there is thus a strong incentive to have circulating beams with a vertical polarization.

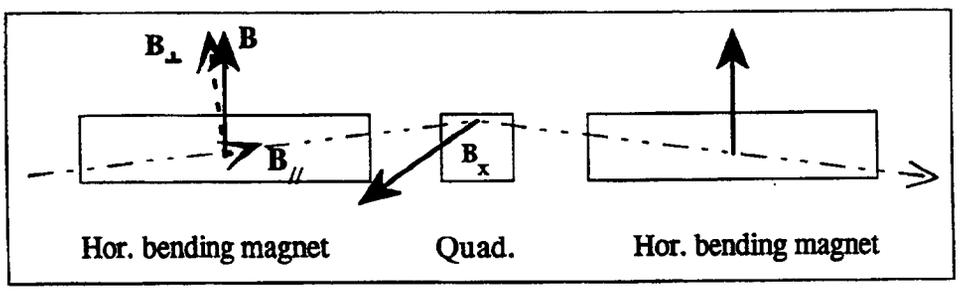


Fig. 2: Fields along a trajectory oscillating in the vertical plane and passing through horizontally bending magnets and quadrupoles.

Even in an ideal accelerator, beams of non-radiating particles are subject to vertical betatron oscillations due to the finite vertical emittance. Along this vertical betatron oscillation, the field $\mathbf{B}_x$ in the quadrupoles is horizontal and proportional to the vertical displacement (see Fig. 2). A small parallel field $\mathbf{B}_{//}$ is also experienced in the horizontally bending magnets

where the trajectory has a vertical slope. The spin is thus bent away from the vertical direction, in a trajectory dependent way. This phenomenon is called *spin-orbit coupling*. At 46 GeV in LEP for instance, the spin precession in an arc quadrupole is $\Delta\psi = a\gamma Kl \cdot \Delta y \approx 2.5$ mrad/mm with Kl integrated gradient and Δy vertical beam displacement in the quadrupole. Even in an ideal accelerator, the spin-orbit coupling may give rise to depolarization. It can be minimized by reducing the vertical beam emittance or by special optics provisions which, by counteracting the spin-orbit coupling, allow to recover the spin transparency of the ideal accelerator. (See sections 4.3, 4.4.)

2.3.2 General case

Although, in general, particle accelerators are designed to lie in a plane, the unavoidable imperfections of alignment and of the magnetic fields break the symmetry about the horizontal plane. In many cases, vertical deflections or solenoidal fields are even introduced on purpose to combine or separate colliding beams, provide means to analyse the products of beam-beam interactions or rotate the spin. On the non-planar closed orbit, the spin is subject to successive rotations in the magnetic elements as the particle makes one revolution starting at azimuth s. Unlike the case of the ideal accelerator, these rotations occur about different axes. The one-turn mapping of the spin is the product of the local spin rotations and is a rotation itself $\mathbf{T}(s)$. The real eigenvector of the rotation matrix is the only vector which, transported by the Thomas-BMT equation once around the ring, reproduces itself:

$$\mathbf{n}_0(s) = \mathbf{T}(s)\mathbf{n}_0(s) \tag{19}$$

It is easy to prove that the one-turn rotation axes $\mathbf{n}_0(s)$ along the machine can all be deduced from each other by virtue of the BMT equation. Let us consider two azimuths s_0 and s_1, where the one-turn mappings are $\mathbf{T}_0$ and $\mathbf{T}_1$ with eigenvectors $\mathbf{n}_{00}$ and $\mathbf{n}_{01}$ (Fig. 3).

Let $\mathbf{R}_{01}$ be the spin transport matrix from s_0 to s_1. The principle is to express the transport $\mathbf{T}_0$ as a function of $\mathbf{T}_1$ in the equation of $\mathbf{n}_{00}$:

$$\mathbf{n}_{00} = \mathbf{T}_0\mathbf{n}_{00} = \mathbf{R}_{01}^{-1}\mathbf{T}_1\mathbf{R}_{01}\mathbf{n}_{00} \tag{20}$$

Hence

$$\mathbf{R}_{01}\mathbf{n}_{00} = \mathbf{T}_1\mathbf{R}_{01}\mathbf{n}_{00} \tag{21}$$

$\mathbf{R}_{01}\mathbf{n}_{00}$ is by definition the real eigenvector $\mathbf{n}_{01}$ of T_1, i.e $\mathbf{n}_{01}$ is obtained from $\mathbf{n}_{00}$ by the spin transport from s_0 to s_1. $\mathbf{n}_0(s)$is thus the spin closed solution. Because rotations conserve angles, any other spin direction will appear to precess about it, much like a betatron oscillation around the closed orbit. As we already saw, the spin closed solution is vertical in an ideal ring lying in a horizontal plane. Usually horizontal bending fields are small and the spin closed solution is nearly vertical. The one-turn spin precession $2\pi\nu$ is given by the phase angle of the two imaginary eigenvalues of the one-turn matrix $\mathbf{T}(s)$. It is independent of the azimuth s since $\mathbf{T}_0$ and $\mathbf{T}_1$ are similar matrices owing to (20). Hence

$$2\pi\nu = \cos^{-1}\left(\frac{\text{Trace}\,(\mathbf{T}(s))-1}{2}\right) \quad , \forall s \tag{22}$$

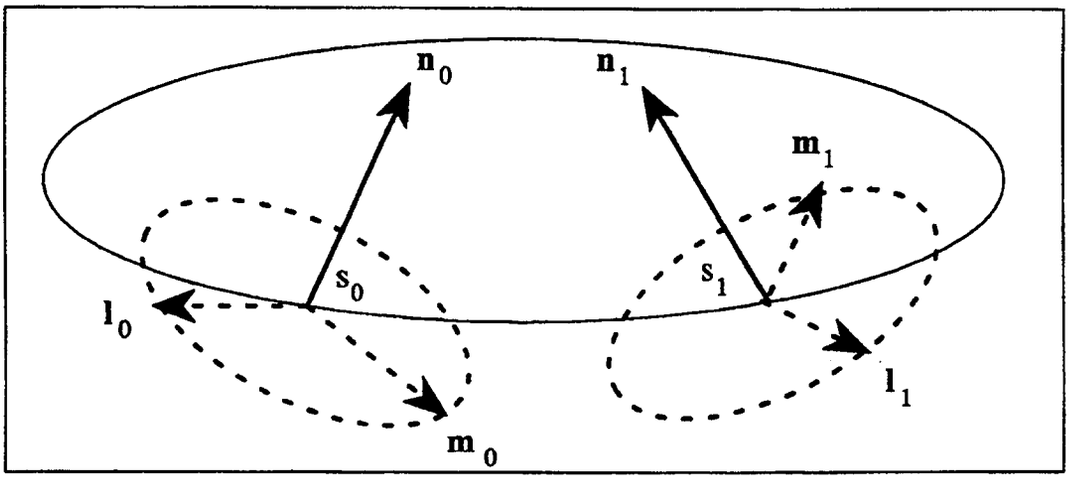


Fig. 3: Transport of the orthogonal spin frame $(\mathbf{l}_0, \mathbf{m}_0, \mathbf{n}_0)$ from s_0 to s_1.

In an ideal ring, the spin tune was shown to be simply related to the energy (18). In a real ring with small vertical bends, this is not true any more. In general, the deviation is very small and can be ignored. It may however set a limit to the accuracy of the calibration by resonant depolarization, as will be shown later. In special cases, the spin precession is perturbed on purpose by large rotations which shift the spin tune. The extreme is the "Siberian snake"[33,34] which enforces half-integer spin values, whatever the particle energy. This special insertion will be analysed later (see section 3.3).

So far we considered the particle motion on the closed orbit. Particles oscillate about it with three degrees of freedom. The oscillation amplitudes being generally small (but not always), the resultant spin precessions may be considered as a perturbation of the spin motion on the closed orbit. It is thus convenient to consider the spin motion in a frame which eliminates the spin motion on the closed orbit. Such an orthonormal frame $(\mathbf{l}_0, \mathbf{m}_0, \mathbf{n}_0)$ is obtained by associating to the spin closed solution $\mathbf{n}_0(s)$ two orthogonal spin solutions $\mathbf{l}_0(s)$ and $\mathbf{m}_0(s)$ precessing around $\mathbf{n}_0(s)$ by $2\pi\nu$ in one revolution. At this point, it should be noted that the calculation of the electron depolarization requires a more general definition of $\mathbf{n}$ on any non-periodic trajectory which coincides with $\mathbf{n}_0$ on the closed orbit. To simplify the notation, $(\mathbf{l}_0, \mathbf{m}_0, \mathbf{n}_0)$ will be noted $(\mathbf{l}, \mathbf{m}, \mathbf{n})$ unless the more general definition is required.

Along the perturbed orbital trajectories, the spin-orbit coupling is responsible for beam depolarization. The spread of energies and trajectories among the particles in a beam leads to a spread in the spin precession. Assuming that all the individual spins would initially point to the same direction (full polarization), after some time they will spread in different directions. The modulus of the beam polarization vector, i.e. the degree of polarization, decreases. As we shall

see later, the depolarization due to the spin-orbit coupling shows resonant properties. Minimizing the depolarization requires the minimization of the spin-orbit coupling and a selective compensation of the most important spin resonances. The conditions to fulfill are referred as the "spin transparency conditions [35]" obtained by "spin matching". (See sections 4.3, 4.4.)

Returning to the spin motion on the closed orbit, one might think that, because the spin and the velocity follow the same dynamics as given by equations (8) and (11), the spin should be restored when the orbit angle is restored. This is not so: unlike the orbit, the spin experiences large rotations in the orbit frame due to the dipole field. For example, the spin precesses by 1.2 rad in each LEP 1/2 cell at 46 GeV. Because of the non-commutativity of finite rotations, the spin is not restored if the orbit is. This is illustrated in Fig. 4: in Fig. 4a, the vertical orbit is locally perturbed between points A and B by exciting a π-bump; the tip of the spin vector, seen from above, is shown not to be restored on Fig. 4b.

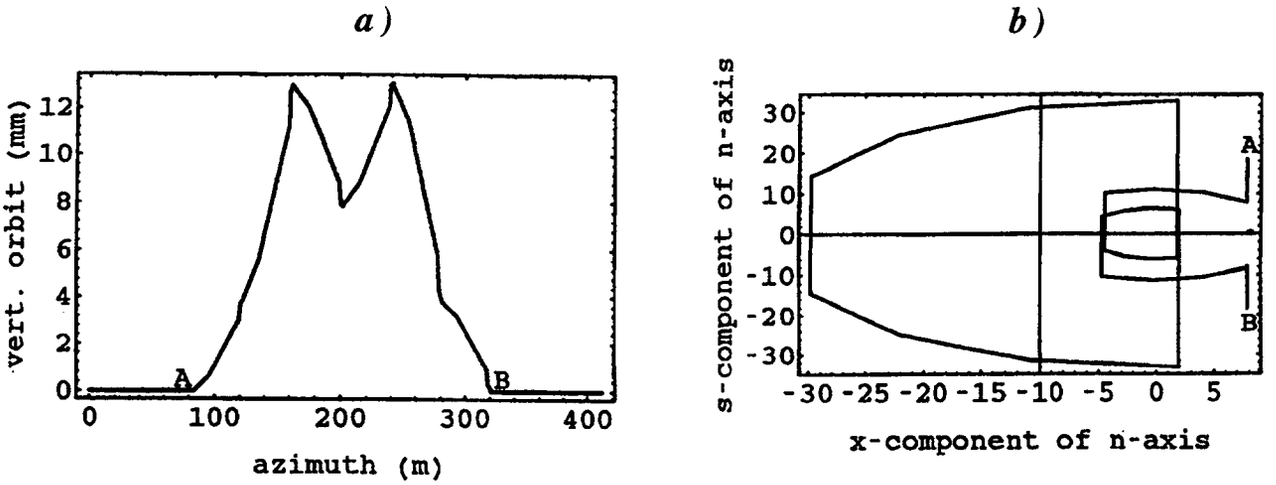


Fig. 4: Rotations of the orbit (closed bump) in *a*) and of the spin in *b)*.

This property allows one to manipulate the spin direction by modifying locally the closed orbit. At high energy where the spin is much easier to rotate than the velocity, these bumps can be made relatively small and can be optically matched to the rest of the ring. Spin rotators and "Siberian snakes" become conceivable insertions.

2.3.3 Spin-dependent orbit perturbation

So far, we considered the consequences of the torque exerted by the magnetic force on the magnetic moment of the particle. This is not the only interaction in addition to the usual Lorenz force. If the gradient of the magnetic field does not vanish, the particle momentum receives a kick $\int \mathbf{F} ds$ due to the interaction of the magnetic moment with the field gradient:

$$\mathbf{F} = \mu \cdot \nabla \mathbf{B} \tag{23}$$

This is the reverse of the spin-orbit coupling. Different spin states lead to slightly different trajectories as observed in experiments of the Stern-Gerlach type. However, the reverse coupling is very weak at high energies: the spin-dependent kick produced by a quadrupole in

LEP is the same as the kick produced by the same quadrupole displaced by $4.5\ 10^{-13}$ mm! The magnetic energy of a spin state is indeed of the order of $ae\hbar/2m \approx 10^{-13}$ MeV/Tesla, much smaller than the kinetic energy. Another tiny effect arises from the slight dependence of the synchrotron radiation emitted by electrons on their spin state. The beam emittance should be unsignificantly modified.

2.4 Spin resonances

2.4.1 Resonant perturbation of the spin motion

When the spin-orbit coupling is not so strong as to depolarize the beam completely, the dependence of the polarization level on the spin tune is observed to be remarkable (e.g. Fig. 23). The polarization vanishes or is minimum at several distinct places where the spin tune is found to fulfill the relationship:

$$\nu = k + k_x Q_x + k_z Q_z + k_s Q_s \tag{24}$$

where Q_x, Q_z and Q_s are respectively the horizontal, vertical betatron tunes and the synchrotron tune, and the k's are signed integers. This is a signature of coupling resonances between the spin degree of freedom and the three orbit degrees of freedom. They are called spin resonances.

Whenever the trajectory is not ideal, i.e. does not go through the magnets on their magnetic axis, a perturbing magnetic field $\mathbf{b}(s)$ is experienced. Its component $\mathbf{b}_\perp$, perpendicular to the main precession axis, bends the spin vector **S** of the oscillating particle away from it. This field perturbation may be static, if due to an imperfect closed orbit, or oscillating with the betatron and synchrotron motions. The perturbation of the spin motion is superimposed on the nominal precession in the guide field. Depending on the phase relationship between spin precession and perturbation, a resonant spiraling motion can occur, generally causing a fast depolarization (Fig.5).

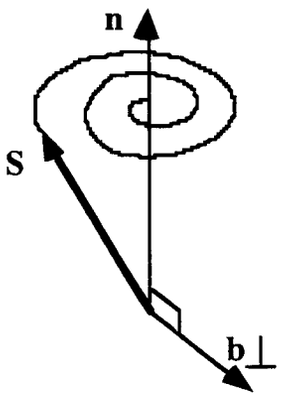


Fig. 5: On a spin resonance the spin vector **S**, bent away from the spin closed solution **n** by a perpendicular perturbing magnetic field $\mathbf{b}_\perp$, spirals out about **n**.

The resonance phenomenon is best understood when using a simple model of a single spin resonance. For that purpose, we specify a perturbing field $\mathbf{b}_\perp(s)$, orthogonal to the unperturbed spin closed solution $\mathbf{n}(s)$, which is assumed to rotate about $\mathbf{n}(s)$ at the frequency ν_r. If ν_r significantly differs from the spin precession frequency (assumed to be uniform), the phase difference between the motions of a spin **S** and of $\mathbf{b}_\perp(s)$ linearly increases with time. The successive kicks on the spin vector **S** point to different directions and average to zero in a finite time. In spite of the perturbing field there is no significant depolarization. On the contrary, when ν_r and ν coincide, the phase difference between the motions of **S** and of $\mathbf{b}_\perp(s)$ is stationary and the kicks add up constructively (Fig. 6). In addition to its precession about **n**, the spin vector **S** is then subject to an orthogonal rotation about $\mathbf{b}_\perp(s)$ at the frequency

$$\varepsilon = (1+\gamma\, a)\,\frac{b_\perp}{B_0} \tag{25}$$

in units of the revolution frequency Ω_c, where B_0 is the field of the nominal bending magnets. The combination of these two rotations gives the spiraling motion of the spin vector.

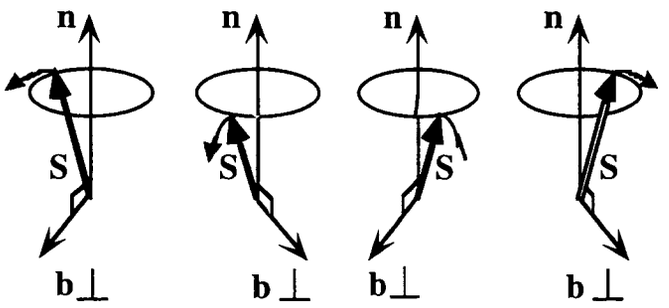


Fig. 6: The directions of a spin vector **S** and of the perturbing field $\mathbf{b}_\perp$ at different times on a spin resonance. The dashed arrows show the directions of the kicks given to **S**, all of them bending it away from **n**.

A better way to study spin resonances is to look at the spin motion in the $(\mathbf{l},\mathbf{m},\mathbf{n})$ frame which rotates about **n** at the frequency ν (section 2.3.2). In this frame, the perturbing field $\mathbf{b}_\perp$ rotates at the frequency $\nu_r - \nu$. On top of the resonance ($\nu_r = \nu$), it becomes stationary and the successive kicks given to a spin **S** add up constructively and lead to a continuous spin rotation about $\mathbf{b}_\perp$. This rotation periodically exchanges the two spin states "up" and "down". Such a spin resonance is exactly what happens in a Nuclear Magnetic Resonance experiment. The polarized beam is the analogue of a substance with spin-aligned nuclei. The field of the normal bending magnets is the analogue of the DC magnetic field applied to the magnetized substance. The perturbing field $\mathbf{b}_\perp$ plays the role of the applied RF field that resonates with the spin precession in a NMR experiment.

The frequency ε in (25) measures the speed at which the spin vector is bent away from the direction of **n**. It is the strength of the resonance. It also measures the width of the resonance, as the tilt of **S** is negligible when the frequency difference $\nu_r - \nu$ becomes larger than ε. In practical cases the perturbing fields can be expanded in sums of rotating components. Because the perturbing fields are caused by the trajectory displacements due to the deviations of the periodic orbit and to the betatron and synchrotron oscillations, the field

components rotate at frequencies which are linear combinations of the revolution, betatron and synchrotron frequencies. Likewise, the spin precession may be expanded in rotating components. A spin resonance occurs whenever a spin and field component rotate at the same frequency, as expressed by (24). The strength of the corresponding spin resonance is obtained by extracting from the perturbing field $\mathbf{b}_\perp$ the component rotating at the resonant frequency ν_r. It is convenient to represent the rotation vector about $\mathbf{b}_\perp$ in the ($\mathbf{l}$, $\mathbf{m}$) plane perpendicular to $\mathbf{n}$ by a complex number ω:

$$\omega = (1 + \gamma a)\,\frac{(\mathbf{m} + i\mathbf{l}).\mathbf{b}}{B_0} \tag{26}$$

following (25). The ν_r rotating component is given by the corresponding Fourier component of ω:

$$\omega_r = \frac{1+\gamma a}{2\pi B_0 \rho}\oint (\mathbf{m} + i\mathbf{l}) \cdot \mathbf{b}\; e^{2i\pi(\nu-\nu_r)s/R} ds \tag{27}$$

where the integration extends over one turn (the $e^{2i\pi\nu}$ phase factor compensates the phase advance of $\mathbf{m} + i\mathbf{l}$). The strength ε is the modulus of the Fourier component ω_r.

Depolarization occurs on a spin resonance when the perturbing field $\mathbf{b}_\perp$ is different for each particle in the beam. The phase of $\mathbf{b}_\perp$ may vary because the phases of the betatron and synchrotron oscillations are randomly distributed in the beam or because particles with different energies must cross the resonance condition at a different time. The amplitude of $\mathbf{b}_\perp$ varies with the oscillation amplitudes. The individual spin directions become scattered in space, leading to a decrease of the beam polarization. The depolarization increases with the resonance strength ε, which itself increases with energy (25). To illustrate the sensitivity of the high energy machines, we show on Fig. 7 the calculated depolarization in an ideal LEP at 45 GeV, due to the horizontal component of the earth magnetic field ($0.2\ 10^{-4}$ Tesla) which only perturbs the beam in the straight-sections. The spin resonances are so large as to require compensation.

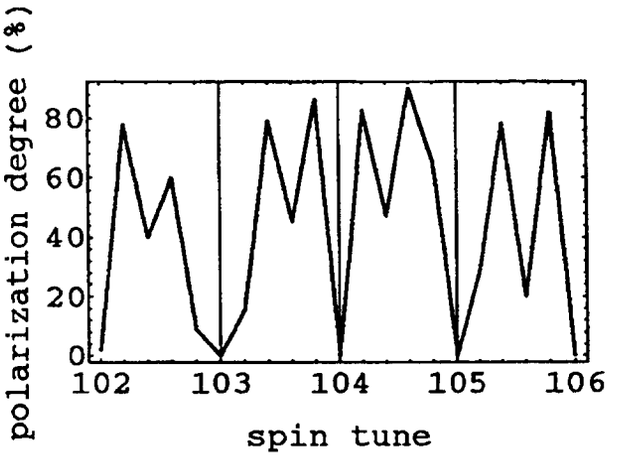


Fig. 7: Calculated depolarization by the uncorrected earth magnetic field in an otherwise ideal LEP.

At high energy, near-by resonances may overlap leading to a more complex polarization behaviour.

2.4.2 Classification of spin resonances

The condition for spin resonances (24) defines a dense web and one might think that one cannot avoid depolarizing resonances. However, calculations as well as measurements show that there is a hierarchy of spin resonances. Depending on the type of spin-orbit coupling, the beam energy or peculiarities of the accelerator, their depolarizing effect can be vastly different. The aim of a classification is to group resonances sharing common properties. The classification follows from the perturbing fields, symmetry or dynamics:

- the dipole fields drive the *integer* resonances $\nu = k$,
- the quadrupole fields drive the *linear* resonances $\nu = k + k_x Q_x + k_z Q_z + k_s Q_s$, with $|k_x| + |k_z| + |k_s| = 1$,
- the multipole fields drive the *non-linear* resonances where $|k_x| + |k_z| > 1$,
- the modulation of the field components by a finite-amplitude synchrotron motion drives *higher-order* synchrotron and synchro-betatron spin resonances with $|k_x| + |k_z| + |k_s| > 1, \quad k_s \neq 0$.

The solenoidal fields drive both integer and linear resonances. Each type of spin resonance is repeated when the integer k is increased by one unit, i. e. every 523 MeV for protons and every 440 MeV for electrons. The various families of resonances are analysed in the following sections.

2.4.3 Integer spin resonances

The resonance condition of the integer (also named imperfection) resonances is

$$\nu = k \tag{28}$$

They are mainly produced by vertical closed orbit distortions when the spin closed solution **n** is not radial and possibly by horizontal closed orbit distortions when **n** is not vertical. These orbit distortions are caused by field errors and tilts of the bending magnets and by misalignment of quadrupoles and sextupoles. Systematic imperfections excite more selectively the resonances $\nu = kP$, where P is the machine superperiod. However, the dominant modulation of the resonance strength is related to the spectrum of the closed orbit. Random imperfections unexpectedly drive systematic integer spin resonances: the orbit spectrum is peaked at the betatron tune $Q_{x/z}$ unless the closed orbit is very carefully corrected. It is then possible to show that the resonances

$$\nu = \text{int}(Q_{x/z}) + k \cdot P_B \tag{29}$$

with P_B superperiod of the *bending* [36] are selectively excited and should be avoided.

On top of the resonance the perturbing dipole field perpendicular to the unpertubed **n** is exactly stationary in the (**l**,**m**,**n**) frame. A spin aligned along the perturbing field is also stationary. It is a particular solution of the perturbed spin motion, periodic in the laboratory frame. As it lies in the plane perpendicular to the unperturbed **n**, the beam polarization will be periodically exchanged between the up and down states. Because a dipole field acts in the same way on all spins, there should be no depolarization. However side-effects may prevent the conservation of the length of the polarization vector. Due to the beam energy spread, all the particles in a beam do not cross an integer resonance at the same time during energy ramping; this causes a spread in precession phase and some depolarization. More important is the interplay between integer and linear resonances. The integer resonance is a resonant tilt of the **n**-axis. This tilt in turn excites near-by linear resonances and causes them to overlap the integer resonances. In the high-energy electron machines, this mechanism causes strong depolarization. This is the reason why the compensation of the integer spin resonances is essential.

2.4.4 Linear spin resonances

These occur when only one of the k_x, k_z and k_s in (24) does not vanish and equals ± 1. They are mainly produced by transverse quadrupole fields :

$$\mathbf{b} = \mathbf{x}\frac{\partial B_x}{\partial z} z + \mathbf{z}\frac{\partial B_z}{\partial x} x \tag{30}$$

where x and z are the trajectory displacements along the radial **x** and vertical **z** unit vector respectively. One distinguishes three types of linear spin resonances, corresponding to the coupling of the spin precession with each of the three degrees of freedom of the orbital motion.

1. *The vertical betatron resonances* (also named intrinsic resonances)

$$\nu = k \pm Q_z \tag{31}$$

are driven by vertical betatron oscillations

$$z = a_z \sqrt{\beta_z} \cos \phi_z \tag{32}$$

whenever the spin closed solution **n** is not pointing to the radial direction **x**. That is the general case as **n** is nearly vertical in horizontal rings. One recognizes that the Fourier spectrum of $\mathbf{b}_\perp$ is made of lines : $(k_o \pm Q_z)\, \Omega_c$, leading to the formula (31). Moreover, the integer k is a multiple of the ring superperiodicity. However, in real rings, small defects generally break the superperiodicity and weaker spin resonances appear at integer values k not multiples of the superperiodicity.

2. *The horizontal betatron resonances*

$$\nu = k \pm Q_x \tag{33}$$

are driven by horizontal betatron oscillations

$$x = a_z \sqrt{\beta_x} \cos \phi_x \tag{34}$$

whenever the spin closed solution is not exactly vertical. In general, this is due to small imperfections of the ring. It also occurs in a ring equipped with spin rotators which bends **n** along the longitudinal direction in some sections of the ring. A horizontal betatron spin resonance can also result from an x-z coupling

$$z = \varepsilon \cos (Q_x \theta + \phi_x) \tag{35}$$

when **n** is not radial.

3. *The synchrotron resonances*

$$\nu = k \pm Q_s \tag{36}$$

are driven by horizontal and vertical synchrotron oscillations :

$$\begin{aligned} x &= D_x \frac{\delta P}{P_o} \cos (Q_s \theta + \phi_s) \\ z &= D_z \frac{\delta P}{P_o} \cos (Q_s \theta + \phi_s) \end{aligned} \tag{37}$$

They are also due to imperfections as the horizontal oscillations only contribute when **n** is not exactly vertical and the vertical oscillations only when the vertical dispersion D_z does not vanish. Because the synchrotron tune Q_s is often small, these resonances are referred to as synchrotron satellites. When the Q_s is very small ($Q_s \ll 1$), one usually does not disentangle the integer resonance from its satellites.

The strength of *linear* resonances can be expressed following (27) in an explicit way which is useful for resonance compensation. The strength of the *integer* resonance is hidden behind the calculation of the perturbed $(\mathbf{l},\mathbf{m},\mathbf{n})$ frame which can only be performed

numerically. Assuming this frame calculated, the strengths ε of the three types of *linear* spin resonances are respectively given by :

$$\begin{aligned}
\varepsilon_{\pm z} &= a_z (1 + \gamma a) \left| \oint (\mathbf{m} + i\mathbf{l}) \cdot \mathbf{x}\, K\sqrt{\beta_z}\, e^{2i\pi(\nu \pm Q_z)s/R \pm i\phi_z}\, ds \right| \\
\varepsilon_{\pm x} &= a_x (1 + \gamma a) \left| \oint (\mathbf{m} + i\mathbf{l}) \cdot \mathbf{z}\, K\sqrt{\beta_x}\, e^{2i\pi(\nu \pm Q_x)s/R \pm i\phi_x}\, ds \right| \\
\varepsilon_{\pm s} &= \frac{\delta P}{P_0} (1 + \gamma a) \left| \oint (\mathbf{m} + i\mathbf{l}) \cdot K(D_x \mathbf{z} + D_z \mathbf{x}) e^{2i\pi(\nu \pm Q_s)s/R \pm i\phi_s}\, ds \right|
\end{aligned} \tag{38}$$

K is the normalized gradient of the quadrupoles. The three integrals, named *spin-orbit coupling integrals*, express the amplitude of the spin-orbit coupling for each linear resonance. The strength is also proportional to the respective amplitude of oscillation a_x, a_z and $\delta P / P_0$ and to the energy. For a distribution of particles, it must be averaged over the amplitude distribution, and also over the relative phases of oscillations ϕ_x, ϕ_z, ϕ_s. It is worth noting the absence of parametric resonances $k + 1/2$. Therefore, the best energy to maintain a high degree of polarization in a storage ring normally corresponds to a half-integer spin tune, that is midway between linear spin resonances. It is also advantageous that the nearest integers should not be multiples of the ring periodicities.

2.4.5 Non-linear spin resonances

Non-linear spin resonances are produced by higher-order multipole fields :

$$\mathbf{b} \propto x^p z^q, \qquad p + q > 1 \tag{39}$$

The frequency analysis of $\mathbf{b}$ leads to a resonant condition of the type :

$$\nu = k_o + k_x Q_x + k_z Q_z \tag{40}$$

with $|k_x| \leq p$ and $|k_y| \leq q$. For instance a sextupole field will drive non-linear resonances with $|k_x| + |k_y| = 2$. The beam-beam interaction in storage rings will also drive series of non-linear resonances.

2.4.6 Higher-order spin resonances

Another family of resonances plays an important role at high energy when the amplitude of the synchrotron oscillations becomes large. In contrast with the preceding resonances, the perturbation of the spin motion is due to the component of the perturbing field along $\mathbf{n}$, mainly produced by the horizontal synchrotron oscillations. It causes a perturbation of the spin precession frequency ω that can be analysed as a modulation of the spin tune at the frequency of the synchrotron oscillations Q_s.

$$\nu_0 - k_s Q_s \qquad |k_s| \geq 1 \tag{41}$$

Several synchrotron satellites appear on each side of any depolarization resonance ($\nu = \nu_r$) of the preceding types:

$$\nu = \nu_r + k_s Q_s \qquad |k_s| \geq 1 \tag{42}$$

This modulation is proportional to the energy spread and increases rapidly with the particle energy.

3. ACCELERATION OF POLARIZED PROTONS IN SYNCHROTRONS

3.1 Experimental observations and results

Polarized protons have been successfully accelerated at high energies (larger than one GeV) in several synchrotrons : Saturne II [13,14] at Saclay (up to 3 GeV), the KEK PS [37] at Tsukuba (up to 7 GeV), the ZGS [12] at Argonne (up to 12 GeV) and the AGS [15] at Brookhaven (up to 22 GeV). The most successful one is Saturne where proton and also deuteron polarized beams are routinely accelerated and used by the physicists. The highest energy (22 GeV) has been reached at the AGS and seems to be the limit of the present technology, beyond which new provisions are needed to counteract large depolarization due to the crossing of numerous and strong spin resonances. One generally considers the use of "Siberian snakes" which, in an experiment [38,39] at the Indiana University Cyclotron Facility (IUCF) Cooler Ring, have been shown to be efficient in suppressing depolarization on spin resonances.

At Saturne II, 2.10^{11} particles per burst, either protons or deuterons, delivered by a polarized source with 90% polarization, are first injected and accelerated in the Mimas booster up to 47 MeV. Then they are extracted and injected in the main ring where they are accelerated up to the top energy. At injection in the rings the beam polarization is vertical. During acceleration the spin tune increases with the energy linearly. (See Eq. (18)). Spin resonances are crossed when the spin tune fulfills one of the conditions given by the Eq. (24). In the booster the vertical betatron tune Q_z must be set at a high value (2.48) to avoid the intrinsic resonance $\nu = 4 - Q_z$ even in the presence of a large tune shift due to space charge effect at injection. In the main ring, depending on the extraction energy, at most 15 resonances are crossed (Fig. 8). Special provisions are made to avoid depolarization at the resonance crossings, as described below. The net effect is only a spin reversal for some of these resonances. At the maximum energy of 3 GeV the overall depolarization during the complete cycle of acceleration is less than 15% as measured by a high-energy polarimeter.

At the AGS (Fig. 9), a polarized source has been used to deliver a 25 μA H^- beam with 75% polarization. This beam was first accelerated in a RFQ, then in a linac up to 200 MeV. After electron stripping through a thin carbon foil, the polarized proton beam was injected in the main ring. At the maximum energy of 22 GeV the proton beam ($2x10^{10}$ ppp) was extracted and transported to the experimental area to bombard a polarized proton target in an experiment to measure the spin dependence of the proton-proton interaction.

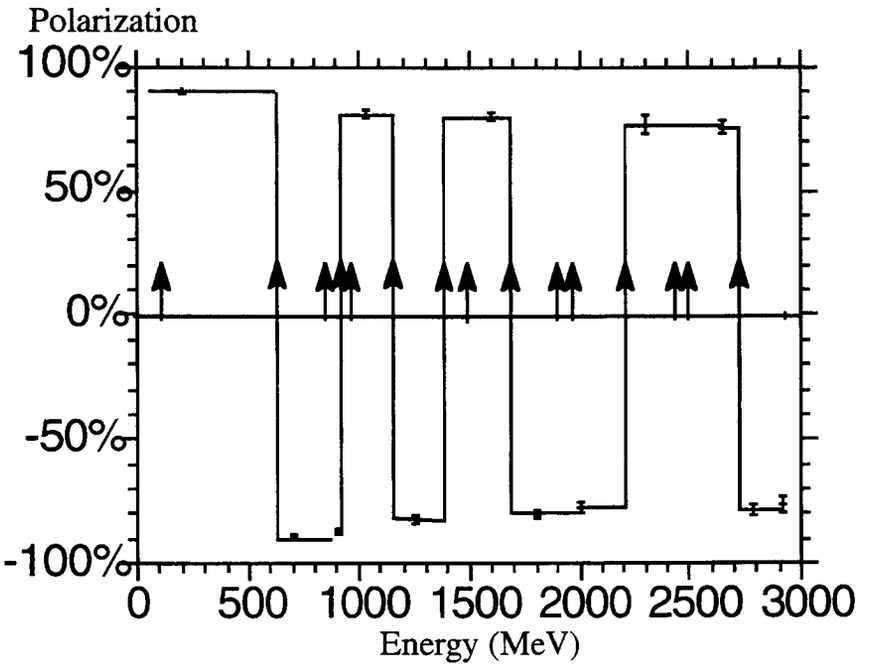


Fig. 8: The vertical polarization as function of the proton kinetic energy during the acceleration cycle in Saturne II [13,14]. The points with error bars are the experimental data. The arrows show the locations of the spin resonances. The solid line is a hand-drawn line showing the spin flips at the crossings of seven resonances.

The polarization was measured by polarimeters at several stages of the acceleration process : downstream of the linac before injection in the ring, inside the ring, and after extraction. During the acceleration cycle, one had to overcome 45 strong spin resonances to maintain the polarization up to 22 GeV, making the commissioning of the proton polarized beam a painstaking effort. The depolarization on the resonances is minimized using 95 dipole correctors and 12 pulsed quadrupoles. The obtained maximum polarization was 45% above a beam momentum of 16 GeV/c (Fig. 10).

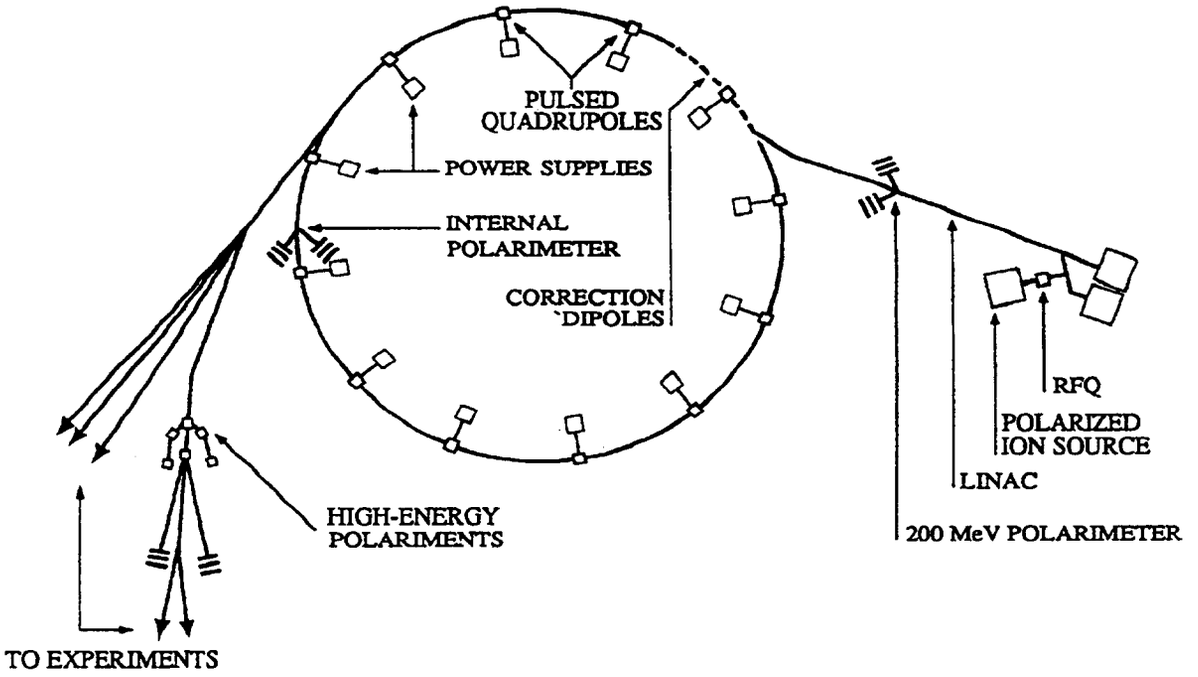


Fig. 9: The AGS [15] lay-out for the acceleration of polarized protons.

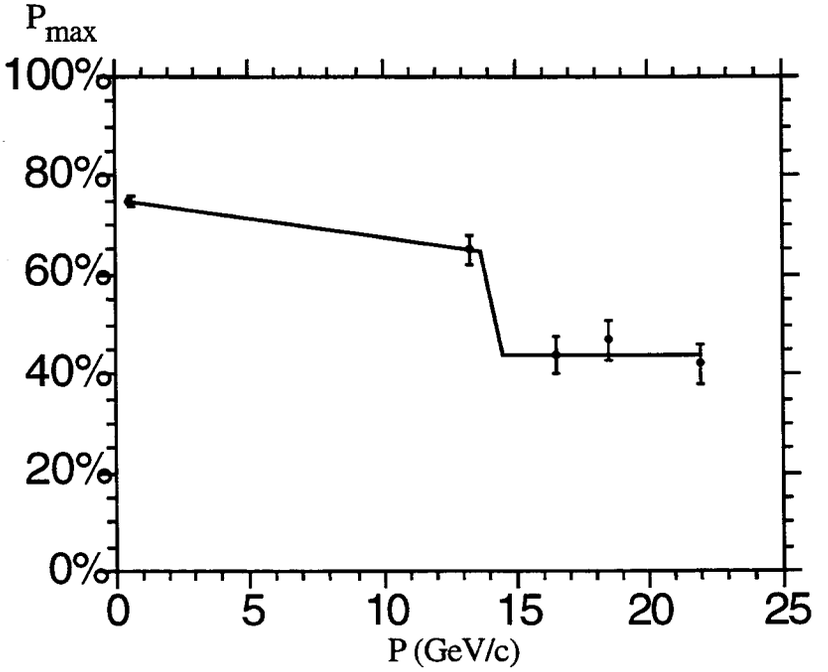


Fig. 10: The maximum AGS beam polarization [15] as function of the beam momentum showing a sharp polarization loss near 14 GeV/c.

3.2 Resonance crossing in proton synchrotrons

As the beam polarization must be maintained vertical to approach spin transparency, only the projection P_z of the polarization vector along the vertical line is considered. Any depolarization is measured as a decrease of the vertical projection. The spin resonances responsible for the most important vertical depolarization are of two types :

- The vertical betatron or intrinsic resonances (section 2.4.4.1) at spin tune values:

$$\nu = kp \pm Q_z$$

  where p is the superperiodicity of the ring assumed to be perfect and k is any integer. When taking into account the ring imperfections, the superperiodicity is broken by the defects. Weaker resonances appear where kp becomes any integer.

- The synchrotron or imperfection resonances (section 2.4.4.3) at spin tune values:

$$\nu = k \pm Q_s$$

In most of the synchrotrons the synchrotron tune Q_s is much less than one. In the spectrum of spin resonances there are two synchrotron lines on both sides of each integer k. The most harmful imperfection resonances are those for which $k = np \pm m$ where n is any integer and m is an integer close to Q_z.

For imperfection resonances, the resonance strength ε scales linearly with the amplitude of the vertical closed-orbit and with the total proton energy (see Eq. 38). It lies in the 10^{-3}-10^{-5} range at the AGS. For the intrinsic resonances the strength scales linearly with the energy and with the square root of the vertical emittance, lying in the 10^{-3}-10^{-2} range.

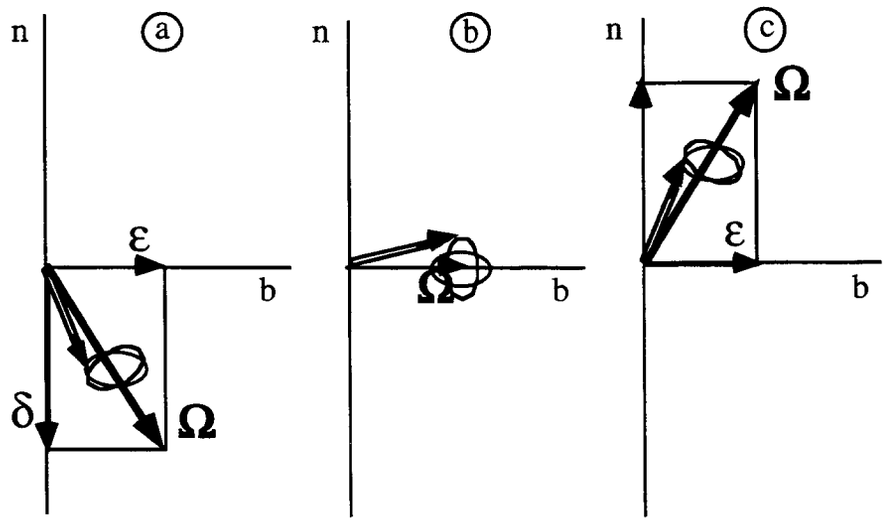


Fig. 11: The spin rotation as seen in the rotating frame near an isolated resonance. The rotation vector is $\mathbf{\Omega}$ with components ε along the driving field **b** and δ along the direction **n**:

a) below the resonance ($\delta < 0$),
b) on top of the resonance ($\delta = 0$),
c) above the resonance ($\delta > 0$).

During the acceleration cycle the spin tune ν increases monotonically and successively crosses several spin resonances. The width of the resonances, also measured by ε, is usually smaller than the resonance spacing. The resonances can be considered as if they would be isolated during their crossing. As in the general study of spin resonances (Section 2.4.1) the best way to study the spin perturbation when crossing a resonance is to consider a rotating frame. The only difference with the $(\mathbf{l},\mathbf{m},\mathbf{n})$ frame, used in Section 2.4.1, is that here this frame rotates about $\mathbf{m}$ at the frequency ν_r of the resonant perturbing field instead of the frequency ν of the spin precession. With respect to this frame the perturbing field is at rest. The spin vector $\mathbf{S}$ rotates about $\mathbf{n}$ at the frequency $\delta = \nu - \nu_r$ and about $\mathbf{b}$ at the frequency ε. Its overall rotation vector Ω has components δ along $\mathbf{n}$ and ε along $\mathbf{b}$ (see Fig. 11). Far below the resonance, δ is negative and much larger than ε ; Ω is nearly antiparallel to $\mathbf{n}$. When approaching the resonance, Ω starts to move into the direction of $\mathbf{b}$. On top of the resonance, Ω is exactly along $\mathbf{b}$. Above the resonance, Ω now moves into the direction of $\mathbf{n}$ and becomes nearly parallel to $\mathbf{n}$ far above the resonance. The resonance crossing leads to a complete reversal of the rotation vector Ω.

The speed of the resonance crossing is measured by the rate α of the spin tune variation:

$$\nu = \nu_r + \alpha\theta \tag{43}$$

where θ is the azimuthal angle of the particle along the ring circumference. The "time" $\Delta\theta$ of the crossing is the width ε of the resonance divided by the rate α : $\Delta\theta = \varepsilon/\alpha$. During that time the rotation angle of the spin about $\mathbf{b}$ is : $\psi = \varepsilon\Delta\theta = \varepsilon^2/\alpha$. There are two extreme regimes of resonance crossing :

- Fast crossing ($\psi << 1$) : either the resonance is too narrow or the crossing rate is very large. There is not enough time to bend the spin vector away from the direction $\mathbf{n}$ and there is no depolarization.

- Slow crossing ($\psi >> 1$) : either the resonance is very broad or the crossing rate is very slow. The spin vector rotates much faster about the rotation vector Ω than the latter moves. The spin vector adiabatically follows the vector Ω in its motion and undergoes a complete reversal as Ω. After the crossing the vertical polarization has changed sign, but there is no depolarization.

Between these two extreme regimes of crossing there is a partial spin flip. The decrease of the vertical component S_z of the spin vector is quantitatively given by the Froissart-Stora Formula [4]:

$$\frac{S_z^{\text{final}}}{S_z^{\text{initial}}} = 2\exp\left(-\frac{\pi\varepsilon^2}{2\alpha}\right) - 1 \tag{44}$$

that also includes the two extreme regimes of fast crossing and adiabatic spin flip. The width ε of a resonance depends on the oscillation amplitude of the considered particle. A large

amplitude may lead to complete spin reversal. On the other hand, very small amplitudes do not cause significant change of S_z. To obtain the decrease of the vertical beam polarization P_z, i.e. the depolarization, one must average the Froissart-Stora formula over the particle amplitudes.

The adiabatic spin flip has been observed at Saturne at the crossing of five imperfection resonances and of two intrinsic resonances. Figure 12 shows detailed measurements [40] of the vertical polarization showing the spin flip when the beam is slowly extracted in the vicinity of the imperfection resonance $\gamma a = 3$.

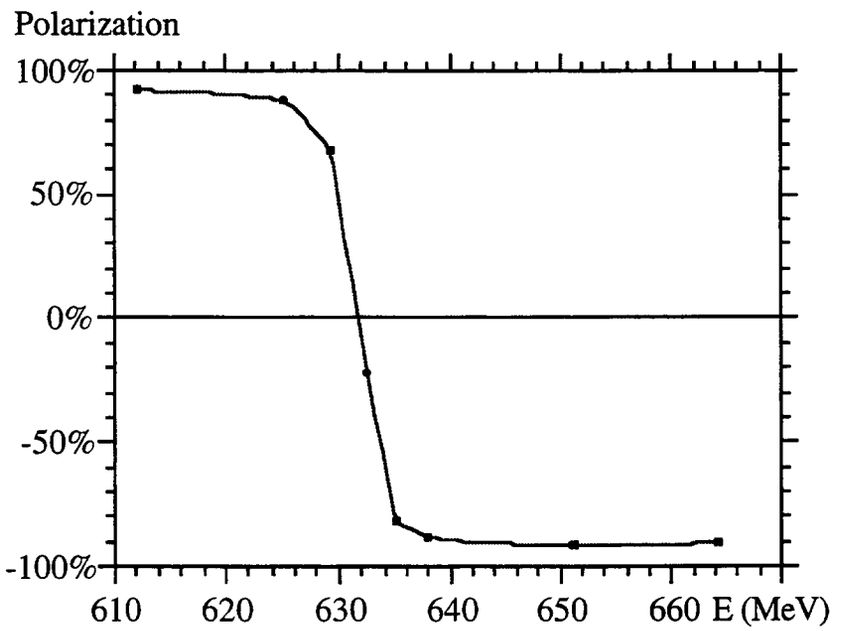


Fig.12: The vertical polarization P_Z as function of the kinetic proton energy E, measured after slow extraction in the close vicinity of the imperfection resonance $\gamma\ a = 3$ (E = 632 MeV) at Saturne II [40].

Two different methods have been applied to avoid depolarization at the crossing of resonances that are not strong enough to obtain a complete spin flip :

- Compensation of the resonance strength ε (named also harmonic spin matching) : with some dipole correctors one can control the harmonics of the vertical closed-orbit distortion that drive an imperfection resonance. After a careful closed-orbit correction which minimizes the deviations of **n** from the vertical, one maximizes the polarization, measured with a polarimeter, by adjusting the dipole correctors. This method has been successfully applied [15] to compensate about 35 imperfection resonances at the AGS using 95 correctors (Fig. 13).

- The same method using dipole correctors is also applied at Saturne II [13,14] to compensate the imperfection resonance $\gamma a = 2$. Similarly some intrinsic resonances are compensated using quadrupolar correctors.

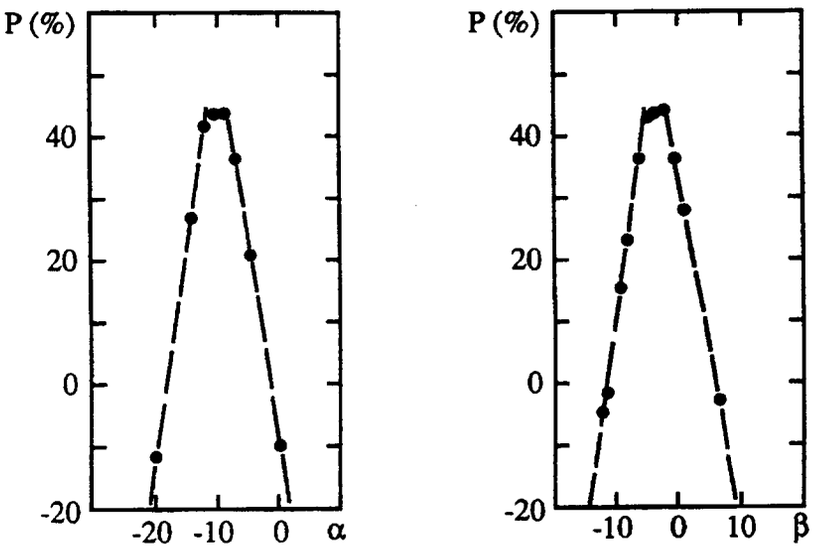


Fig. 13: The vertical polarization as function of the amplitudes α and β of the sine and cosine (resp.) harmonics that compensates the strong imperfection resonance $\gamma\, a = 9$ at the AGS [15].

- Speed increase of resonance crossing (also named resonance jumping) : when approaching an intrinsic resonance during acceleration, the vertical betatron tune is abruptly varied such that the resonance is crossed more rapidly (Fig 14). Thereafter the initial tune is slowly restored. At the AGS ten pulsed quadrupoles have been used to jump seven intrinsic resonances. The betatron tune was lowered by 0.25 with a 1.6 μs risetime and a 20 μs falltime.

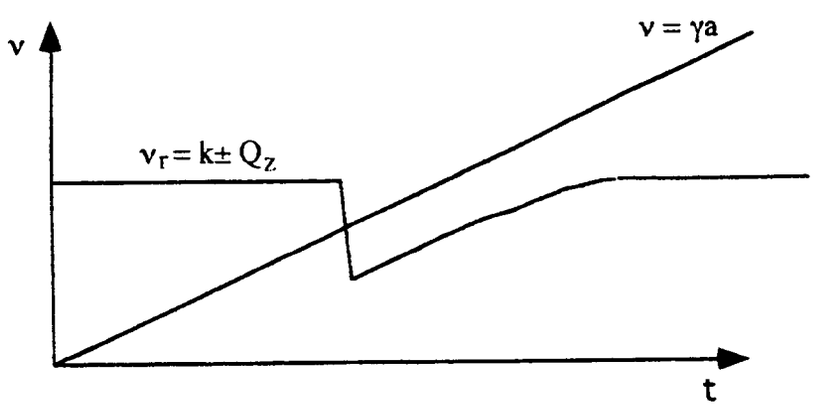


Fig. 14: Variation with time of the spin tune ν and of the tune ν_r of an intrinsic resonance, when the vertical betatron tune Q_Z is rapidly shifted.

3.3 Siberian snakes

The number and the strengths of the spin resonances, crossed during the acceleration cycle, increase with energy. To accelerate polarized protons at the AGS up to 22 GeV a considerable effort has been required to overcome depolarization with the preceding methods when crossing 45 strong spin resonances. It would become unpractical at higher energies. A long time ago Ya. Derbenev and A. Kondatrenko [33,34] proposed to equip rings with special magnet arrangements, later named "Siberian snakes", that would avoid the crossing of all the resonances. With them the fractional spin tune becomes equal to 1/2 and independent of energy. During acceleration the spin tune stays constant and never crosses tune values of spin resonances.

In principle a Siberian snake rotates any spin vector **S** by π about an axis **u** lying in the horizontal plane of the ring. In a ring equipped with one Siberian snake, the spin closed solution $\mathbf{n}(s)$ lies in the horizontal plane, apart inside the magnets of the Siberian snake. Figure 15 shows the motion of $\mathbf{n}(s)$ and of a spin vector **S**. At the point 0, opposite to the snake, $\mathbf{n}(0)$ is parallel to the axis **u**. Its direction $\mathbf{n}(\pi R)$ at the snake entrance is rotated into $\mathbf{n}'(\pi R)$ at the exit. The motion of $\mathbf{n}(s)$ before and after the snake are symmetric, such that $\mathbf{n}(s)$ again becomes parallel to the axis **u** after one turn. The motion of another spin vector **S** is also symmetric before and after passing through the snake. In particular at the point 0 its directions at the beginning and at the end of one turn are symmetric. Therefore they are connected by a rotation about the axis **u** showing that the spin tune is effectively 1/2 (the spin tune is the rotation angle in units of 2π.).

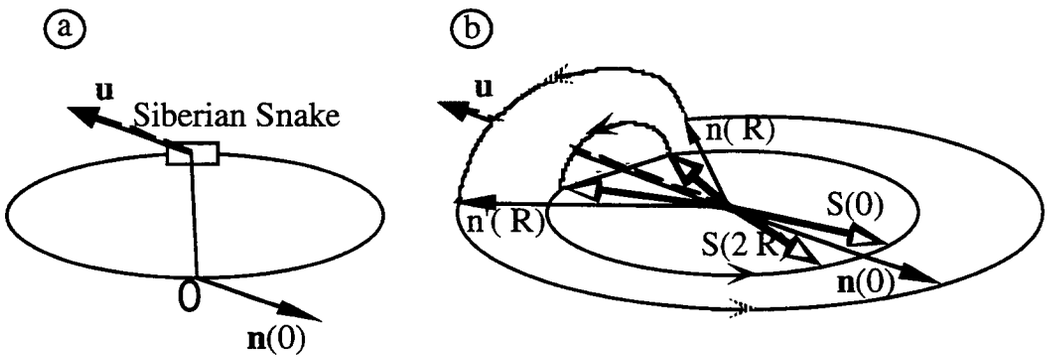


Fig. 15: a) A ring equipped with a Siberian snake located at $s = \pi R$. Its axis **u** lies in the horizontal plane. The spin closed solution $\mathbf{n}(0)$ is parallel to **u** at the origin 0, opposite to the snake.

b) Horizontal precession of the spin closed solution $\mathbf{n}(s)$ and of a spin vector **S** lying in the ring plane. $\mathbf{n}(\pi R)$ and $\mathbf{n}'(\pi R)$ are the directions of $\mathbf{n}(s)$ at the entrance and at the exit of the snake respectively. $\mathbf{S}(0)$ and $\mathbf{S}(2\pi R)$ are the directions of the spin vector at the beginning and at the end of one turn respectively. The circles show the tracks of these vectors in the horizontal plane and about the snake axis **u**.

The polarization behavior in a ring equipped with a snake is similar to the spin echo phenomenon used in NMR techniques. In a simple spin echo experiment a nuclear magnetic substance is magnetized and its magnetization vector **M** precesses about a stationary magnetic field $\mathbf{B}_0$ perpendicular to **M** (Fig. 16). Due to local field inhomogeneities, the magnetic moments **u** of different nuclei precess at slightly different frequencies. If they were initially aligned in the same direction, they rapidly spread out and magnetization decreases. At time T a transient field is applied which rotates all the magnetic moments μ by π about the axis **u**. The fastest moments which were in advance before the time T become the latest after that π rotation. At time $2T$ all the magnetic moments are again aligned together and the initial magnetization is restored. The spin vectors of the particles in a polarized beam are similar to the nuclear magnetic moments. They precess about the magnetic field of the ring bending magnets. The polarization vector **P** is the analogue of the magnetization **M**. The snake plays the role of the transient field and the spin-orbit coupling plays the role of the field inhomogeneities. The polarization decreases in the half turn preceding the snake and is restored at the end of the following half turn.

Generally one considers two kinds of Siberian snakes [34,41]. A snake of type I rotates the spin by π about an axis parallel to the reference orbit. At low momentum p (Gev/c) it can be made of a single solenoid with a field integral of $3.752xp$ (Tm) according to the Formula (13). A snake of type II rotates the spin by π about an axis perpendicular to the reference orbit. At high energies all kinds of snake would be made by sandwiches of several horizontally and vertically bending magnets, eventually combined into helical magnets.

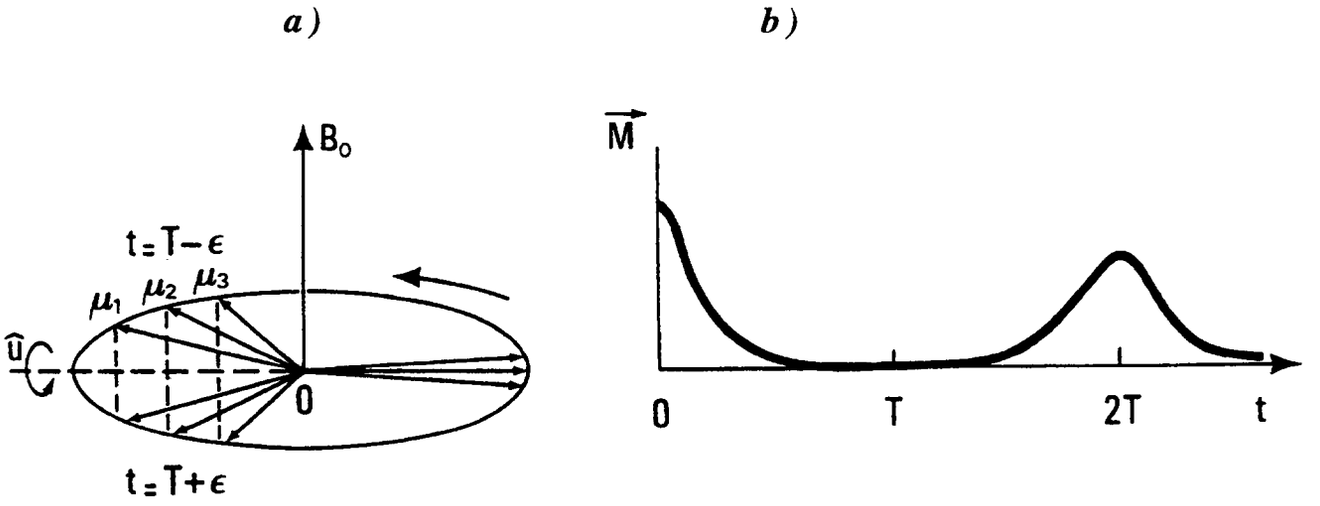


Fig. 16: Scheme of a NMR spin echo experiment :

a) Precession of three magnetic moments $\mu_{1,2,3}$ about the magnetic field $\mathbf{B}_0$ with π rotation at time T. (Positions at times $T-\varepsilon$ and $T+\varepsilon$ are shown).
b) The variation of the magnetization **M** with time t.

The first experimental test [38,39] of a Siberian snake to overcome depolarizing resonances was made at the Indiana University Cyclotron Facility (IUCF) Cooler Ring. A type-I Siberian snake, made of a superconductive solenoid and correcting quadrupoles, was

installed in the Cooler Ring. Other solenoids were used to produce a variable longitudinal field operating as a controlled defect driving imperfection resonances. Proton beams were injected and stacked in the ring with either vertical or horizontal polarization to match the direction of the spin closed solution **n** (vertical when the snake is off and horizontal when the snake is on). An internal polarimeter measured the vertical and radial components of the polarization. Fig. 17 shows the polarization data in the vicinity of the imperfection resonance $\gamma a = 2$ and the intrinsic resonance $\gamma a = -3 + Q_z$. For the imperfection resonance the proton beam energy was fixed at 104 MeV (4 MeV below the top of the resonance) and the field integral of the imperfection solenoids was varied. With the snake off, the vertical polarization was consistent with zero except when the defect field integral was nearly vanishing. On the contrary, with the snake on, the measured radial polarization was independent of the defect and close to its expected maximum. For the intrinsic resonance the proton beam energy was held to 177 MeV and the vertical betatron tune Q_z was varied. With the snake off, the vertical polarization was found to fall to zero for tune values close to the resonance, while radial polarization was maintained at its maximum for all tune values when the snake was turned on. In another test the vertical tune was ramped in a way equivalent to an energy ramping across the intrinsic resonance. With the snake off, the resonance crossing led to 75% polarization loss, but no observable depolarization with the snake turned on, proving the ability of Siberian snakes to overcome depolarization during acceleration. This positive result opens the possibility to accelerate polarized proton beams at higher energies than the AGS.

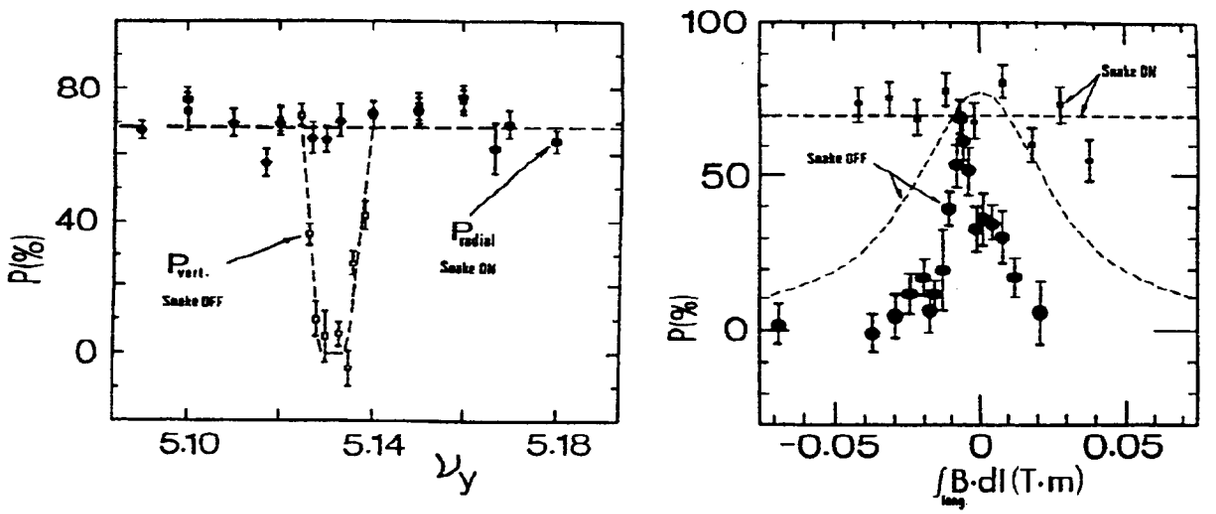


Fig. 17: Polarization data at the IUCF Cooler Ring [38,39] with a Siberian Snake. The vertical and the radial polarizations are measured when the snake is respectively turned off and on :
a) At 177 MeV near the intrinsic resonance $\gamma a = -3+Q_Z$, varying the vertical betatron tune Q_Z.
b) At 104 MeV near the imperfection resonance $\gamma a = 2$, varying the field integral of the imperfection solenoids.

4. POLARIZATION OF ELECTRONS IN STORAGE RINGS

The orbital dynamics of electron and proton beams in high-energy storage rings are very different. Because of their much smaller mass, the electrons, subject to the central acceleration of the guide field, radiate energy. The quantum fluctuations associated with the photon emissions introduce a noise in the orbital motion. Conversely, the replacement of the lost energy by the RF cavities and differential radiation effects cause a damping of the excited orbital motion. The combination of the two effects determines the electron beam emittance, while the emittance of a proton beam depends on its value at injection.

Similarly, the dynamics of polarization are totally different. A polarized source must be used to inject polarized protons and the final polarization is at most equal to its initial value. Electrons become transversely polarized by a quantum mechanical effect which occurs when synchrotron photons are emitted. Conversely, the noise which perturbs the electron dynamics causes a diffusion of the spins. These two processes compete and lead to an equilibrium polarization.

4.1 Spontaneous polarization of the electrons

4.1.1 Spin-dependent synchrotron radiation

Ternov, Lostukov and Korovina [6] discovered that the probability for an electron to emit a photon depends slightly on the initial spin state of the electron. Sokolov and Ternov [7], by solving the Dirac equation in an homogeneous magnetic field, worked out the complete formula giving the rate of photon emission for an electron with given initial and final spin states in the direction of the magnetic field.

$$w\left(s_i, s_f\right) = \left(\frac{5\sqrt{3}}{6}\frac{e^2}{\hbar c}\frac{c}{\rho}\frac{E}{m_0 c^2}\right)\left[\begin{array}{l}\left(\text{if } s_i = s_f\right)\left\{1 - \frac{16\sqrt{3}}{45}\xi_0 + \frac{25}{18}\xi_0^{\,2} - \frac{s_f}{5}\left(1 - \frac{20\sqrt{3}}{9}\xi_0\right)\xi_0\right\} \\ \left(\text{if } s_i \neq s_f\right)\left\{\frac{1}{6}\left(1 - s_f\frac{8\sqrt{3}}{15}\right)\xi_0^{\,2}\right\}\end{array}\right] \quad (45)$$

where ξ_0 is the ratio of the critical photon energy $\hbar\omega_c$ to the electron energy E. It is in general a very small value, e.g. 10^{-6} in LEP at 45 GeV. This formula shows that the vast majority of photon emissions does not involve spin flip. It is amusing to note that the beam radiates slightly less if it is polarized! The message of formula (45) is clarified by forming the asymmetries:

Spin flip versus non-spin flip events, whatever the spin state:

$$\frac{\left(w_{\uparrow\uparrow} + w_{\downarrow\downarrow}\right) - \left(w_{\uparrow\downarrow} + w_{\downarrow\uparrow}\right)}{\left(w_{\uparrow\uparrow} + w_{\downarrow\downarrow}\right) + \left(w_{\uparrow\downarrow} + w_{\downarrow\uparrow}\right)} = \frac{\xi_0^{\,2}}{3} \approx 10^{-12} \quad (46)$$

Dependence on the final spin state for spin flip events:

$$A = \frac{\left(w_{\uparrow\downarrow} - w_{\downarrow\uparrow}\right)}{\left(w_{\uparrow\downarrow} + w_{\downarrow\uparrow}\right)} = \frac{8}{5\sqrt{3}} \approx 0.924 \qquad (47)$$

The probability for a spin flip is indeed very small. In case of a spin flip however, the preference for the spin state which is antiparallel to the magnetic field is almost one. There is thus a mechanism which forces the electron spins to the same state. A stored positron beam becomes polarized as well as an electron beam. However, the direction of polarization is opposite, i.e. parallel to the magnetic field. It is worth noting that in both cases the direction of polarization corresponds to a minimum of the magnetic energy (3). However, it has been shown that it is more a coincidence than a real cause [42].

4.1.2 Build-up of the polarization

In the uniform magnetic field we consider, the depolarizing effects are absent if one neglects the very small electron recoil. The dynamics of the polarization may thus be calculated from (45). At any given time, the beam polarization and its time derivative are given by:

$$P = \frac{N_\uparrow - N_\downarrow}{N} \qquad \frac{dP}{dt} = \frac{1}{N}\left(\frac{dN_\uparrow}{dt} - \frac{dN_\downarrow}{dt}\right) \qquad (48,49)$$

where $N_\uparrow$ and $N_\downarrow$ denote the number of electrons in each spin state. Their rate of change may easily be deduced from the transition probabilities (45):

$$\frac{dN_\uparrow}{dt} = N_\downarrow w_{\downarrow\uparrow} - N_\uparrow w_{\uparrow\downarrow} = N w_{\downarrow\uparrow} - N_\uparrow\left(w_{\uparrow\downarrow} + w_{\downarrow\uparrow}\right) \qquad (50)$$

and a similar expression for $dN_\downarrow / dt$. Replacing in (49) and integrating yields an exponential law for the growth of the polarization:

$$P = A\left(1 - e^{-t/\tau_p}\right) \qquad (51)$$

The maximum degree of polarization is equal to the asymmetry A, i.e. 92.4%. Under very special circumstances, it has been shown [43] that the ultimate degree of polarization would reach 99.2% due to the contribution arising from the electron recoil in the photon field in the vicinity of a vertical betatron spin resonance. This enhancement of the polarization has not been observed so far.

The characteristic time τ_p of the polarization build-up is given by :

$$\tau_p^{-1} = \left(w_{\uparrow\downarrow} + w_{\downarrow\uparrow}\right) = \frac{5\sqrt{3}}{8} c\lambda\!\!\!^{-}_c r_e \frac{\gamma^5}{\rho^3} \tag{52}$$

where r_e is the electron classical radius, $\lambda\!\!\!^{-}_c$ is the Compton wavelength divided by 2π and ρ is the bending radius in the homogeneous magnetic field. The polarization time decreases very rapidly when the energy is increased. It is due to the very fast increase of radiation rate that counteracts the low probability of spin flip in photon emission (Fig. 18).

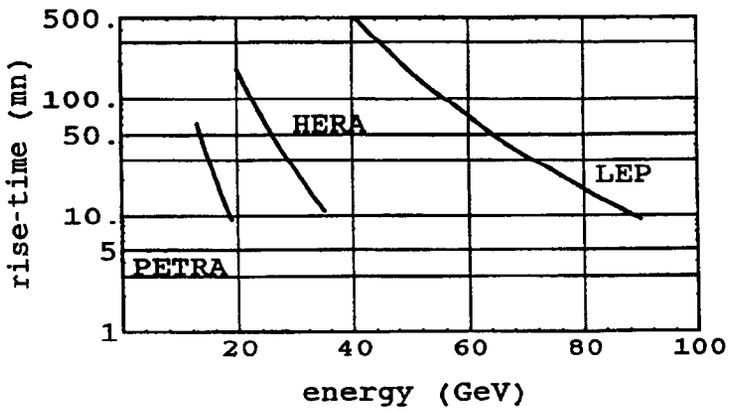


Fig. 18: Calculated polarization time versus energy for PETRA, HERA and LEP

4.1.3 Accelerator Guide Field

The guide field of real storage rings is in general piece-wise constant. It may in dipole insertions such as injection or wiggler magnets, change its strength or polarity. In this most common case, the expressions for the polarization level and rise-time can be generalized:

$$P_\infty = \frac{8}{5\sqrt{3}} \frac{\oint \rho^{-3} \mathrm{d}s}{\oint \left|\rho^{-3}\right| \mathrm{d}s} \tag{53}$$

$$\tau_p^{-1} = \frac{5\sqrt{3}}{8} c\lambda\!\!\!^{-}_c r_e \gamma^5 \frac{\oint \left|\rho^{-3}\right| \mathrm{d}s}{\oint \mathrm{d}s} \tag{54}$$

The difference with the uniform field case (47) (52) is in general negligible.

4.1.4 Experimental observations

The polarization build-up by the Sokolov-Ternov effect has been observed in all the electron storage rings where it has been sought. Table 2 gives the polarization data for some of these rings.

Table 2

Polarization experimental data for some electron storage rings. The time τ_p *is* the *theoretical* polarization time given by Eq.(54) at energy *E*. The polarization degree *P* is a typically *measured* value.

	VEPP[10]	VEPP2-M[11]	ACO[8,9]	BESSY[44]	SPEAR[45]	VEPP4[46]
E(GeV)	0.640	0.625	0.536	0.800	3.70	5.0
τ_p(min)	50	70	160	150	15	40
P(%)	52	90	90	>75	>70	80
	DORIS II[47]	CESR[48]	PETRA[49]	HERA[19]	TRISTAN[50]	LEP[51]
E(GeV)	5.0	4.7	16.5	26.7	29	46.5
τ_p(min)	4	300	18	40	2	300
P(%)	80	30*	80**	70**	75**	57**

Figure 19 shows the polarization build-up as function of time at 26.7 GeV in HERA, following closely equation (51). The fifth power law of the polarization time versus energy is observed when the beam energy is changed in a storage ring. In particular the polarization time at LEP becomes very long in the lower part of its energy range : five hours at the energy (46 GeV) of the Z_0 vector boson production. The polarization time also increases as the third power of ρ, explaining why several rings have different polarization times at the same energy.

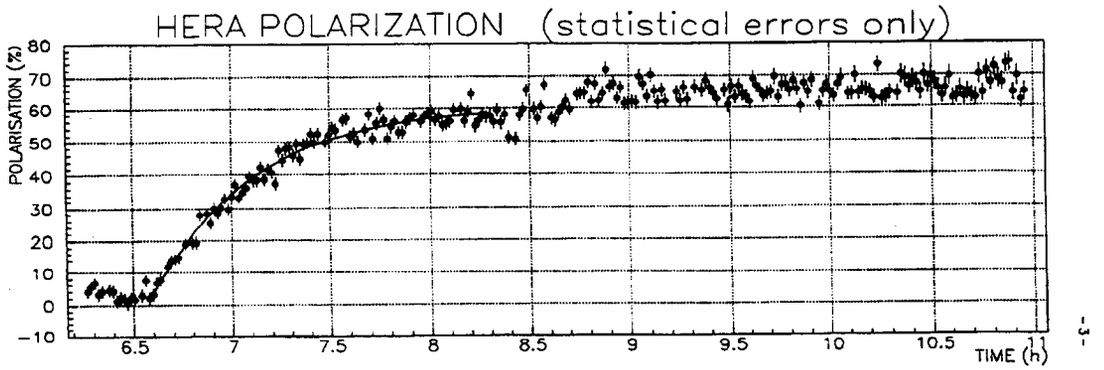


Fig. 19: Polarization *P* versus the time *t* in the storage ring HERA at 26.7 GeV.

The experimental values (see Table 2) indicate that the equilibrium polarization is in general lower than the theoretical 92.4% value, especially in multi-GeV rings.

4.1.5 Enhancement of the polarization build-up

---

* after 120 min.

** after harmonic correction (harmonic spin matching).

In CESR and LEP1, the design of the machine is such that the polarization time constant is as long as 5 hours. This is a serious drawback in practice: it takes a long time before the beams are highly polarized and empirical adjustments of the polarization are difficult given the long response time. It appears from formula (54) that it should be possible to reduce significantly the polarization rise-time by increasing locally the bending radius. This can be done with an asymmetric dipole wiggler as shown on figure 20.

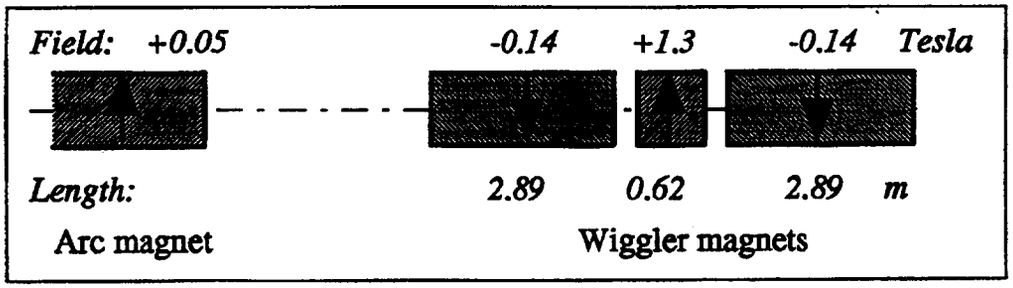


Fig.20: The LEP asymmetric wiggler to enhance the Sokolov-Ternov effect

The central part of the wiggler is designed with the highest magnetic field pointing in the direction of the main guide field; it is compensated on each side by the weakest field compatible with a reasonable magnet length. In this way, the asymptotic polarization level P_∞ is hardly decreased (Eq.(53)) while the rise-time is drastically reduced. Such a scheme has been designed for LEP and successfully used to decrease the rise-time by a factor of 2 [51]. Beyond, the enhanced synchrotron radiation which increases the beam energy spread, causes other phenomena which depolarize the beam.

4.1.6 Non-uniform magnetic field

The results of Sokolov and Ternov for a uniform magnetic field were generalized to arbitrary field configurations by Baier, Katkov and Strakhovenko [52] . The expressions for the asymptotic polarization level and the polarization rise-time can be deduced [53, 54] provided the spin dynamics is not perturbed by spin resonances: the approximation is valid at low energy, for weak (compensated) spin resonances, for a spin tune far from resonances.

$$P_\infty = \frac{8}{5\sqrt{3}} \frac{\left\langle \left| \rho^{-3} \right| \mathbf{n}\cdot\mathbf{b} \right\rangle}{\left\langle \left| \rho^{-3} \right| \left[1 - \frac{2}{9}(\beta\cdot\mathbf{n})^2 \right] \right\rangle} \tag{55}$$

$$\tau_p^{-1} = \frac{5\sqrt{3}}{8} c\lambda r_e \gamma^5 \left\langle \left| \rho^{-3} \right| \left[1 - \frac{2}{9}(\beta\cdot\mathbf{n})^2 \right] \right\rangle \tag{56}$$

The brackets $\langle\,\rangle$ stand for the average along the machine circumference; **b** is the unit vector in the direction of the magnetic field of bending radius ρ. These formulae show that the useful component of the magnetic field is along **n**, i.e. that **n** is the equilibrium polarization direction.

In general, non-uniform magnetic fields give an asymptotic polarization lower than the maximum of 92.4%.

4.2 Resonant spin diffusion and equilibrium polarization

Equation (46) shows that, in the vast majority of the cases (6000 times per turn in LEP at 45 GeV), the photon emission is not associated with a spin flip. The sudden energy jump causes the electron to initiate additional synchrotron and betatron oscillations if the dispersion functions do not vanish at the azimuth of the photon emission. The resulting orbital motion is thus "noisy", causing a diffusion eventually balanced by radiation damping along the perturbed trajectory. The spin experiences extra rotations and diffusion as well. The synchrotron and betatron oscillations are damped with a characteristic time (10^{-2} - 10^{-3} sec) much shorter than the polarization time of the Sokolov-Ternov effect (10^{+2} - 10^{+4} sec). It is thus legitimate to consider separately the spin diffusion due to the non-spin-flip events and the Sokolov-Ternov mechanisn.

4.2.1 Ideal storage ring

In an ideally planar ring, only synchrotron and horizontal betatron oscillations are excited by the photon emission. The small vertical angle of the electron recoil at photon emission is totally negligible at high energy. The diffusion due to the quantum excitation results in equilibrium longitudinal and horizontal emittances. The particle motion being constrained to remain in the horizontal plane, the quantum excitation only causes stochastic spin precessions about the vertical axis; they are of no consequence as long as the spin is vertical, i.e. in the **n** direction. (See Section 2.3.1) This is indeed the case if the polarization is due to the Sokolov-Ternov effect. This situation is different from a proton beam which has a finite vertical emittance and may be depolarized when crossing an intrinsic resonance even if there are no defects. (See Section 3.2) The only consequence of the quantum excitation is a partial randomization of the spin precession phases. The spin tune is defined by averaging over several turns.

4.2.2 Spin perturbation by quantum excitation

The general case is best explained by Figure 21. It shows the evolution of the phase space and spin coordinates of a reference particle which emits a photon. The first line represents the state of the particle just before the photon emission: the particle is at rest, at the origin of the coordinates, i.e. on the closed orbit; the spin is along **n** assumed to be vertical.

The second line represents the state just after the photon emission: the energy coordinate suddenly became negative. The position of the particle could not change during the short photon emission. However, the break-down of the position into closed orbit, dispersion orbit and betatron amplitude does change if the dispersion function does not vanish. The momenta change due both to the photon emission and for the same reason as above. This causes the particle to start oscillating in the three phase planes. Along the new trajectory, the spin motion is slightly different, precessing about a new spin axis **n**, in general tilted with respect to the initial spin direction.

The third line shows the evolution of the coordinates over a few damping times. The spatial and momentum coordinates are damped down by the well-known damping mechanism (see the CAS lecture on electron dynamics). As the trajectory gradually converges towards its unperturbed position, the spin axis **n** is gradually restored to its initial vertical position. The particle spin however, which precesses very rapidly about it, follows adiabatically the **n** axis and finds itself tilted at an angle when the orbital coordinates have been restored. The horizontal component of the spin precesses in a stochastic way as photons are emitted. Its average over all particles of the beam vanishes. The remaining polarization is the projection of the initial polarization vector onto the spin axis after it has been tilted by the photon emission. This rather simple picture of depolarization due to quantum excitation arises because of the very large difference between the time constants of the three relevant phenomena: $1\,\mu$s for one spin precession, 50 ms for the damping of the orbital coordinates and the gradual change of the **n** axis, 5 hours for the Sokolov-Ternov polarizing mechanism (LEP at 45 GeV).

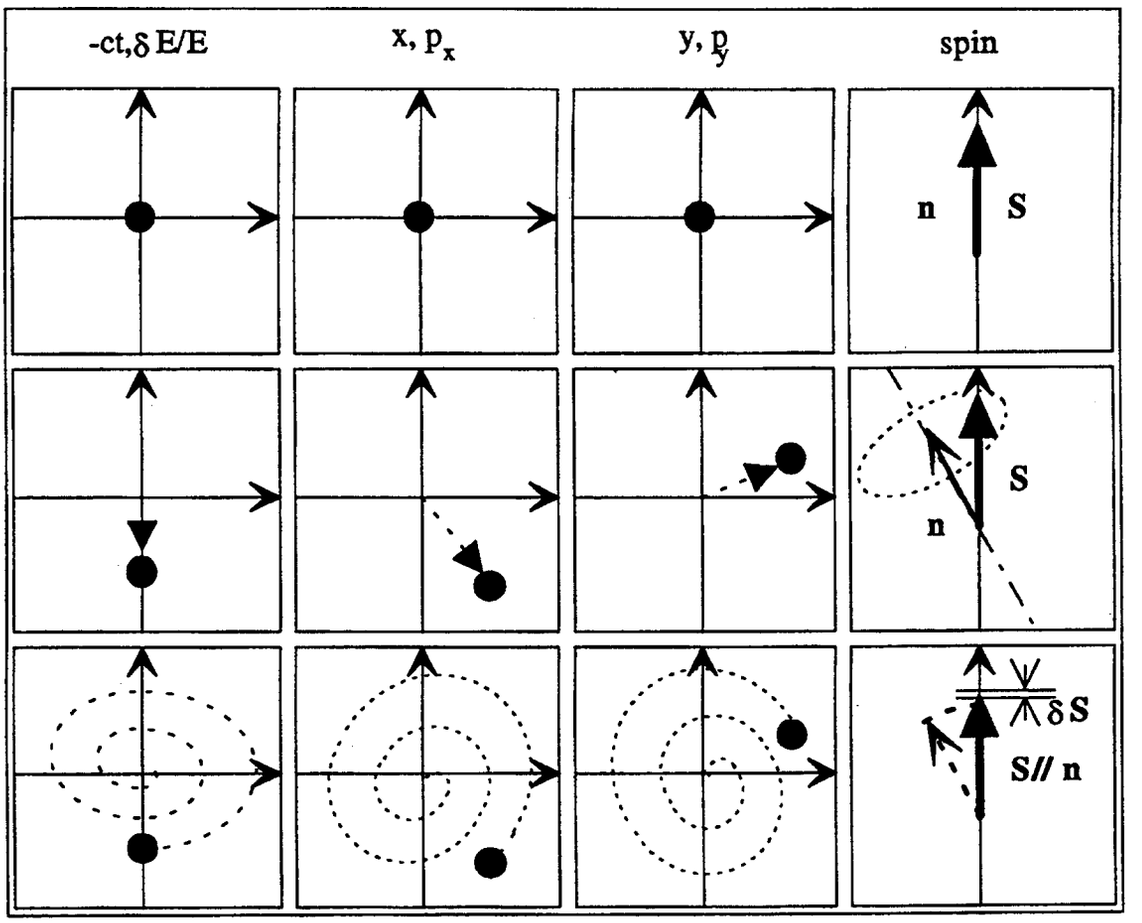


Fig. 21: Model of a quantum excitation

4.2.3 Spin diffusion and equilibrium polarization

This simple model of the spin diffusion process leads to a good estimate of the polarization level. Let us assume that the tilt $\delta\mathbf{n}$ of the **n** axis just after the emission of photon of energy δE is proportional to the relative energy loss (Figure 21):

$$\delta \mathbf{n} = \mathbf{d}(s)\frac{\delta E}{E} \tag{57}$$

The proportionality constant $\mathbf{d}(s)$ is a vectorial quantity which depends on the azimuth. It summarizes the contributions from the horizontal and vertical betatron oscillations as well as from the synchrotron oscillation.

For a single energy jump the decrease of the polarization P along the spin closed solution $\mathbf{n}$ is given by :

$$\frac{\delta P}{P} \approx \frac{1}{2}\left|\mathbf{d}\frac{\delta E}{E}\right|^2 \tag{58}$$

To combine in a simple way the effect of the many energy jumps we assume linearity of the orbital motion of the particles. In this case, which is generally a good approximation, the perturbations add up linearly in an incoherent way. As the times of photon emission and the energy jumps are random, the spin motion becomes a random process, i.e. diffuse away from the spin closed solution $\mathbf{n}$. Within the linear approximation the rate of the polarization decrease is obtained by summing the effects of all the successive energy jumps per unit time. The average rate $1/\tau_d$ of decrease of the spin component is found by averaging over the coordinate s and the energy jump δE:

$$\frac{1}{\tau_d} = \frac{1}{\mathrm{d}t}\frac{\delta P}{P} \approx \frac{1}{2}\dot{N}\left\langle \left|\mathbf{d}\frac{\delta E}{E}\right|^2 \right\rangle \tag{59}$$

where $\dot{N}$ is the photon emission rate and the brackets $\langle\ \rangle$ indicate the averaging. The quantity $\dot{N}\left\langle (\mathrm{d}E/E)^2 \right\rangle$ happens to be closely related to the polarization rise-time τ_p [55], yielding

$$\frac{\tau_p}{\tau_d} = \frac{11}{18}\left\langle |\mathbf{d}|^2 \right\rangle \tag{60}$$

When taking into account this depolarizing process Eq. (51) becomes:

$$P = P_\infty\left(1 - e^{-\frac{t}{\tau_p+\tau_d}}\right) \tag{61}$$

$$P_\infty = \frac{8}{5\sqrt{3}}\frac{1}{\left(1+\frac{\tau_p}{\tau_d}\right)} = \frac{8}{5\sqrt{3}}\frac{1}{\left(1+\frac{11}{18}\left\langle |\mathbf{d}|^2 \right\rangle\right)} \tag{62}$$

where the averaging $\langle\ \rangle$ is made over the azimuth s. The knowledge of $\mathbf{d}(s)$ is sufficient for a complete description of the polarization dynamics.

In a more thorough study of the equilibrium polarization Ya. Derbenev and A. Kondratenko [56] derived a formula for an arbitrary configuration of magnetic fields:

$$P_\infty = \frac{8}{5\sqrt{3}} \frac{\left\langle \left| \rho^{-3} \right| \mathbf{b} \cdot (\mathbf{n} - \gamma \frac{\delta \mathbf{n}}{\delta \gamma}) \right\rangle}{\left\langle \left| \rho^{-3} \right| \left(1 - \frac{2}{9} (\beta \cdot \mathbf{n})^2 + \frac{11}{18} \left| \gamma \frac{\delta \mathbf{n}}{\delta \gamma} \right|^2 \right) \right\rangle} \tag{63}$$

where ρ is the bending radius, $\mathbf{b}$ a unit vector along the transverse component of the field and β a unit vector along the reference orbit.

The vector $\mathbf{n}$ i*s no longer the spin closed solution attached to the closed orbit.* It is defined for any location or synchrotron trajectory as the solution of the BMT equation (11) which is quasi-periodic in the six-dimensional phase space, i.e. periodic in each angle variables and azimuth. The general $\mathbf{n}$ vector reduces to the spin closed solution when evaluated on the closed orbit.

Comparing formulae (62) and (63), it appears that $\gamma \delta \mathbf{n} / \delta \gamma$, named the spin-orbit coupling vector, can be identified with the deviation vector $\mathbf{d}$ of the simple diffusion model. The averaging in the Derbenev-Kondratenko formula (63) must be done over all six orbital coordinates since $\gamma \delta \mathbf{n} / \delta \gamma$ depends on oscillation phases and amplitudes, in addition to the averaging over the azimuth s. The $\gamma \delta \mathbf{n} / \delta \gamma$ linear term in the numerator of formula (63) usually gives a negligible contribution compared to the quadratic term in the denominator. The contributions of the electron recoil during photon emission, that have also been included in other calculations [57, 58], appear to be negligible at high energy.

The effective polarization rate in (61) is the sum of the Sokolov-Ternov rate τ_p^{-1} and of the depolarization rate τ_d^{-1}. Therefore in the presence of imperfections the polarization time τ is shorter than the ideal value τ_p . When the polarization time is too long to reach the equilibrium polarization, one can derive its value P_{eq} just by measuring the rate of increase of the polarization at the beginning and by using equation (61). It can also be used to calibrate in polarization degree the experimental data of a polarimeter (Fig. 22).

This analysis shows how different the polarized electrons and protons are. The spin motion of a proton is deterministic and beam depolarization results from the incoherent spread of trajectories and energies. For an electron the spin motion is random and beam depolarization results from the random quantum fluctuations of synchrotron radiation. Moreover, the electron beam depolarization is an irreversible process. On the contrary proton beam polarization can be restored. A Siberian snake is an example of polarization restore after one turn. That also explains why Siberian snakes are not efficient in overcoming depolarization in electron rings [59].

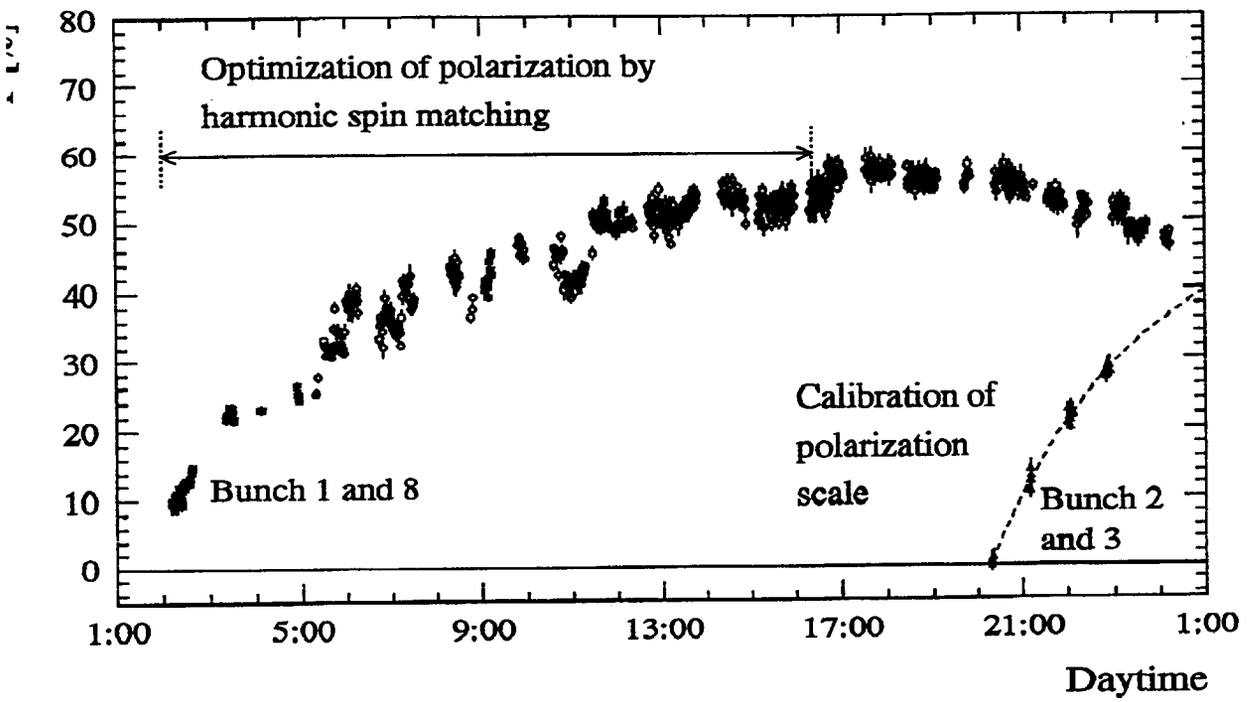


Fig. 22: Polarimeter data versus time at an energy of 46.55 GeV in LEP. The bunches 2 and 3 were selectively depolarized to allow the absolute calibration of the polarization scale [51]

4.2.4 Spin-orbit coupling in the linear approximation

In a small amplitude linear model of orbital motion, the result of the first-order calculation of the spin-orbit coupling vector $\mathbf{d}(s)$ is a linear combination of contributions from the three oscillations [60]:

$$\mathbf{d}(s) = \frac{1}{2}\mathrm{Im}\left[(\mathbf{m}+i\mathbf{l})^{*}\left(\Delta_x + \Delta_{-x} + \Delta_z + \Delta_{-z} + \Delta_s + \Delta_{-s}\right)\right] \tag{64}$$

Each contribution can be expressed by the product of a resonant term with an integral over s of the coupling between the spin motion and the considered orbital oscillation.

$$\Delta_{\pm x,z}(s) = (1+a\gamma)\frac{\left[-D \pm i(\alpha D + \beta D')\right]_{x,z}}{\sqrt{\beta_{x,z}}}\frac{1}{e^{2i\pi(\nu\pm Q_{x,z})}-1}\mathrm{J}_{\pm x,z}(s)e^{\mp i\phi_{x,z}} \tag{65}$$

$$\Delta_{\pm s}(s) = (1+a\gamma)\frac{1}{e^{2i\pi(\nu\pm Q_s)}-1}\mathrm{J}_{\pm s}(s)e^{\mp i\phi_s} \tag{66}$$

where D and D' are the dispersion and its derivative, and $\phi_{x,z,s}$ the phases of the betatron and synchrotron oscillations. On resonance, the spin-orbit coupling vector $\mathbf{d}(s)$ becomes arbitrarily large in this linear approximation and the polarization vanishes according to (63). The condition is:

$$\nu = k \pm Q_{x,z,s}$$

corresponding to the linear resonances described in section 2.4.4. The real keys for controlling the polarization level are the spin-orbit coupling integrals $\mathbf{J}_{\pm x,z,s}(s)$, for, if they can be made to vanish, spin-transparency can be recovered, even close to the resonance conditions. These integrals already appeared in the expression of the resonance strengths (see Eq.(38)).

$$\begin{aligned}
\mathbf{J}_{\pm z}(s) &= \int_s^{s+C} (\mathbf{m} + i\mathbf{l}) \cdot \mathbf{x}\, K\sqrt{\beta_z}\, \mathrm{e}^{\pm i\phi_z}\, \mathrm{d}s' \\
\mathbf{J}_{\pm x}(s) &= \int_s^{s+C} (\mathbf{m} + i\mathbf{l}) \cdot \mathbf{z}\, K\sqrt{\beta_x}\, \mathrm{e}^{\pm i\phi_x}\, \mathrm{d}s' \qquad (67,68,69) \\
\mathbf{J}_{\pm s}(s) &= \int_s^{s+C} (\mathbf{m} + i\mathbf{l}) \cdot K\left(D_x \mathrm{z} + D_z \mathrm{x}\right)\mathrm{e}^{\pm i\phi_s}\, \mathrm{d}s'
\end{aligned}$$

As already noted in section 2.4.3 , $\mathbf{J}_{\pm x,s}(s)$ are sensitive to the tilt of the **n** axis. The spin-orbit coupling vector $\mathbf{d}(s)$, just like the tilt of the closed spin solution **n**, shows a large increase close to the integer resonances $\nu = k$

To summarize, the spin-orbit coupling vector $\mathbf{d}(s)$ becomes large on resonance, causing a significant reduction of the polarization. It scales linearly with energy. Its amplitude, for small oscillation amplitudes, is conveniently expressed in terms of the spin-orbit coupling integrals. There are several implementations of the calculation of $\mathbf{d}(s)$ and of the equilibrium polarization in the linear approximation. The most popular is SLIM [61]. Its approach is very elegant and has been widely used in many other codes. The limitations are mainly technical: SLIM assumes thin lenses and does not allow optics imperfections to be described in a natural way. The program SITF [62] integrates approximately the spin motion in thick elements. It allows for a complete description of optics errors, in the FORTRAN-like PETROS syntax. SITF was integrated as a module in MAD [63]. There it benefits from the powerful MAD language, which allows a natural and flexible description of the optics imperfections and, for example, of the harmonic orbit bumps. This powerful implementation has contributed to the success of the polarization program in LEP, largely based on calculations.

4.2.5 Spin-orbit coupling at higher orders

As the amplitudes of the betatron and synchrotron oscillations scale linearly with energy in a given machine, the assumption of vanishingly small oscillations is bound to break down at some energy and so does the analysis we performed, only considering the **n** axis on the closed orbit. Qualitatively, we already saw in sections 2.4.5 and 2.4.6 that new resonance conditions emerge, both due to the influence of the non-linear fields, e.g. of the sextupoles, and to the FM modulation of the beam oscillations by the finite amplitude synchrotron motion. In addition, the spin-orbit coupling vector also increases linearly with energy. The spin resonances become wider and wider until their separation becomes comparable with their width. A run-away is expected at an energy threshold above which polarization would become practically impossible.

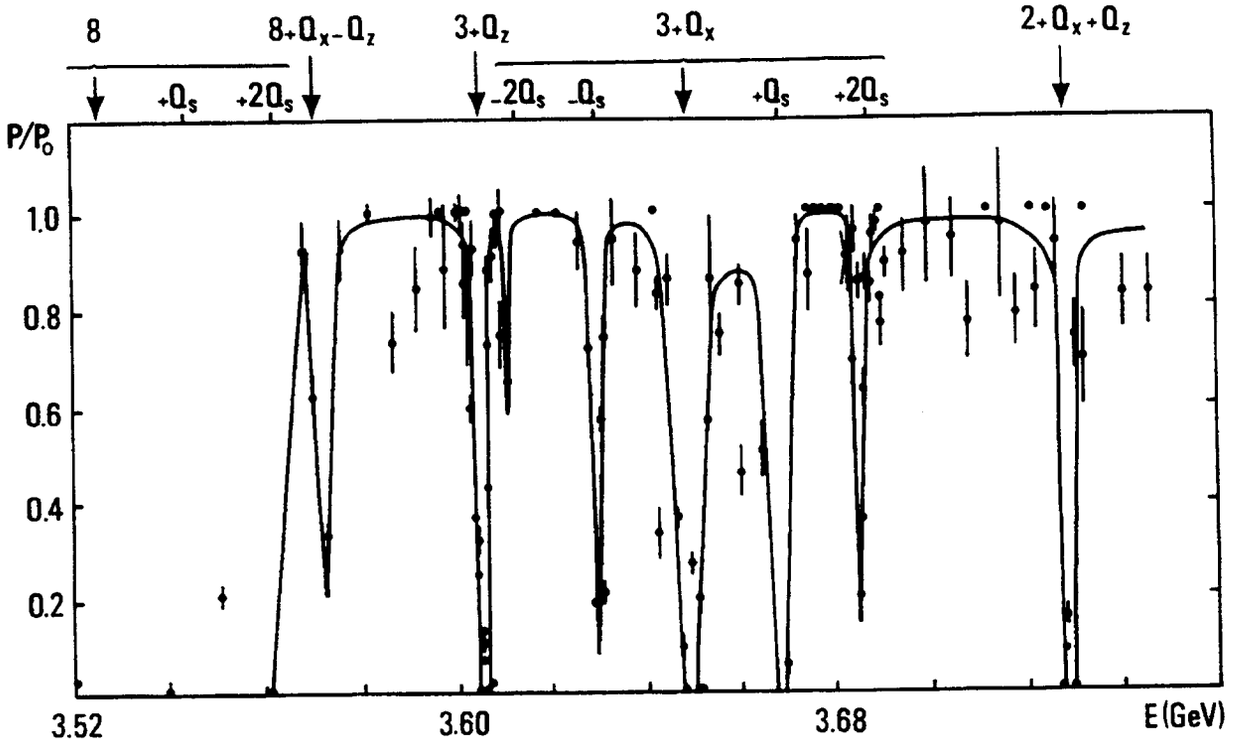


Fig.23: Relative polarization P/P_0 (P_0 = 92.4 %) versus beam energy at SPEAR [45]. The solid line is a hand-drawn line to guide the eye. The spin resonances labelled : $k + k_x\, Q_x + Q_z + k_s\, Q_s$ are indicated above the figure.

Figure 23 shows the most detailed scan of the polarization as a function of beam energy at the SPEAR storage ring[45]. It shows the depolarization on many spin resonances in a small energy interval, particularly near the synchrotron resonance γa = 8 at 3.5 GeV. Despite the low energy, several non-linear and higher-order spin resonances appear harmful, particularly synchrotron satellites of the horizontal betatron resonance γa = 3 + Q_x. This evidence, and some early calculations, left the impression that polarization might not be possible in HERA and LEP. Given the renewed interest in polarized beams, several attempts were made to calculate higher-order effects.

Several analytical approaches [64 and e.g. 65] have singled out the modulation index x as the relevant parameter for higher-order synchrotron effects:

$$x = \left(\frac{\nu}{Q_s} \frac{\sigma_e}{E} \right)^2 \tag{70}$$

σ_e / E is the relative energy spread of the beam. The modulation index ranges from 3% in SPEAR to 50% in LEP at 46 GeV and about 200% in HERA at 27 GeV and in LEP with wigglers. The asymptotic polarization level can be estimated from an enhancement function $C(x)$ of the depolarization as given by formula (62):

$$P_\infty = \frac{8}{5\sqrt{3}} \frac{1}{1 + C(x)\frac{\tau_p}{\tau_d}\Big|_{x=0}} \tag{71}$$

$C(x)$ is a combination of Bessel functions encountered in FM modulation. It summarizes the contribution to the depolarization of the synchrotron satellites which surround each linear spin resonance. (See section 2.4.6) The amplification factor for the strongest synchrotron spin resonances is typically between 1 and 2 at HERA over its energy range and between 1.5 and 20 at 46 GeV at LEP depending on the wiggler excitation (Fig. 24).

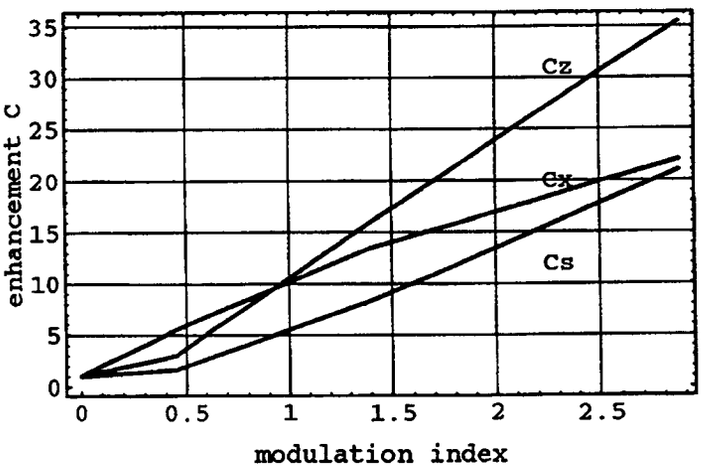


Fig. 24: Variation of the enhancement function with the modulation index in LEP [65]

The importance of the beam energy spread is confirmed by an experiment made at LEP where it was increased at constant energy (46 GeV) by means of asymmetric wiggler magnets (Fig. 25). The polarization level decreases with the beam energy spread, in agreement with calculations [51].

A purely analytical approach is limited to the calculation of the enhancement factor, i.e. of the relative strength of the synchrotron satellites. The full spectrum of the generic linear resonances must be calculated numerically. Various numerical or semi-numerical approaches have been carried out lately, all requiring heavy computations:

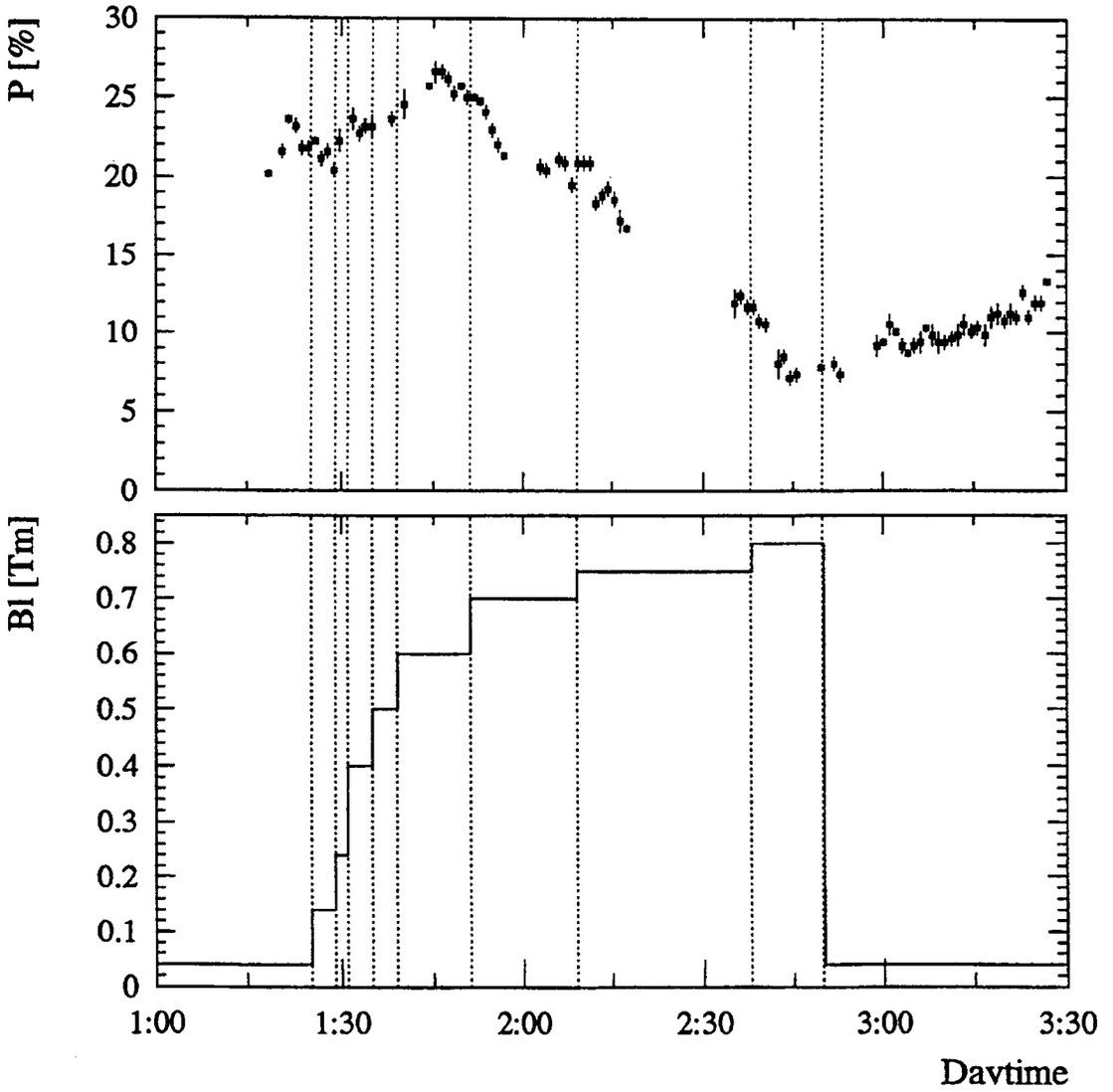


Fig. 25: Measured polarization as a function of beam energy spread at LEP [51], enhanced by a wiggler of strength B ℓ

- SMILE [66] implements a recursive perturbative solution of the BMT equation, where the small parameters are taken to be the amplitudes of the oscillations. This program could predict the non-linear and higher-order spin resonances at relatively low energy (SPEAR at 3.6 GeV). However a critical analysis of its prediction for HERA (27 GeV) and LEP (46 GeV) showed that the very pessimistic results obtained were due to the non-convergence in the calculation of the synchrotron spin resonances at these energies. The program was modified to compute the latter using the analytical approach mentioned above. SMILE is available in a stand-alone version and as a module of the optics program MAD.

- SODOM [67] implements a non-perturbative approach. Advantage is taken from the known periodicities of the general **n**-axis and of its transport around the machine to compute the most significant harmonics of **n**. The variation of **n** with respect to a finite energy difference yields its derivative $\gamma \delta \mathbf{n} / \delta \gamma$, allowing the computation of the equilibrium polarization level. The program is available in a stand-alone version and as a module of MAD [68]. Preliminary checks against measured higher-order depolarization phenomena in LEP (depolarization by an asymmetric wiggler, Fig. 25) and prediction of the strength of synchrotron satellite (Fig. 26) seem to validate this approach [51].

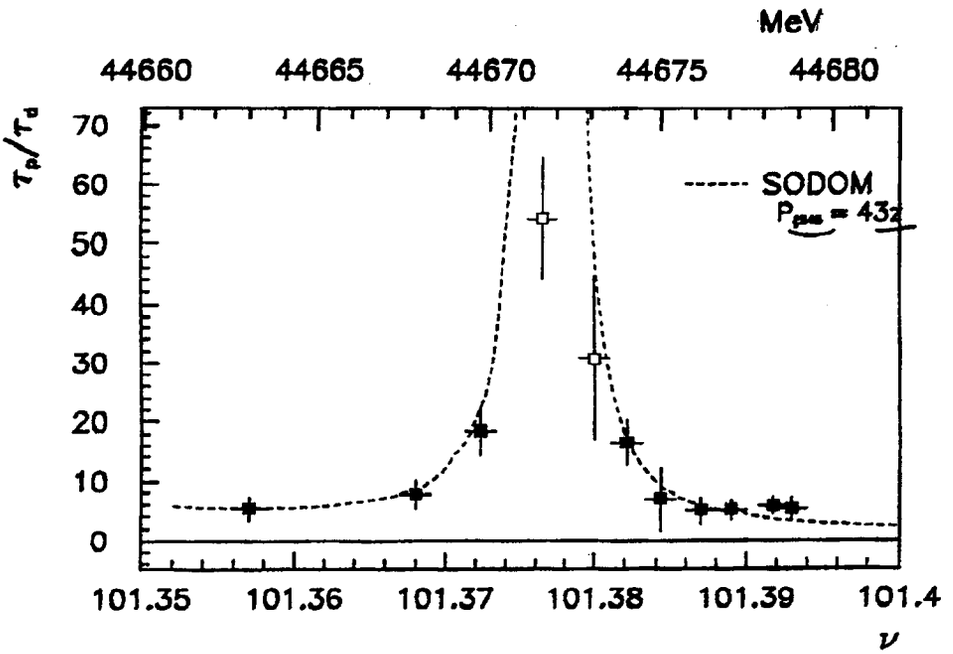


Fig.26: Measured and predicted synchrotron satellite in LEP [51]

- SITROS [62] is a Monte-Carlo tracking of the orbital and spin coordinates of a sample of particles. It was used to assess the possibility of observing polarization at the LEP energy and is presently used in HERA. The limitations of this approach are related to the requirements in computer time. To keep it reasonable, it is necessary to simplify grossly the quantum excitation by lumping the photon emissions in only a few azimuths. The number of tracked particles is necessarily limited. The core of the distribution is therefore well simulated but not the tails where strong depolarization is expected. Yet the rather good agreement between SITROS predictions and measured polarization at HERA seems to validate the model [19] (Fig. 27). At LEP, the disagreement between SITROS and SODOM results, the SITROS approach yielding more polarization, have not yet been resolved.

This rapid overview of the non-linear or higher-order phenomena shows that this field is still the subject of research and experimentation. Other approaches are now being developed, such as using Lie algebraic techniques to carry out perturbation expansions [69]. More

experiments at high energy are necessary to probe the theories. An accurate estimate of the threshold energy at which polarization would disappear is not yet possible.

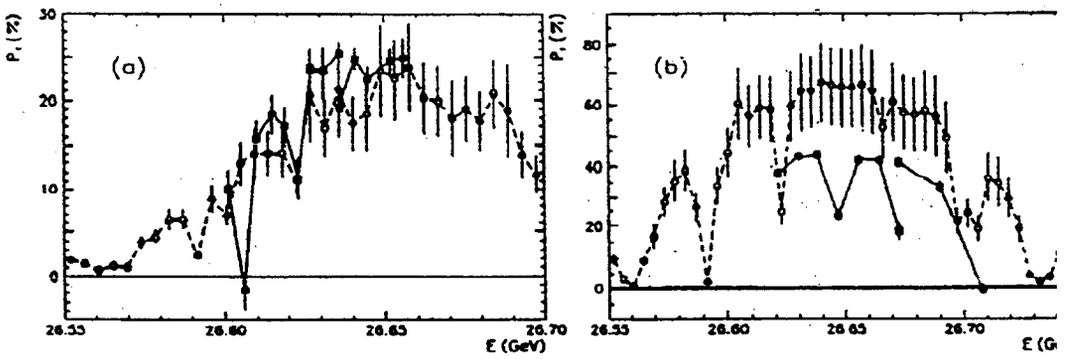


Fig. 27: Comparison between the measured polarization in HERA before and after optimization [19] and the SITROS predictions.

With the search and maximization of the polarization in HERA, Tristan and LEP, a wealth of new data is suddenly available. It is instructive to compare these data with the scaling law for the polarization *in the linear approximation*. According to formulae (65, 66), the asymptotic polarization should scale like:

$$P_{\infty} = \frac{8}{5\sqrt{3}} \frac{1}{1+(\alpha E)^2} \tag{72}$$

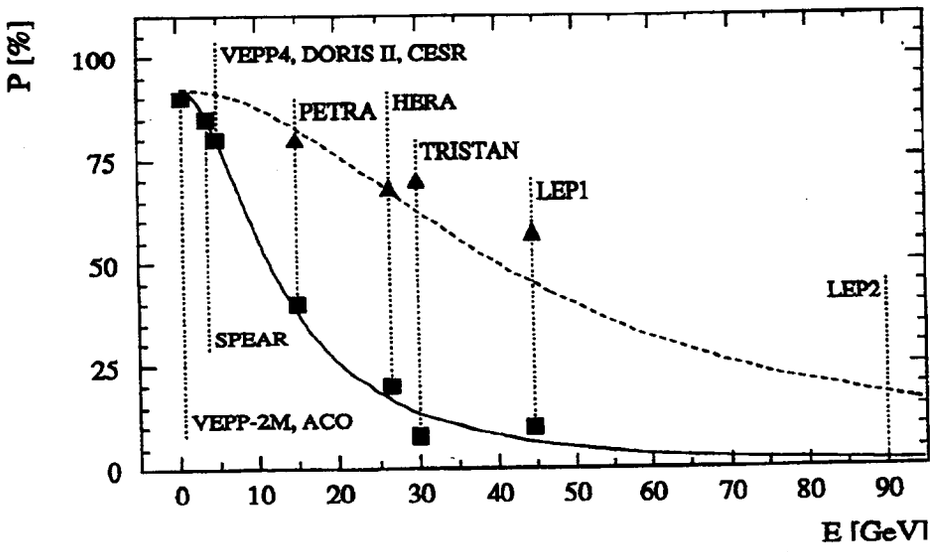


Fig. 28: Comparison between the predictions of the linear approximation and the measured polarization data, before and after harmonic spin matching.

It is remarkable that the measured data (figure 28) are almost perfectly compatible with the linear approximation. Yet the higher-order resonances are already observed at 3.7 GeV (Fig. 23) and above. This shows that the width of the higher-order spin resonances is still small compared to their spacing up to the highest energy reached so far, i.e. 46 GeV in LEP. If the working point is carefully chosen away from higher-order spin resonances, the asymptotic polarization is well predicted by the linear theory.

4.2.6 Depolarization by the beam-beam effect

In a beam-beam collision, the electro-magnetic field of one beam causes a precession of the spins of the particles in the second beam. Since SPEAR and VEPP-2M, collisions of polarized beams have not been required by physics and the expected depolarization not studied in depth. There was an experimental indication [43] at PETRA that at higher energy the beam-beam interaction can reduce the equilibrium polarization above the beam-beam limit. Nowadays, at energies comparable to the mass of the Z boson, where the electro-weak interactions dominate in the e^+e^- or ep collisions, there is a new interest for experiments with polarized beams. More information on the depolarization by the beam-beam effect is awaited before deciding on their feasibility.

4.3 Maximization of the polarization

4.3.1 Minimization of the depolarization

As can be seen from figure 28, the natural polarization decreases rapidly with increasing energy. For example, the expected polarization in LEP is less than 2% in standard physics conditions. It is possible to increase this expectation by minimizing the machine imperfections, optimizing their correction and by a careful choice of the machine parameters:

- In large machines, the alignment of the magnets is critical. The mere correction of the orbit with respect to beam monitors generally locally aligned on the near-by quadrupoles does not guarantee a good compensation of the spin rotations. Figure 29 shows the calculated polarization in LEP for given misalignments and residual orbits after correction. The results are averaged over a sample of random imperfections [70]. Realigning LEP vertically with the tighest tolerance has indeed provided the calculated increase in natural polarization.

- The most efficient correction of the vertical orbit is essential. The quality of the beam position monitors and their number set the limits of the correction capabilities. For example, in LEP, the 90° lattice was abandoned for a lattice with 90° in the horizontal plane but 60° in the vertical plane, thus increasing the sampling of the vertical orbit. This, together with improved beam monitors, allowed the rms residual orbit deviation to be decreased from a typical 1 mm to 0.3 mm. The correction of the orbit with a number of orbit correctors larger than strictly necessary reduces the vertical dispersion function significantly. This in turn reduces the excitation of the synchrotron spin resonance as may be inferred from (69).

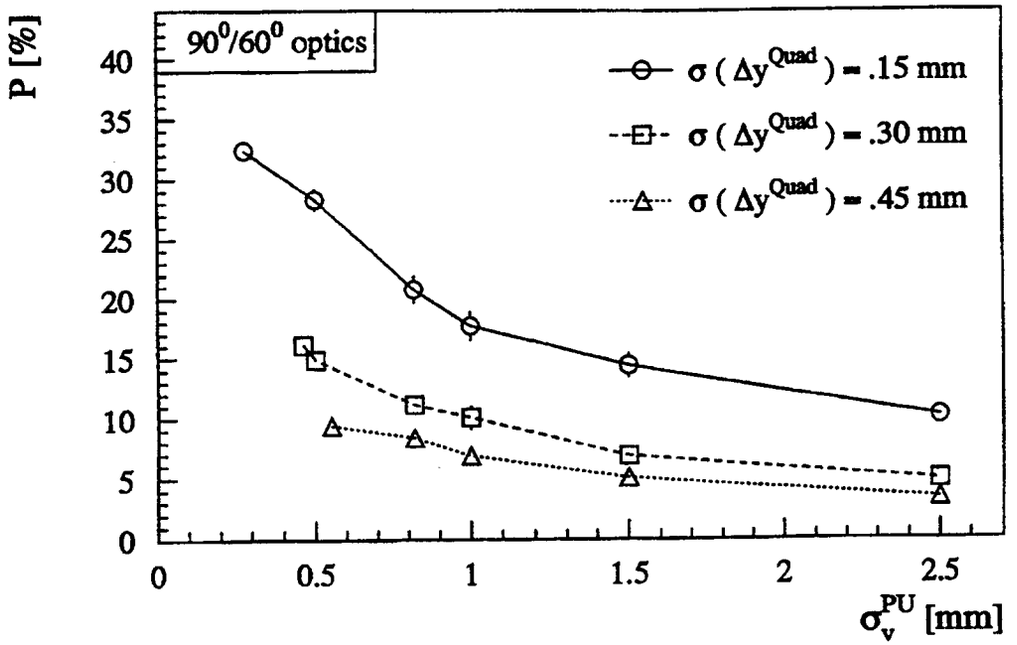


Fig.29: Dependence of the polarization on the machine alignment in LEP [70]

- A careful choice of the beam eigentunes $\{Q_x, Q_z, Q_s, \nu\}$ maximizes the 'natural' polarization level. The spin tune, i.e. the beam energy, should be selected with an integer part far from multiples of the ring superperiodicity (see section 2.4.4). The requirement of obtaining polarization at a given beam energy to calibrate a particle mass may lead to changing the integer parts of the betatron tunes, so as to avoid the strong integer systematic resonances (see section 2.4.3). At the LEP energy, these resonances do not allow polarization to develop; the physics optics was matched to obtain betatron tunes optimal for obtaining polarization at the Z mass. The fractional part of the spin tune should be close to 1/2, yet away from higher-order satellites of linear resonances. Their density was minimized in LEP by choosing a synchrotron tune such that the upper and lower sidebands of the linear resonances are superimposed. The fractional parts of the betatron tunes should be as low as possible so that low-order satellites are as far as possible from the half integer.

4.3.2 Minimization of the tilt of the **n**-axis

In the high-energy storage rings, the global minimization of the depolarization sources is not sufficient to reach a high polarization level. It becomes necessary to improve the equilibrium polarization by the so-called spin-matching procedure, which aims at making the ring more spin-transparent. The first step is to minimize the deviation of the spin closed solution **n** from the vertical. We saw in section 2.4.3 that this deviation is large on integer spin resonances. It is furthermore possible to show that the depolarizing effect associated with the tilt of the **n**-axis decreases with the 4th power of the distance between the spin tune and the integer spin resonance [e.g. 54]. One is thus naturally guided to a compensation of the two near-by integer spin resonances. The method is known as harmonic spin matching. Its principle is to generate

an horizontal field which is stationary in the (**l,m,n**) spin frame and to adjust its amplitude and phase so as to compensate the driving term of each integer resonance. In other words, the perturbing horizontal magnetic field

$$\Delta B(s) = K(s) \cdot y_{co}(s) \tag{73}$$

due to the vertical displacement of the orbit $y_{co}(s)$ in the quadrupole gradient $K(s)$ is Fourier analysed. The angle in this analysis is the angle of the orbit in the laboratory frame α. The spin precession ψ is then

$$\psi = h\alpha \tag{74}$$

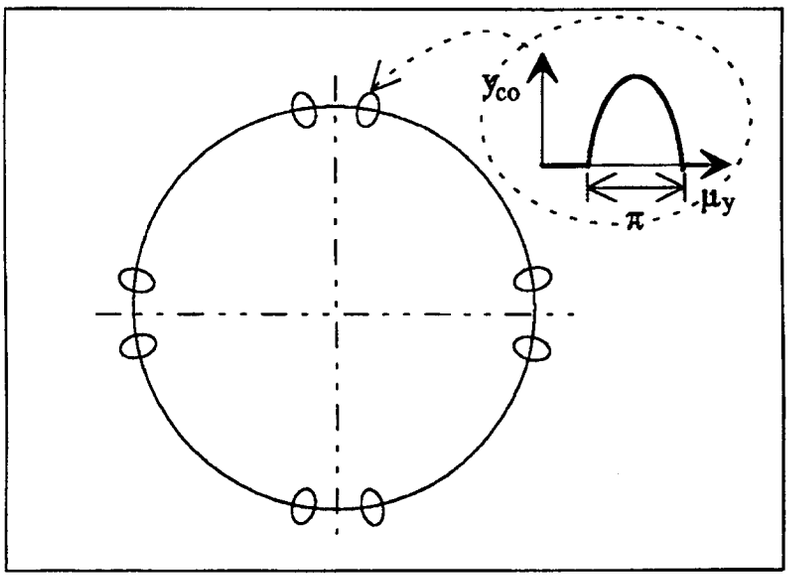


Fig.30: In HERA, LEP and Tristan, a pattern of vertical π-bumps in the arcs is used to produce the sine and cosine magnetic field harmonics necessary to compensate the near-by integer spin resonances

with the h harmonic number. The spin resonance compensation is achieved by generating opposite harmonics using a pattern of vertical orbit correctors. However the offending harmonics do not affect the closed orbit in a visible way (typically less than 0.1 mm) as they are far from the betatron tunes. Depending on the sensitivity of the beam position monitors, their compensation can be either empirical or computed from the measured orbit. The empirical procedure has been successfully applied at PETRA [49], where it allowed the polarization to be increased from 40% to 80%. The formalism was generalized to non-flat machines, e.g. including spin rotators[70]. Since then, the method was improved in LEP by observing that the

generation of horizontal field harmonics could be economically achieved by a few short vertical closed orbit bumps rather than by an orbit perturbation all around the ring (fig. 30].

Figure 31 shows the optimization in HERA where the sine and cosine components of the two near-by interger resonances are compensated. In LEP the very slow rise-time of the polarization would jeopardize an empirical compensation. This was a strong incentive to upgrade the beam monitoring system with the best technology. Figure 32 shows a harmonic analysis of the vertical closed orbit in the vicinity of the spin tune (103.47) before and after a calculated correction. Figure 33 shows the abrupt change of slope of the polarization build-up after the compensation was applied. The same technique was used at TRISTAN [50], where it allowed the polarization to be increased to 75%, and in proton rings (see section 3.2).

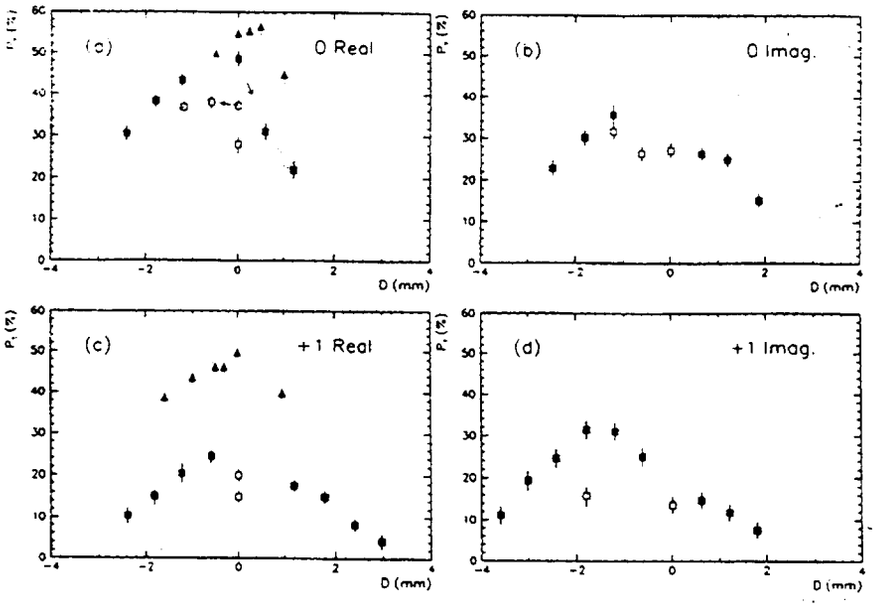


Fig.31: Optimization of the polarization at HERA by harmonic spin matching; the two parameters of the two closest integer resonances are compensated by trial and error. [19]

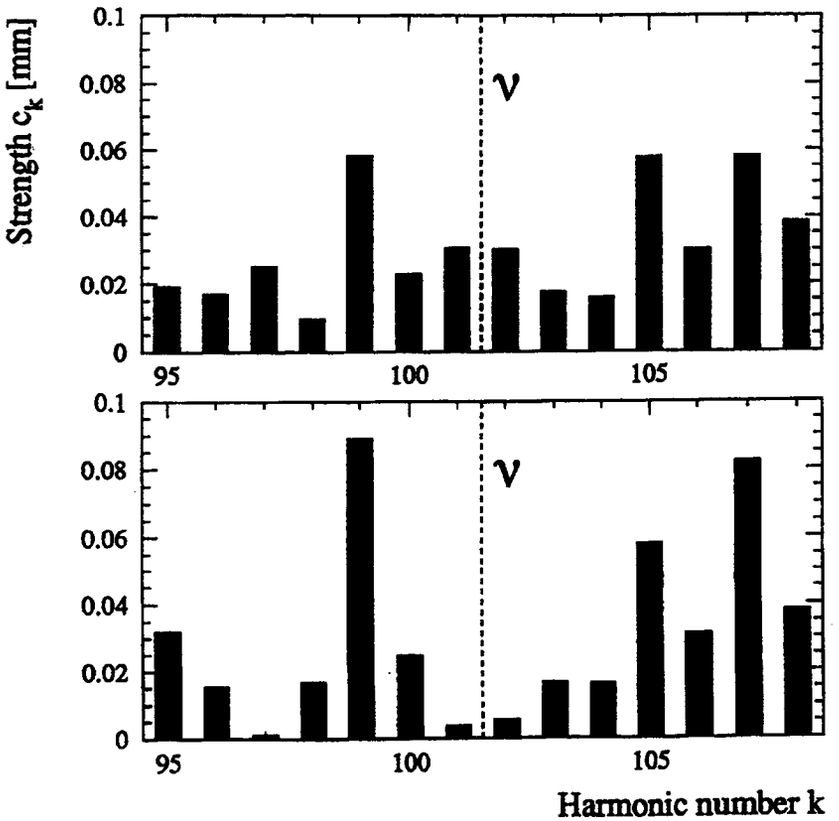


Fig.32: Harmonic analysis of the LEP vertical closed orbit in the spin precession frame. Figure a) shows the spectrum before harmonic correction, figure b) after correction. [51]

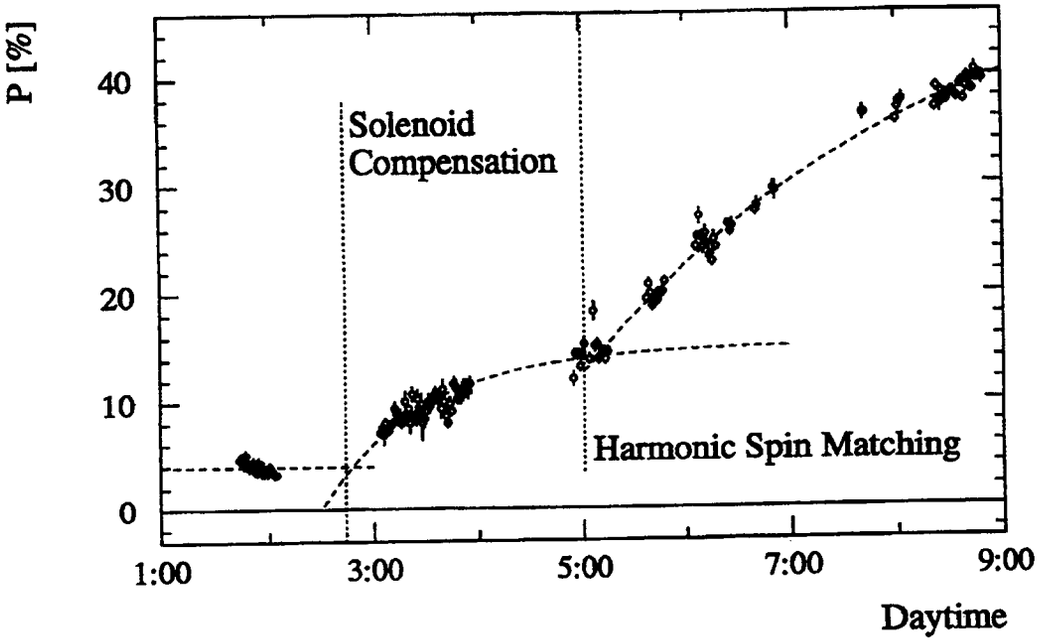


Fig. 33: Improvement of the polarization build-up in LEP following a calculated harmonic correction of the vertical orbit. [51]

4.3.3 Minimization of the Spin-orbit Coupling Vector **d**

The high polarization levels recorded in Table 2 were all obtained by harmonic spin matching of the integer resonances. Yet, formulae (63, 64, 65, 66) shows that the depolarization is due to the betatron and synchrotron spin resonances and not to the integer resonances. As already introduced in section 2.4.3, the reason for the efficiency of the harmonic spin matching can be found by a close examination of the spin-orbit coupling integrals (67), (68), (69). In the general case of a moderate deviation of the **n** axis from the vertical, they can be approximated as follows:

$$\mathbf{J}_s \approx \oint K e^{i\psi}\left(D_y - D_x|\delta\mathbf{n}|e^{i\psi_0}\right) ds \tag{75}$$

$$\mathbf{J}_x \approx \oint |\delta\mathbf{n}| e^{(i\psi+\psi_0)}\left(K\sqrt{\beta_x}e^{\pm i\phi_x}\right) ds \tag{76}$$

$$\mathbf{J}_y \approx \oint e^{i\psi}\left(K\sqrt{\beta_y}e^{\pm i\phi_y}\right) ds \tag{77}$$

At high energy, because of the large beam energy spread, the synchrotron spin resonances $\mathbf{J}_s$ are overwhelming. The horizontal betatron resonances $\mathbf{J}_y$ are also somewhat excited by betatron coupling while the vertical betatron resonances $\mathbf{J}_y$ are rather weak. The two main depolarizing resonances are thus directly dependent on the tilt of the **n**-axis.When the orbit is well corrected, this term is generally dominant in $\mathbf{J}_s$.

Although the polarization levels reached with harmonic spin matching are rather respectable, there is an incentive to compensate more exactly the spin-coupling integrals at high energy. It can indeed be shown that the strength of the higher-order spin resonances are proportional to the strength of the corresponding linear resonances. By compensating the former beyond the requirement for negligible linear depolarization, one can hope to weaken the higher-order depolarization and increase the threshold above which polarization disappears. Figure 34 shows the polarization calculated both linearly and by tracking with SITROS for LEP with very strong asymmetric wigglers, causing a beam energy spread at 45 GeV equivalent to the natural energy spread at 90 GeV. The four near-by integer spin resonances are corrected, assuming perfect beam position monitors. In which case, the spin-orbit coupling integral $\mathbf{J}_s$ becomes the Fourier transform of the vertical dispersion function modulated by the quadrupole gradient, assuming an integer spin tune (formula (75)). This term was corrected for $\nu = 104$ by exciting compensating dispersion harmonics. In the vicinity of the integer, the polarization is calculated to reach then very high values. The difficulty of this compensation lies in the requirement of generating dispersion harmonics without orbit harmonics. Fortunatly these harmonics are evaluated in the spin precession frame which does not rotate in the straight sections. Dispersion generated by orbit bumps in the straight sections meets the above-mentioned requirement.

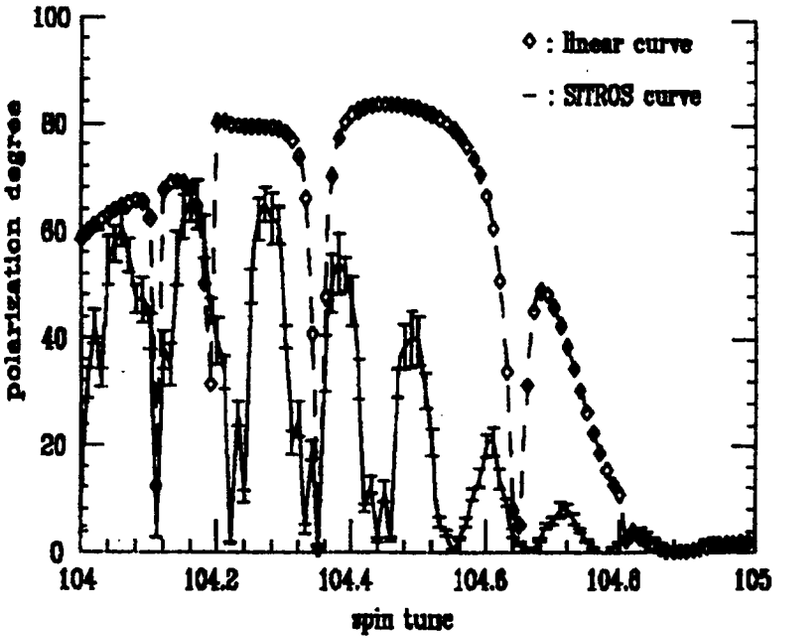


Fig. 34: Calculated polarization in LEP with strong asymmetric wigglers after complete compensation of the synchrotron spin-orbit coupling integral at $\nu = 104$.

It is equally possible to create harmonics of the focusing to compensate the depolarization by the betatron spin resonances. The interested reader is referred to Ref.[49].

4.4 Spin rotators

Spin rotators are special insertions which allow the **n** vector to be rotated. They take advantage of the non-commutativity of finite rotations to break the proportionality relationship between the spin precession and the orbit rotation (Eq. (15)). Spin rotators may be used to restore the **n** vector in its vertical position if it was tilted away by, e.g. experimental solenoids. A second aim is to decouple the spin tune and the energy, so as to allow easy acceleration without spin resonance crossing. These insertions, called Siberian snakes, were analysed in section 3.3. The main purpose of the 90° spin rotators is to rotate the polarization vector from its natural vertical position to the longitudinal position at the interaction point and back to the vertical position before entering the accelerator arc. Indeed, the experimental study of the electroweak interactions in e^+e^- annihilations or collisions calls for longitudinal helicity states.

4.4.1 Compensation of the solenoids

The field of the experimental solenoid is generally along the nominal orbit. It therefore does not disturb the closed orbit, but causes the spin to rotate about the horizontal direction. The skew quadrupoles generally used to cancel the betatron coupling do not restore the spin direction. The tilt of the **n** vector propagates all around the machine and causes, in the case of LEP, an almost complete depolarization. Because it was necessary to obtain polarization in

physics conditions, small spin rotators were positionned on each side of each solenoid to restore the vertical position of the **n**-axis in the arcs. This was in fact done merely by using four vertical closed orbit bumps positionned at the beginning of the arc [72]. Figure 35 shows the projection of the **n**-axis onto the horizontal direction. The bump amplitudes (some 6 mm) were adjusted for an exact cancellation of the precession in the solenoid. The calculated polarization increased from 2% to more than 75%. Although the depolarization is here sufficiently and elegantly corrected, this example shows that restoring the **n**-axis is very effective but not sufficient to achieve complete spin transparency.

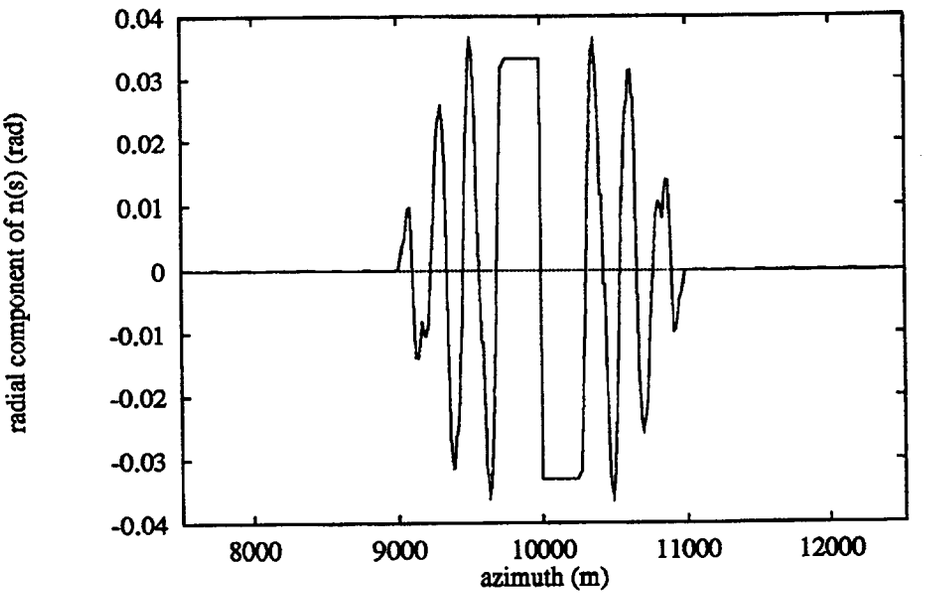


Fig.35: Two spin rotators per solenoid compensate the precession of the **n** vector. The solenoid is positioned at the antisymmetry point.

4.4.2 Spin rotation at the interaction points

The requirement of a longitudinal spin is incompatible with the Sokolov-Ternov polarizing mechanism. For that reason, the polarization must remain vertical in the arcs, where most of the synchrotron radiation is emitted. In an interaction region, the spin closed solution **n** is first bent into the beam direction by a first 90° spin rotator before the collision point. Then a second rotator bends it back to the vertical direction before entering the next arc. These two rotators are two sets of bending magnets with an antisymmetric configuration w.r.t. the collision point in the vertical plane. The antisymmetry has the advantage to leave the vertical closed orbit unperturbed in the arcs. There are many ways to bend the spin by a combination of horizontal, vertical and possibly longitudinal rotations. The choice of a solution involves an optimization process taking into account the depolarizing effect of the rotators, the allowable vertical emittance increase, the energy range where the spin is to be rotated, the synchrotron radiation background to the experiment, the possibility of tilting the detectors to match an inclined orbit.

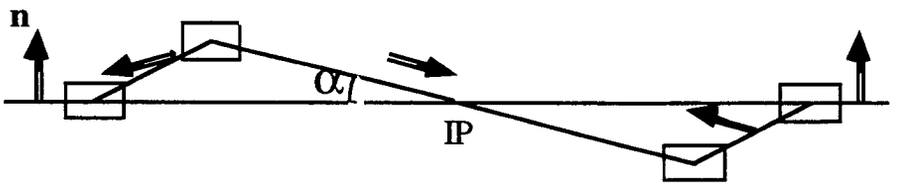


Fig. 36: Schematic side view of an S-bend configuration of two spin rotators antisymmetric w.r.t. the collision point IP. Each rotator consists of two magnets vertically bending the electron trajectory by an angle α and -2α respectively. With $\alpha = \pi/(2\gamma a)$ the spin closed solution **n** is longitudinal at the IP.

The simplest configuration [73] of rotators is an S-bend, made of four vertically bending magnets (Fig. 36). It produces a vertical orbit displacement around the collision point, where the orbit crosses the median plane at a vertical angle α. Since the spin rotation angle is γa larger than the bending angle (15), the 90° rotation required to align the spin vector along the orbit results from the inclination of the latter by $\alpha = \pi/(2\gamma a)$. In the example of the S-bend rotators proposed [74] for LEP at 46 GeV, the vertical orbit slope is $\alpha = 15$ mrad, requiring a corresponding tilt of the physics detectors.

If for some reason the spin rotation is not exactly 90°, the spin is not exactly longitudinal at the interaction point: the asymmetry of the scheme guaranties however that the spin is rotated back to the vertical in all cases. This example illustrates a common difficulty of the spin rotators. Energy changes or the need to change the helicity state to avoid systematics requires the geometry of the accelerator to be changed locally. In HERA, where a large flexibility is required, a more involved rotator, the so-called mini-rotator [75], was designed. It is made of sandwiches of horizontally and vertically bending magnets arranged to match the separation of the electron and proton beams on both sides of the interaction point where they collide head-on on the nominal orbits. The displacement of the orbit in its tuning range from 27 GeV to 35 GeV was minimized to 20 cm by a proper sequence of horizontal and vertical rotations. The helicity can be reversed. This flexibility is made possible by installing the spin rotators on movable jacks.

Inside any of the possible rotators, the spin closed solution **n** is bound not to be parallel to the magnetic field of the rotator dipoles. In the latter, the Sokolov-Ternov mechanism tends to polarize the beams along a field direction different from **n**. The result is an ultimate polarization degree lower than the maximum 92.4%. The reduction is quantitatively given by the formula (55) in which the scalar product **b. n** becomes lower than one in the rotator magnets. The polarization decrease can only be minimized by an optimization of the rotator magnet configuration, in particular by using long magnets with low fields. Practical designs allow for polarization levels in excess of 80% at the highest energy. The real difficulty with spin rotators is that, unlike the case of the solenoid compensation, the **n** vector is far from vertical in the spin rotator insertion. This causes the spin-orbit coupling vector **d** not to vanish all along the ring circumference. At the energy of HERA, a single spin rotator reduces

significantly the polarization. At LEP it almost disappears. To overcome this effect, spin transparency must be restored.

4.4.3 Spin transparency conditions

The spin rotator can be considered as a very large defect which causes the spin-orbit coupling integrals $\mathbf{J}_{\pm x,z,s}(s)$ to become large. The conditions for spin transparency are, on the contrary, that, whatever the azimuth s of a photon emission, the spin-coupling integrals integrated from s should vanish. This translates into ten conditions per magnet. Fortunatly, unlike the ring imperfections, the defects introduced by the spin rotator are known. The same conditions are valid for most magnets and optics symmetries decrease further the number of constraints. Furthermore perfect spin transparency entails that all spin resonances are compensated. This would be a council of perfection, as we saw that the depolarization is mainly caused by a limited number of close-by spin resonances. In practice, it is possible to reduce the number of conditions to a manageable number [76]. The spin matching involves an adjustment of the quadrupole gradients to minimize the spin-orbit coupling integrals at least in some energy range, i.e. spin tune range. In LEP the spin transparency conditions were calculated analytically and converted into optics constraints handled by the matching module of the MAD program. Figure 37 shows the polarization calculated before and after spin matching of a possible LEP spin rotator [74].

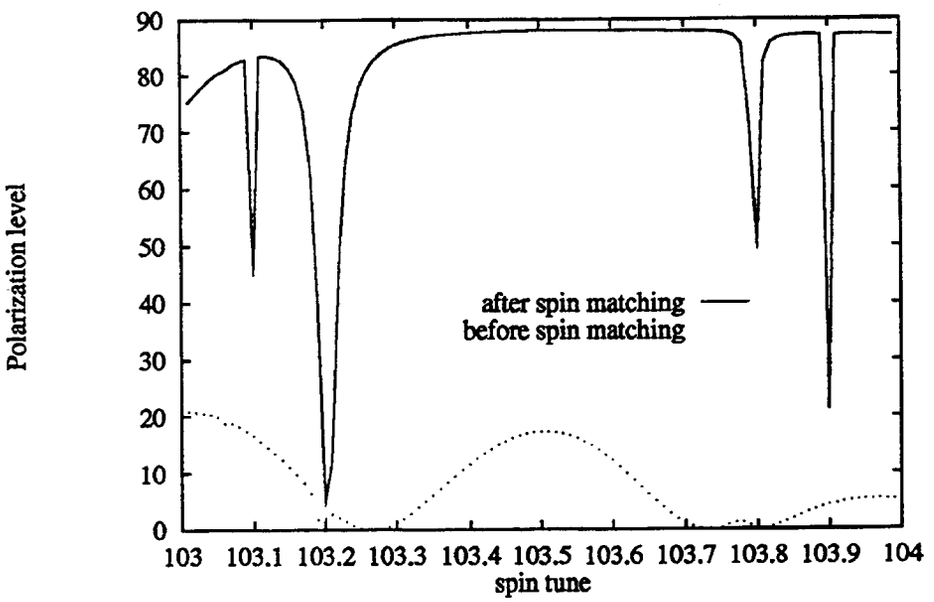


Fig.37: Calculated polarization in LEP equipped with a spin rotator before and after spin matching [74]

For HERA, the dependence of the spin coordinates on the orbital coordinates was computed in the linear approximation. The requirement was then to adjust the focusing so as to minimize the submatrix expressing the spin-orbit coupling. Figure 38 shows the calculated polarization in HERA with the spin-matched rotator [77]. In spring 1994, the HERA spin

rotator was installed and tested. Its spin matching was sufficiently effective to avoid any loss of polarization at a level of 70% [78].

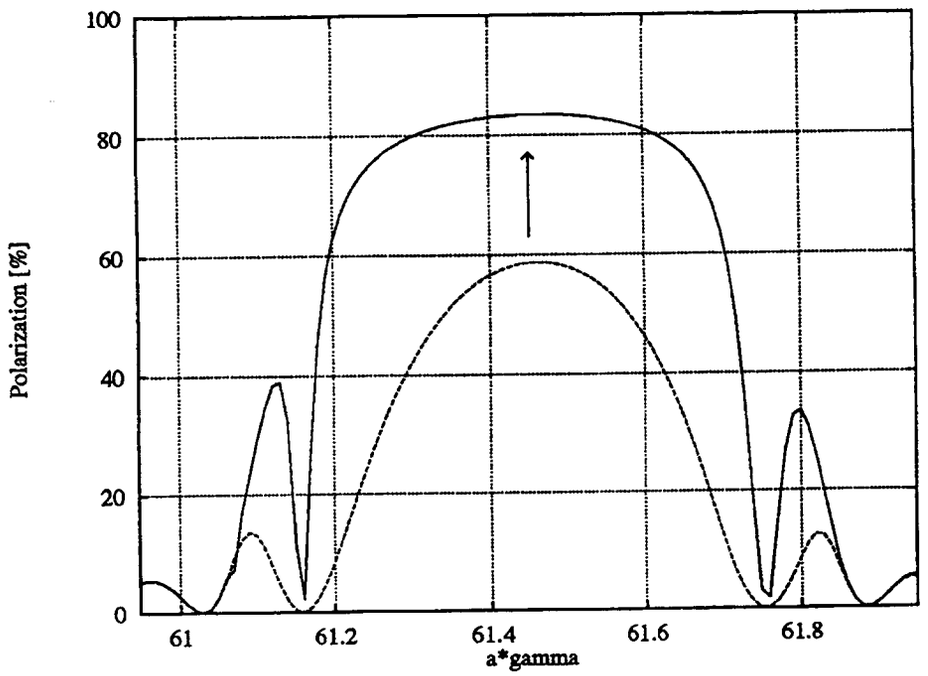


Fig.38: Calculated polarization in HERA equipped with a spin rotator before and after spin matching [77].

4.5 Energy calibration by depolarization on an artificial RF spin resonance

The first application of transverse beam polarization in an electron storage ring is the accurate calibration of the beam energy. This has been carried out in several rings : VEPP2-M [79], BESSY [44], VEPP4 [80], DORIS II [81], CESR [48] and LEP [82]. The method used is based on a RF resonance technique. It can be done even with a modest degree of polarization. The principle is similar to a NMR experiment, although the magnetic energy absorbed by the beam is too small to be directly measured as in NMR experiments.

At LEP the RF resonance technique has been applied to measure the Z mass and width at 44.7, 45.6 and 46.5 GeV corresponding to a spin tunes of 101.5, 103.5 and 105.5. As soon as the polarization reaches about 10%, a small radial RF magnetic field, produced by a kicker, is turned on. The applied field excites a RF spin resonance. The tune value of that resonance is varied by slowly changing the frequency of the RF field. When the scanned range of tune overlaps the spin tune of the beam, the RF resonance is crossed, and the beam is depolarized (Fig. 39). To determine the RF frequency at which depolarization occurs the central frequency is varied and the polarization is continuously monitored with the polarimeter. From that measurement the spin tune and the energy of the beam are deduced. In contrast to a proton beam accelerated in a synchrotron, the spin tune of the beam is fixed and the tune of the RF spin resonance is varied.

The RF magnetic field has a very small integral value, about 4 Gm only. For the spin motion this radial field is a perturbing field which can be analyzed in frequency to find the spin

resonances that it excites, as explained in Section 2.4. The tunes of these RF resonances are given by :

$$v = n \pm \frac{f_{RF}}{f_{rev}} \tag{78}$$

where f_{RF}, f_{rev} are the applied RF and revolution frequencies respectively and n is any integer. The frequency span is typically 112 Hz per minute corresponding to a spin tune range of 0.002 and an energy range of 0.9 MeV. The beam energy E is deduced from the frequency f_{RF} at which depolarization is observed :

$$E = \frac{m_0 c^2}{a}\left(n \pm \frac{f_{RF}}{f_{rev}}\right) \tag{79}$$

after lifting the sign ambiguity (obtained by observing the change of the depolarization frequency when the energy is varied). The limit in accuracy of this calibration method can be as small as 10^{-5}. It is limited by the imperfections of the real storage rings: if the spin rotations are not all about the vertical axis, the spin precession per turn is not anymore strictly proportional to the energy and formula (78) is not exact [83]. The perturbation is less than 0.5 MeV at LEP for a polarization level of 10%. Other limitations are related to non-linearities (sextupoles), interference with other spin resonances,....

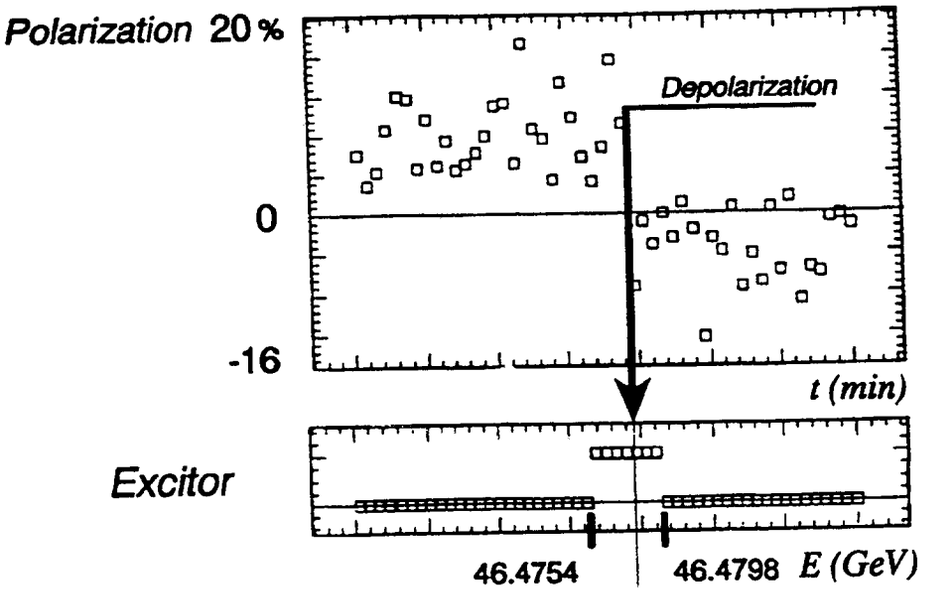


Fig. 39: The polarization versus the time in a calibration experiment at LEP. Depolarization is observed when the RF field (excitor) is turned on with a frequency span corresponding to an energy range (4.4 MeV) overlapping the beam energy E.

The energy calibration by resonant depolarization has also been used to measure the mass of the K mesons [84,85], ϕ meson [86], ψ meson [87] and of the $\Upsilon, \Upsilon', \Upsilon''$, mesons [46,48,79,88,89]. One must also mention the very accurate comparison [90] of the gyromagnetic anomaly of electrons and positrons at VEPP2-M by again a RF technique similar to an NMR experiment showing that the sophisticated spin manipulations in Nuclear Magnetic Resonance can also be used in electron storage rings.

ACKNOWLEDGEMENTS

The authors wish to thank the colleagues of the different laboratories who provided the referenced data and figures. J. P. K. thanks R. Assman, A. Blondel, D. Barber, A. Hofmann and the LEP Polarization Collaboration for many discussions and comments.

BIBLIOGRAPHY

[I] J.S. Bell, Introduction to the concept of polarization in particle beam, Report CERN 75-11 (1975).

[II] A. Chao, Calculation of the electron polarization in the linear approximation (SLIM), Fermilab Summer School 1981, AIP Conf. Proc. n° **87**, p. 395

[III] E.D. Courant and R. Ruth, Acceleration of polarized protons, Report BNL-51270 (1980)
R. Ruth, Proc of the 12th Int. Conf. on High-Energy Accelerators (Fermilab, 1983) p. 286.

[IV] B.W. Montague, General review on polarization in high-energy rings, Physics Reports, Vol. **113** (1984).

[V] A.A. Sokolov and I.M. Ternov, Synchroton radiation and spontaneous polarization, `Radiation from Relativistic Electrons', AIP Translation Series, New-York (1986).

REFERENCES

[1] R. F. Schwitters et al, Phys. Rev. Lett., 35 (1975), 1320.

[2] F. Lehar, Colloq.Phys. (France) n0 C-6 (1990), 19.

[3] A.N. Skrinskii and Yu. M. Shatunov, Sov. Phys. Usp. 32(6) (1989) p. 548.

[4] M. Froissart and R. Stora, Nucl. Instr. Meth., 7 (1960) 297.

[5] V. Bargmann, L.Michel, V.L.Telegdi, Phys. Rev. Lett. 2 (1959) 435.

[6] I.M-.Ternov, Yu. M. Loskutov, L.I. Korovina, Sov. Phys. JETP, 14 (1962), 921.

[7] A. A. Sokolov and I. M.Ternov, Sov. Phys.-Doklady, 8 (1964) 1203.

[8] R. Belbéoch et al, in Proc. USSR Nat. Conf. Particle Accelerators (1968) p. 129.
[9] The Orsay Storage Ring Group, in Proc. 8th Int. High Energy Accelerators CERN, Geneva, 1971) p. 127.
[10] V. N. Baier, Sov. Phys.-USPEKHI, 14 (1976) 1063.
[11] S. I. Serednyakov, Sov. Phys.-JETP, 44 (1972) 695.
[12] T. Khoe et al, Part. Accel., 6 (1975) 213.
[13] T. Aniel et al., Colloq.Phys. (France) N0 C-2 (1985), 499.
[14] P. A. Chamouard et al., Colloq. Phys. (France) n0 C-6 (1990), 569.
[15] F. Z. Khiari et al., Phys. Rev. D, 39 (1989) 45.
[16] W. Brefeld et al, Nucl. Instr. Meth., 228 (1985) 228.
[17] C. Y. Prescott et al., Phys. Lett., 77B (1978) 347.
[18] C. Y. Prescott, SLAC-PUB-6242 and Proc. 2nd Intl. W. on Physics and Exp.at Linear Colliders, Waikoloa, Hawaï (1993)
[19] D.P. Barber et al. Nucl. Instr. Meth. A338 (1994), 166.
[20] K.Nakajima et al, Proc. 3rd EPAC, vol.1 (1992), 919.
[21] L. Knudsen et al, Phys. Lett., 270B (1991) 97.
[22] E. D. Courant and R. D. Ruth, BNL Report, 51270 (1980).
[23] R. D. Ruth, IEEE Trans. Nucl. Science, 30 (1983) 286.
[24] Tkatchenko, 1990 EPAC Conf., ed. P. Marin and P. Mandrillon (Nice,1990) p. 159.
[25] A. W. Chao, AIP Conf. Proc., 87 (1981) 395.
[26] A. W. Chao, IEEE Trans. Nucl. Science, 30 (1983) 2383.
[27] D. P. Barber, 1990 EPAC Conf., Ed. P. Marin and P. Mandrillon (Nice,1990) p. 164.
[28] Particle Data Group, J. J. Hernandez et al., Phys. Lett. B, 239 (1990).
[29] CODATA Recommended values, Journ. of Research of NBS, 92 (1987) 85.
[30] J. S. Bell, Report, CERN 75-11 (1975).
[31] L. T. Thomas, Phil. Mag. 3 (1927) 1.
[32] J. D. Jackson, Classical Electrodynamics (J. Wiley and sons, 2nd ed.,1975) p.541.
[33] Ya. S. Derbenev and A. M. Kondratenko, Sov. Phys.-Dokl., 20 (1976) 562.
[34] Ya. S. Derbenev et al, Part. Acc., 8 (1978) 115.
[35] K. Steffen, DESY HERA report, 82/02 (1982).
[36] J.P.Koutchouk, Proc. 8th Symp.High Energy Spin Phys. vol. 2 (1988), 1004.
[37] N. Horikawa, in Proc. 9th Int. Symp. on High Energy Spin Physics, ed. K.-H. Althoff and W. Meyer (Springer-Verlag, 1991) p. 172.
[38] A. D. Krisch, Phys. Rev. Lett., 63 (1989) 1137.
[39] J. E. Goodwin, Phys. Rev. Lett., 64 (1990) 2779.
[40] P. A. Chamouard, private communication.
[41] Ya. S. Derbenev and A. M. Kondratenko, Novosibirsk preprint, 78-74 (1978).

[42] J. D. Jackson, Rev. Mod. Phys., 48 (1976) 417.
[43] J. S. Bell and J. M. Leinaas, Nucl. Phys. B, 284 (1987) 488.
[44] K. Ott, private communication and BESSY Annual Report 1988.
[45] J. R. Johnson et al, Nucl. Instr. Meth., 204 (1983) 261.
[46] A. S. Artamonov et al, Phys. Lett., 118B (1982) 225.
[47] D. P. Barber, DESY report, M-83-29 (1983).
[48] W. Mackay et al, Phys. Rev., 29 (1984) 2483.
[49] H. D. Bremer et al, DESY report, 82-026 (1982).
[50] K. Nakajima et al, Paper TUS04A at the 1992 EPAC Conf. (Berlin 1992).
[51] R. Assmann et al., CERN Report, CERN SL 94-08 (AP), (1994).
[52] V.N. Baier et al., Sov. Phys. JETP, 31 (1970), 908.
[53] J. Buon, LAL/RT/83-01 (1983).
[54] B. W. Montague, Physics Reports, 113 (1984)
[55] M. Sands, SLAC-121, (1970), 118.
[56] Ya. S. Derbenev and A. M. Kondratenko, Sov. Phys.-JETP, 37 (1973) 968.
[57] L. N. Hand and A. Skuja, Phys. Rev. Lett., 59 (1987) 1910.
[58] D. P. Barber and S. R. Mane, Phys. Rev. A, 37 (1988) 456.
[59] J. Buon, LAL/RT report, 84-05 (1984).
[60] A. Chao, K. Yokoya, DESY M-82/09, (1982), K9, R19.
[61] A. W. Chao, Nucl. Instr. Meth., 180 (1981) 29.
[62] J. Kewisch et al., Phys. Rev. Lett. 62, (1989), 419.
[63] H. Grote, CERN Report, SL-DI 91-20, (1991).
[64] Ya. Derbenev et al., Part.Accel. 9, (1972), 247.
[65] J. Buon, Part. Acc., 32 (1990) 153.
[66] S. R. Mane, Fermilab report, FN-466 (1987).
[67] K. Yokoya, KEK Report 92-6, (1992).
[68] H. Grote, CERN Report SL/92-56 (AP), (1992).
[69] Yu. Eidelmann, V. Yakimenko, Proc. 1991 IEEE Part. Conf. , San Francisco, (1991), 269.
[70] R. Assman, Proc. 3rd Work. on LEP Perf., CERN SL/93-19 p.341 (1993)
[71] D.P. Barber et al., DESY 85-044, (1985).
[72] A. Blondel, LEP Note 629, (1990).$
[73] V. N. Baier, Sov. Phys.-USPEKHI, 14 (1972) 695.
[74] H. Grote et al, 1990 EPAC Conf., Ed. P. Marin and P. Mandrillon (Nice,1990),1 497.
[75] J. Buon and K. Steffen, Nucl. Instr. Meth., A 245 (1986) 248.
[76] J. Buon, Nucl. Instr. Meth., A 275 (1989), 226.
[77] D. Barber, M. Boege, E. Gianfelice, private communication.

[78] D. Barber et al., to be published.

[79] Ya. S. Derbenev et al, Part. Acc., 10 (1980) 177.

[80] A. E. Bondar et al, in Proc. 12th Int. Conf. on High Energy Accelerators, ed. T. Cole and R. Donaldson (Fermilab 1983) p. 240.

[81] D. P. Barber et al, Phys. Lett., 135B (1984) 498.

[82] L. Arnaudon et al,CERN Report, SL/92-16 (1992), to be published in Phys. Lett. B, (1992).

[83] R. Assmann, J.P. Koutchouk, CERN SL/()ç-13 (AP) (1994).

[84] L. M. Barkov et al, Nucl. Phys., B148 (1979) 53.

[85] L. M. Barkov et al, Novosibirsk preprint, 83-85 (1983).

[86] A. D. Bukin et al, Yad. Fiz., 27 (1978) 976.

[87] A. A. Zholents et al, Phys. Lett., 96B (1980) 214.

[88] A. S. Artamonov et al, Phys. Lett., 137B (1984) 272.

[89] S. E. Baru et al, Z. Phys. C, 30 (1986), 551.

[90] I. B. Vasserman et al, Phys. Lett., 198 (1987), 302.

SPACE CHARGE DOMINATED BEAM TRANSPORT

I. Hofmann
GSI Darmstadt, D-64220 Darmstadt 11

Abstract
We consider beam transport systems where space charge forces are comparable in strength with the external focusing force. Space charge then plays an important role for beam transmission and emittance growth. We use the envelope model for matching and the generalized field energy equations to study emittance growth. Analytic results are compared with numerical simulation.

1 INTRODUCTION

In high-current linear accelerators and in transport systems for protons or heavier ions the repulsive force due to the space charge carried by the beam itself plays an important role for the design of the focusing system and for conservation of beam emittance. In proton or heavy ion linear accelerators the actual space charge bottle-neck is at injection, where the ion is slow and correspondingly the space charge density and resulting forces large. Much can be learnt from transport experiments [1, 2, 3], which have been performed to study emittance growth under stationary conditions, i.e. no acceleration. Interest in these transport experiments was largely stimulated by the idea of using heavy ions for inertial confinement fusion [4], which requires transport of large currents over long distances. We emphasize, however, that the role of space charge for emittance growth is equally important for high-current proton or heavy ion linacs. One way of discussing emittance growth is by relating it to the electrostatic energy of the space charge distribution ("nonlinear field energy"). This concept was originally derived in Refs. [5, 6],

for 2-D beams. The analytical theory can be generalized to the 3-D case [7], but we confine this lecture to the 2-D problem of beam transport of long beams, ignoring the longitudinal degree of freedom.

2 BASIC PROPERTIES

2.1 Incoherent Effect of Space Charge

For comparison we recall that in circular accelerators space charge is always a relatively weak perturbation described by the betatron tune shift $\Delta\nu$

$$\frac{d^2x}{d\Theta^2} = -(\nu_0 - \Delta\nu)^2\, x = -\nu^2 x \qquad (1)$$

where ν_0 is the betatron tune in the absence of space charge and $\Delta\nu < \frac{1}{4} << \nu_0$ in order to use a gap free of machine resonances. The corresponding slight increase in betatron wavelength is contrasted by a large effect in high-current linear accelerators, where one tries to compensate the external focusing force (given by ν_0^2) as much as possible by space charge, hence $\nu^2 < \nu_0^2$ (Fig. 1).

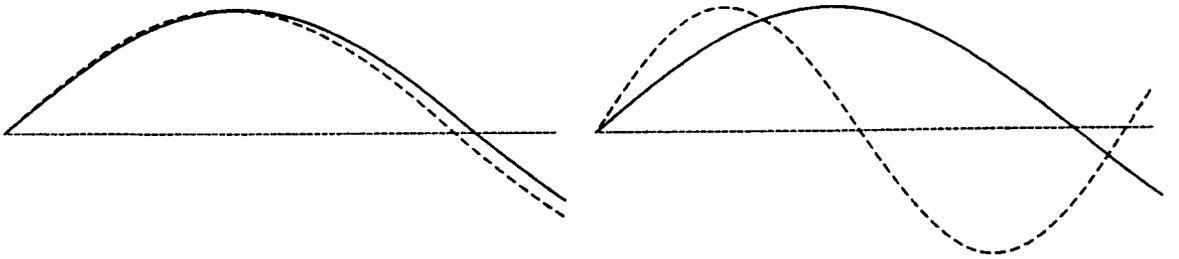


Figure 1: Space charge effect on betatron oscillation in circular accelerators (left) and high-current linear accelerators (right).

2.2 Coherent Effects

In practice the ideal space charge limit $\nu = 0$ with an exact balance between the applied focusing and the defocusing force due to space charge can be approached only to a limited extent. With decreasing ν/ν_0 the beam is increasingly dominated by collective behaviour, which can lead to emittance growth. Coherent modes of oscillation have been studied analytically for the highly specialized Kapchinskij-Vladimirskij distribution [8]. For more realistic beam models one depends on computer simulation, which shows various modes of emittance growth due to such coherent oscillations (reviewed in Ref. [9]). It will be shown below that in a periodic focusing system coherent modes of oscillation can be in resonance with the focusing period. "Harmful" structure resonances can be largely avoided by choosing sufficiently small values for σ_o.

2.3 Space Charge Limited Current

Following Maschke [10] the relationship between beam current and σ/σ_0 for a periodic quadrupole channel can be expressed in the following way [11]:

$$I = 3.66 \cdot 10^6 (\frac{A}{Z})^{1/3} B_0^{2/3} (\beta\gamma)^{7/3} \varepsilon^{2/3} F(\sigma/\sigma_0) \tag{2}$$

with B_0 the quadrupole pole-tip field, ε the transverse emittance and $F(\sigma/\sigma_0)$ a factor, which depends only weakly on the type of focusing (see Fig. 2 for a symmetric quadrupole channel). Equation (2) is the condition that the beam is matched to the transport channel. For small σ/σ_0 the dependence of beam current on σ for a given channel can be seen more directly from the approximately valid scaling [12]

$$I \sim a^2 \sim \frac{\varepsilon}{\sigma} \tag{3}$$

where a is the channel radius. Knowledge of the minimum σ for stable transport is thus essential for optimum current transmission. This will be studied in more detail further below.

3 MATCHING WITH UNIFORM SPACE CHARGE

The defocusing effect of space charge has to be compensated by an appropriate change of the α- and β-functions at the entrance to a transport channel. For large space charge it is often necessary to increase the quadrupole strength to avoid hitting the aper-

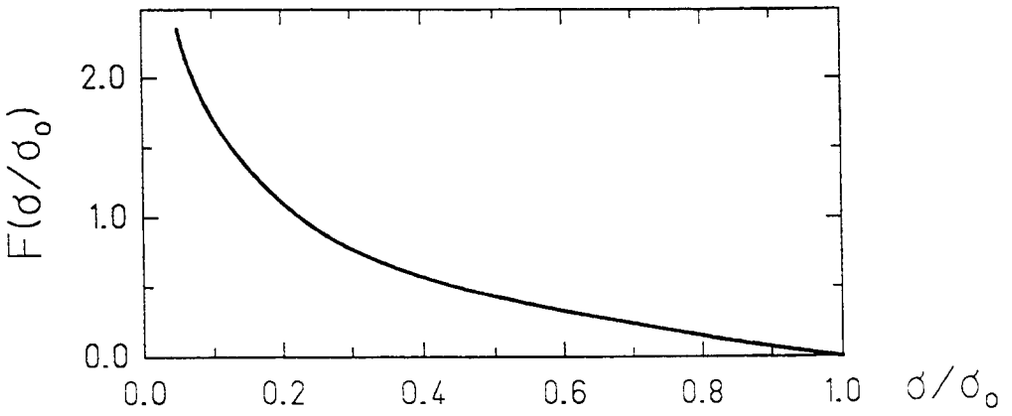


Figure 2: Current transmission factor as function of σ/σ_0.

ture. For uniform space charge distribution within an elliptic boundary the repulsive Coulomb force varies linearly with distance from the beam center:

$$E_x = \frac{I}{\pi\varepsilon_0 v_0} \frac{x}{a(a+b)} \tag{4}$$

and

$$E_y = \frac{I}{\pi\varepsilon_0 v_0} \frac{y}{b(a+b)} \tag{5}$$

with $v_0 = \beta c$ and a, b the transverse semi-axi (i.e. envelopes). As is shown in the appendix, following the derivation by Sacherer, [13] one can then describe the beam envelopes by the equations

$$\frac{d^2}{ds^2}a + k_x(s)a - \frac{\varepsilon_x^2}{a^3} - \frac{qI}{\pi\varepsilon_0 m\gamma^3 v_0^2(a+b)} = 0 \tag{6}$$

and similarly

$$\frac{d^2}{ds^2}b + k_z(s)b - \frac{\varepsilon_z^2}{b^3} - \frac{qI}{\pi\varepsilon_0 m\gamma^3 v_0^2(a+b)} = 0, \tag{7}$$

with $\varepsilon_x, \varepsilon_z$ the emittances and I the current. These equations are the basis for matching of a beam to a transport channel in the presence of space charge.

As an example we consider a symmetric FODO transport channel with the periodic cell 4 m long and a beam of protons at 100 MeV with $\varepsilon = 10^{-6}\pi$m-rad. A matched envelope for vanishing space charge ($I = 0$) is shown in Fig. 3 (top), where the upper trace is the horizontal, and the lower trace the vertical envelope. The trajectory of the test particle shown gives a full oscillation over six periodic cells, i.e. $\sigma_0 = 60^0$. We also show the increased envelope and reduced betatron phase advance σ for large currents, but focusing gradients unchanged. Results are summarized in Fig. 4, along with the increase of β-function values (here at the center of a drift section) for decreasing σ.

For an aperture limited beam transport it is necessary to increase the focusing gradients. In Fig. 5 we have used the same currents as in Fig. 3, but increased the

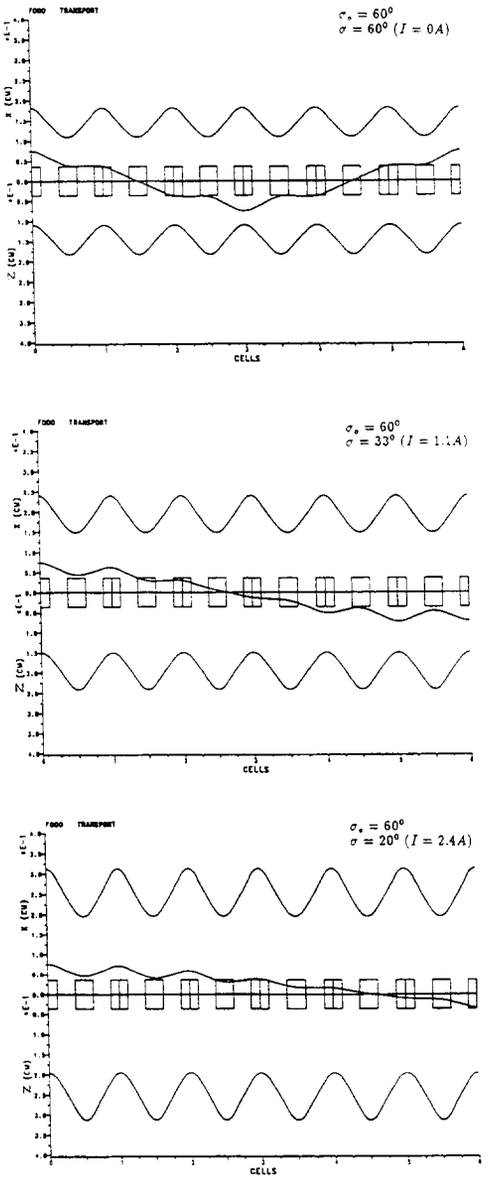


Figure 3: Space charge effect on matched envelopes in symmetric FODO channel with $\sigma_0 = 60^0$.

quadrupole gradients (shown by the height of square boxes) so that $\sigma = 60^0$ is maintained for the depressed tune. As a consequence the average envelope has about the same size as in the zero-current case of Fig. 3, whereas the excursions of the envelope increase with current.

4 EMITTANCE AND FIELD ENERGY

4.1 General Principle

We are interested to understand under what conditions and to what extent the emittance can grow. Our goal is to predict emittance growth from general principles - as much as possible - without actually computing the dynamical process responsible for it. Here we present an approach, which allows a qualitative discussion of the various sources of emittance growth, as well as quantitative estimates for the expected growth in special cases. The general idea of this approach is to consider emittance growth as an increase

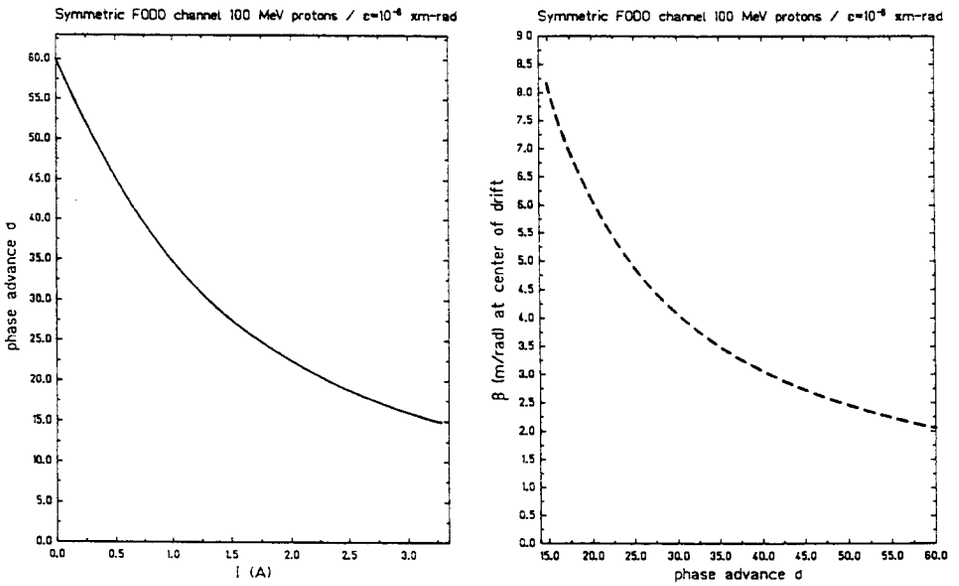


Figure 4: Phase advance σ as function of I and β-function as function of σ for fixed $\sigma_0 = 60^0$.

of thermal energy, which must be taken from the overall available beam energy. If we succeed in estimating the available source of energy we can also estimate the maximum possible emittance growth.

We assume the total energy of the beam can be written in the following way

$$E_{total} = E_0 + E_{th,x} + E_{th,z} + E_{ext} + E_{field} \tag{8}$$

where

x, z	= transverse coordinates
E_0	= forward kinetic energy
$E_{th;x,z}$	= thermal energy in x, z (equivalent to emittance)
E_{ext}	= potential energy due to externally applied focusing force
E_{field}	= energy of space charge induced electric field

A growth of thermal energy (in x or z) and thus emittance is possible at the expense of either E_0 or E_{field}. Due to the fact that the beam is always at a minimum of the external potential (in time average), E_{ext} is not a source of energy available for emittance growth and we expect E_{ext} also to increase during growth of the emittance. In addition, it is possible that thermal energy is transferred between the transverse plane (x, z) and the longitudinal direction (y), if there is a significant initial imbalance. This important subject ("equipartitioning") is treated in Ref. [14], with an analytical model in Ref. [15]. In all cases E_{field} plays a crucial role. It should be noted here that in a constant focusing system $E_{th} + E_{ext} + E_{field}$ is constant, hence we have no coupling with E_0. In periodic focusing, on the other hand, such a coupling can be significant as will be shown below. This subject has been treated in detail in Ref. [8].

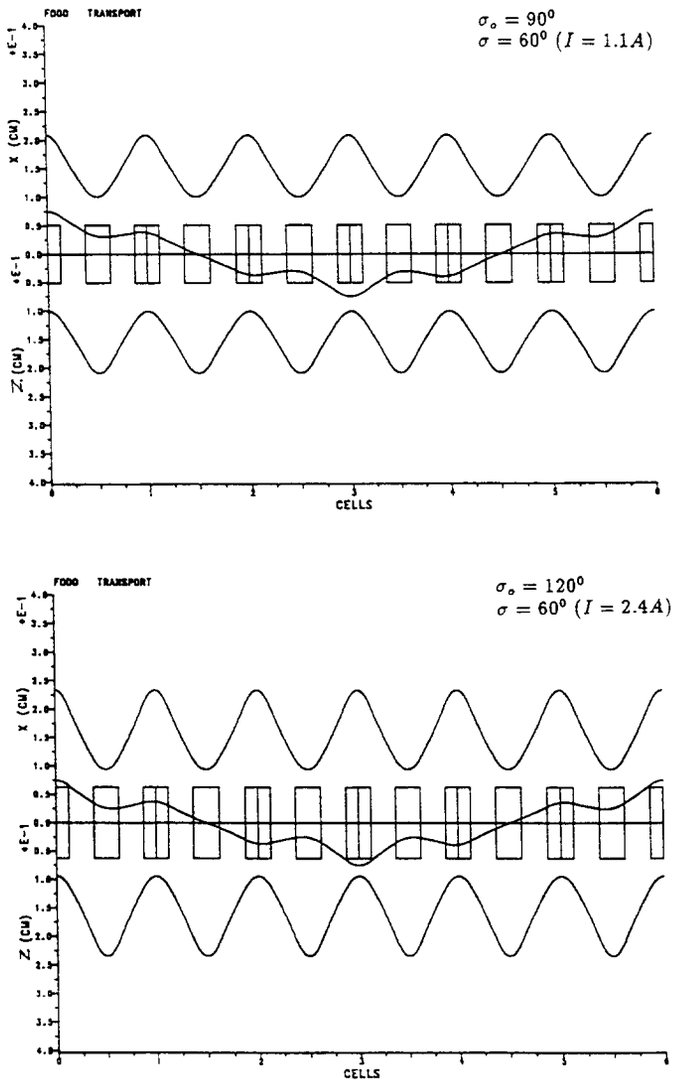


Figure 5: Increased focusing gradients keeping constant $\sigma = 60^0$, which results in roughly constant average envelope.

4.2 Generalized Emittance Equation

It can be shown that the emittances $\varepsilon_x, \varepsilon_z$ for the two transverse planes x, z are related to the field energy by the following equation, valid for constant or periodic focusing (the derivation is given in Appendix A; see also Refs. [5, 6, 7]:

$$\frac{1}{\overline{x^2}} \frac{d}{ds} \varepsilon_x^2 + \frac{1}{\overline{z^2}} \frac{d}{ds} \varepsilon_z^2 = -4K \frac{d}{ds} \frac{W - W_u}{w_0} \tag{9}$$

with

$$K = \frac{Nq^2}{2\pi\varepsilon_0 m\gamma^3 v_0^2} \quad (generalized \ \ perveance)$$

$\overline{x^2}, \overline{z^2}$ mean squares of x, z

W actual field energy
W_u field energy of equivalent uniform density beam
w_0 normalization constant (= field energy inside actual beam boundary of a uniform beam)

The significance of this equation is that the change of emittance is related to a change of the "nonlinear field energy" $W - W_u$, i.e. the excess field energy due to the non-uniformity of the density. A uniform density beam — even with large mismatch — therefore has constant emittance.

4.3 Minimum Field Energy of Uniform Density Beam

A practically important property of the field energy is that it adopts its minimum for a uniform density beam. This can be shown by defining a variational expression

$$S \equiv W + \alpha_1 x^2 + \alpha_2 z^2 \tag{10}$$

with α_1, α_2 Lagrangian multipliers to keep the r.m.s. size constant. For a minimum we require the variation of S to vanish, hence

$$\delta S = \int\int \left[\varepsilon_0 E \delta E + N^{-1}\left(\alpha_1 x^2 + \alpha_2 z^2\right)\delta n\right] dx\ dz. \tag{11}$$

The density perturbation is defined by an arbitrary displacement

$$\delta n(x,z) = \nabla n \cdot \delta \mathbf{x}, \tag{12}$$

thus we obtain

$$\delta S = \int\int \left[\phi + N^{-1}\left(\alpha_1 x^2 + \alpha_2 z^2\right)\right](\nabla n \delta \mathbf{x})\, dxdz = 0 \tag{13}$$

requiring either $\phi = -N^{-1}(\alpha_1 x^2 + \alpha_2 z^2)$ (interior of beam, hence uniformly charged ellipsoid) or $n = const. = 0$ (exterior).

4.4 Field Energy of Different Beam Models

Calculation of the field energy for a parabolic density profile

$$n = \frac{2N}{\pi ab}\left(1 - \frac{x^2}{a^2} - \frac{z^2}{b^2}\right) \tag{14}$$

leads to the following result

$$W = \frac{N^2 q^2}{16\pi\varepsilon_0}\left(\frac{11}{6} - 4\ \ln\sqrt{6} + 4\ \ln\frac{2R}{\tilde{x}+\tilde{z}}\right) \tag{15}$$

with the r.m.s. envelope $\tilde{x} = a/\sqrt{6}$ and $\tilde{z} = b/\sqrt{6}$. Due to the minimum field energy property of a uniform beam of the same r.m.s. size it is convenient to calculate the difference energy in normalized units:

$$\frac{W - W_u}{w_0} = \frac{5}{6} - 4\ \ln\frac{\sqrt{6}}{2} \simeq 0.0224. \tag{16}$$

This normalized "nonlinear field energy" hence is always positive according to the previous section. For a Gaussian density profile truncated at 4 standard deviations one obtains the considerably larger value of 0.154.

4.5 Stationary Distributions and Debye Shielding

It must be noted that a Gaussian phase space distribution is consistent with a Gaussian density profile only in the low-current limit. From computer simulation one finds that practically any phase distribution functions leads to uniform density in the limit of high current. In Fig. 6 this is shown for a Gaussian distribution.

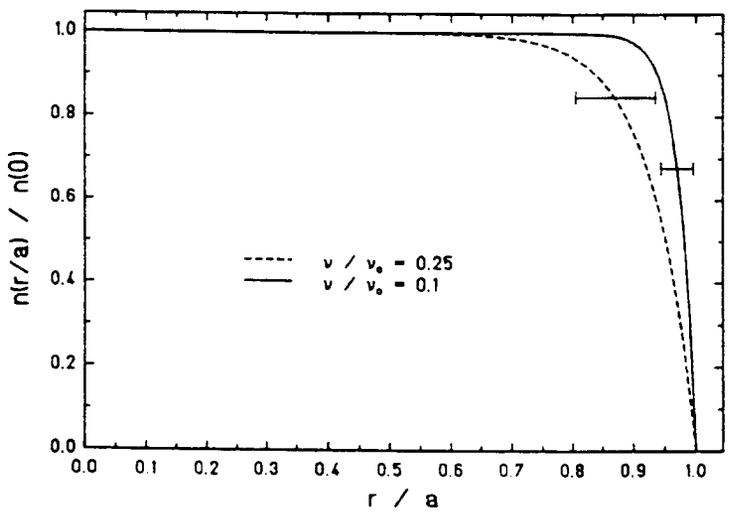


Figure 6: Density profiles for Gaussian phase space distribution near space charge limit. Bars indicate λ_D according to Eq. (20).

This is readily shown for a Gaussian distribution ($r^2 = x^2 + z^2$)

$$f \sim \exp\left[-\left(\frac{x'^2 + z'^2}{2} + \frac{k}{2}\, r^2 + \frac{q}{m\gamma^3 v_0^2}\phi\right)/\mu\right] \tag{17}$$

which, after insertion into Poisson's equation and expansion of the exponential yields

$$n \simeq n_0\left[1 - \left(\frac{a}{r}\right)^{1/2} \exp\left((r-a)/\lambda_D\right)\right] \tag{18}$$

provided that the total potential in Eq. (17) is small as compared with the average transverse energy μ. This condition can be expressed in terms of the Debye-length λ_D, i.e.

$$\frac{\lambda_D}{a} << 1 \tag{19}$$

where λ_D is given by

$$\lambda_D^2 = \frac{\mu}{\omega_p} \simeq 2\frac{\overline{x'^2}}{\omega_p^2} \simeq \frac{v_{th}^2}{\omega_p^2} \tag{20}$$

We can also express λ_D/a in terms of the tune depression

$$\frac{\lambda_D}{a} \simeq \frac{1}{\sqrt{8}} \frac{\nu}{\nu_0} \tag{21}$$

where ν/ν_0 is replaced by σ/σ_0 for periodic focusing.

The physical meaning of the Debye-length is the one familiar in plasma physics: a high-current beam shields the external focusing force from its interior by building up a uniform density with a space charge force that cancels the focusing force. The shielding is ineffective in the boundary layer of a thickness given by λ_D (Fig. 6). Hence, for small ν/ν_o there is practically no restoring force in the interior of the beam and particles are "reflected" by the uncompensated force in the boundary layer.

5 APPLICATION TO EMITTANCE GROWTH

Using Eq. (9) we can now discuss the possible sources for emittance growth. The relationship between emittance and thermal rsp. kinetic energy suggests the analogy with a ball in a potential trough. The following situations can be envisaged:

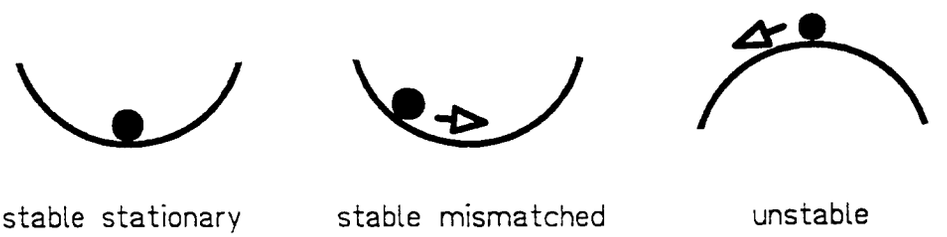


Figure 7: Analogy with ball in a potential trough

5.1 Stable Stationary Beam

This requires that the emittances ε_x and ε_z are conserved along the channel and the envelopes in x and z vary periodically. According to Eq. (9) a necessary condition for this is that $W - W_u$ is a constant. This can be satisfied if the beam density profile remains self-similar. Such an observation has been made in numerical simulation; an analytical proof for beams with non-uniform density and periodic focusing is given in Ref. [16].

5.2 Mismatched Beam

An initial excess field energy is transferred into emittance according to Eq. (9). For a round beam and constant focusing we can integrate Eq. (9) and find

$$\frac{\Delta\varepsilon^2}{\varepsilon^2} = -\frac{1}{2}\left(\frac{\nu_0^2}{\nu^2} - 1\right)\Delta U \tag{22}$$

with $U \equiv (W - W_u)/w_0$ the normalized nonlinear field energy.

As an example we inject a beam with Gaussian profile into a channel with $\nu_0/\nu = 6$. Since we know from equation (14) that close to the space charge limit a stationary

solution has nearly uniform density, nearly all of the excess nonlinear field energy (0.15 in normalized units) is found as emittance growth during less than a betatron oscillation. Hence we can estimate the predicted emittance growth by ignoring the final nonlinear field energy and thus obtain

$$\frac{\Delta\varepsilon^2}{\varepsilon^2} \leq 2.6 \tag{23}$$

which is equivalent to an emittance growth of 90 % and agrees very well with the result of simulation. This is shown in Fig. 8a for a multiparticle computer simulation in a periodic solenoid channel. We observe that the mismatch emittance growth formula applies equally well to periodic focusing also - growth factors found in simulation are practically identical with those in constant focusing. It is only the ratio ν_0/ν rsp. σ_0/σ, which is important.

A different result is obtained if a strictly uniform beam is injected, hence we start at the minimum field energy. Such a beam has the tendency of smoothing its sharp boundary, which requires a small amount of field energy leading to a small decrease in r.m.s. emittance. This is again confirmed by simulation (Fig. 8b). We observe that such an r.m.s. emittance decrease is not violating Liouville, which still applies to the four-dimensional phase space. We thus conclude that injection of as uniform a density as possible minimizes emittance growth of high-current beams.

In this context we also need to mention the possibility of an initial envelope mismatch, which might result from an improper entrance matching. The associated "mismatch energy" can be translated into emittance growth if the beam density is nonuniform. This case is considered in more detail in Ref. [17].

5.3 Coherent Instabilities

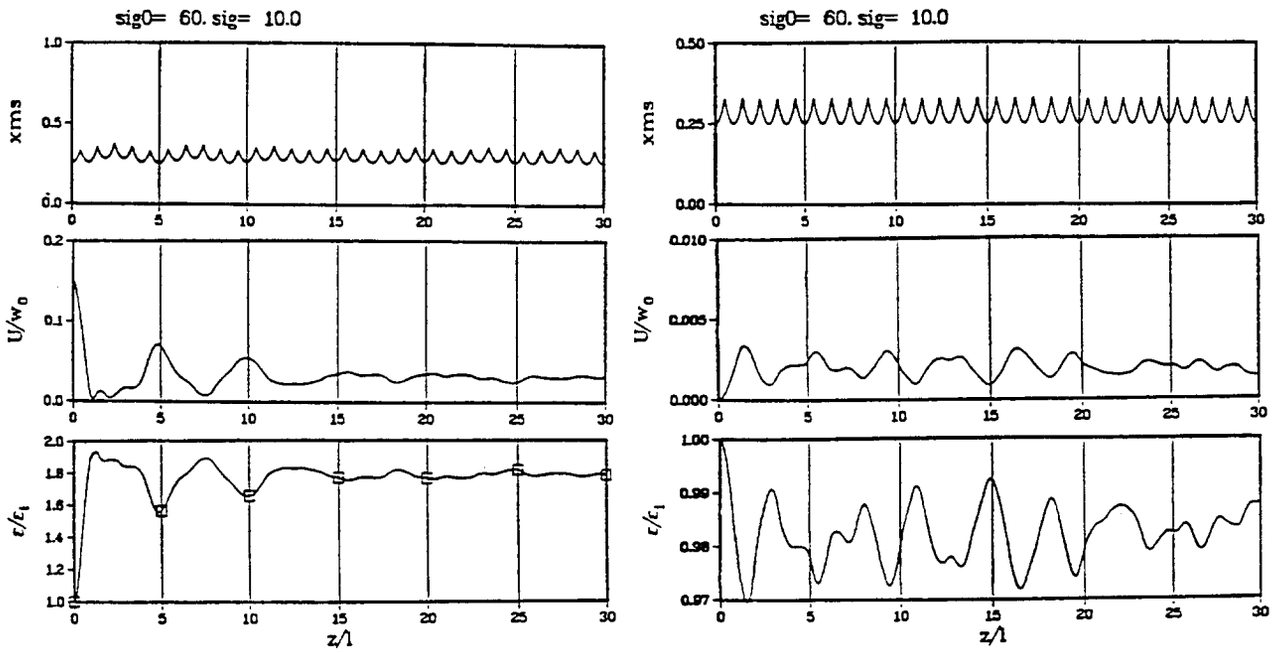
In linear beam transport, coherent instabilities occur due to a local interaction of an ensemble of particles via the space charge force. This is in contrast with beams in circular accelerators, where coupling to the surrounding structure or to other bunches plays a dominant role.

In the present framework coherent instabilities can be discussed qualitatively by writing Eq. (9) for a round beam in periodic solenoidal focusing ($k_x = k_z$).

$$\frac{d}{ds}\varepsilon^2 = -2K\,\overline{x^2(s)}\frac{d}{ds}\,\frac{W - W_u}{w_0}. \tag{24}$$

We assume a matched envelope, hence $\overline{x^2}$ oscillates with the focusing periodicity. Furthermore, we assume a small coherent oscillation of the beam, such that $W - W_u$ oscillates with the period of oscillation of the particular mode. We now anticipate a steadily growing ε^2, if $\overline{x^2}$ and $W - W_u$ oscillate with the same frequency (phase shifted by $\sim 90^0$). In this case the r.h.s. of Eq. (24) has a non-oscillatory part, which adds up to a finite emittance growth until the resonance condition is lost by detuning. The resulting "structure resonance" is thus a coupling between the zero-order oscillatory envelope and a non-uniform coherent mode of oscillation as illustrated by the example in Fig. 9 of a round beam in periodic focusing with $\sigma_0 = 120^0$ and $\sigma = 10^0$.

A direct calculation of the emittance growth is not possible from Eq. (24); this would require to determine the actual time dependence of $W - W_u$, which is not possible analyt-



(a) (b)

Figure 8: Squared envelope (top), nonlinear field energy U/w_0 (center) and emittance growth factors (bottom) versus number of focusing periods obtained from multiparticle computer simulation in a periodic solenoid channel with $\sigma_0 = 60^0, \sigma = 10^0$. (a) Gaussian density profile, (b) Semi-Gaussian: uniform density profile and Gaussian velocity distribution.

ically. For illustration of the principle of a structure resonance we can, however, derive a scaling expression for the emittance growth by making a simple harmonic approximation

$$\overline{x^2(s)} \simeq \hat{x}^2 + \delta x^2 \sin(\omega s) \tag{25}$$

$$U \simeq \hat{U} + \delta U \cos(\omega s) \tag{26}$$

where $\omega = 2\pi/L$ and L the focusing period. We then find from Eq. (24)

$$\frac{d}{ds}\varepsilon^2 \simeq 2K\delta x^2 \omega \ \delta U \sin^2(\omega s) \tag{27}$$

In smooth approximation K follows from the envelope equation (with ε_i the matched initial emittance)

$$K = \frac{1}{4}\frac{\varepsilon_i^2}{\hat{x}^2}\left(\frac{\sigma_0^2}{\sigma^2} - 1\right). \tag{28}$$

We obtain

$$\frac{\varepsilon}{\varepsilon i} \leq \left[1 + \frac{\pi}{2}\left(\frac{\sigma_0^2}{\sigma^2} - 1\right)\frac{\delta x^2}{\hat{x}^2}\delta U \frac{s}{L}\right]^{1/2}, \tag{29}$$

where we have used the approximation $\int \sin^2(\omega s) ds \approx \frac{S}{2}$.

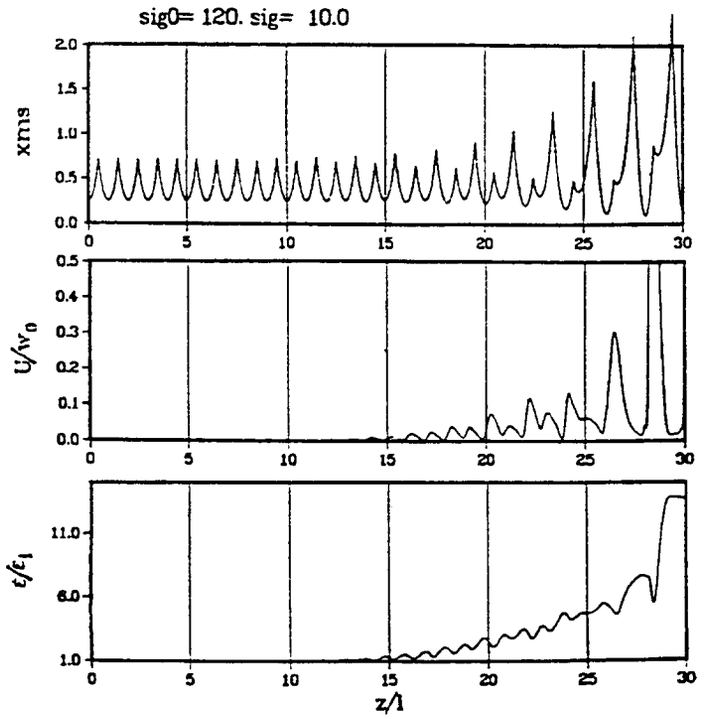


Figure 9: Structure resonance in computer simulation of initially uniform beam (KV-distribution) in periodically interrupted solenoid focusing.

From this expression we conclude that in an advanced stage the emittance rise is weaker than linear in time. It is enhanced by small σ and large envelope excursions as in strong focusing. In the example of Fig. 9 the excited mode has an octupole type field perturbation, which grows from noise on the initially uniform beam (KV-distribution) and saturates between focusing period 15 and 20. In this region the average normalized nonlinear field energy δU is estimated as 0.02 and Eq. (29) yields $\varepsilon/\varepsilon_0 \simeq 3$ which agrees reasonably well with the numerical result. After focusing period 20 the situation becomes very nonlinear: a different mode with half the oscillation frequency starts growing and leads to further strong emittance growth.

These particular modes can be avoided if $\sigma_0 \leq 90^0$, in which case they are out of resonance. As was shown in Ref. [8] another type of mode can appear for $\sigma_0 = 90^0$ and $\sigma \simeq 45^0$ (sextupole type field perturbation). This mode, however, was found to be practically negligible in its effect on emittance growth. This result was also confirmed by experiment [1].

References

[1] A. Faltens et.al., Proc. of the 1984 Linear Accelerator Conference, Seeheim, Germany, p. 312 (1984)

[2] J. Klabunde et. al., ibid., p. 315 (1984)

[3] M. Reiser et. al., ibid., p. 309 (1984)

[4] see Proc. of the INS International Symposium on Heavy Ion Accelerators and Their Applications to Inertial Fusion, INS-Report, Tokyo, Jan. 23-27, 1984 (1984)

[5] J. Struckmeier, J. Klabunde and M. Reiser, *Part. Accel.* 15, 47 (1984)

[6] T.P. Wangler et. al., *IEEE Trans. Nucl. Sci.* NS-32, p. 2196 (1985)

[7] I. Hofmann and J. Struckmeier *Part. Accel.* 21, 69 (1987)

[8] I. Hofmann, L.J. Laslett, L. Smith and I. Haber, *Part. Accel.* 13, 145 (1983)

[9] I. Hofmann in *Applied Charged Particle Optics* (ed. A. Septier), part C, Academic Press, N.Y., p. 99 (1983)

[10] A.W. Maschke, Proc. ERDA Summer Study of Heavy Ions for Inertial Fusion, Berkeley, 1976.

[11] G.R. Lambertson et. al., *IEEE Trans. Nucl. Sci.* NS-24, 933 (1977)

[12] M. Reiser, *Part. Accel.* 8, 167 (1978)

[13] F.J. Sacherer, *IEEE Trans. Nucl. Sci.* NS-18, 1105 (1971)

[14] R.A. Jameson, *IEEE Trans. Nucl. Sci.* NS-28, 2408 (1981)

[15] I. Hofmann, *IEEE Trans. Nucl. Sci.* NS-28, 2399 (1981)

[16] J. Struckmeier and I. Hofmann, *Part. Accel.* 39, 219 (1992)

[17] M. Reiser, *J. Appl. Phys.* 70, 1919 (1991)

A APPENDIX

A.1 Derivation of R.M.S. Envelope Equation

In order to derive analytically a relationship between emittances and field energy we consider x and z as transverse coordinates and start from the single-particle equation of motion in x (similar for z)

$$x'' + k_x(s)x - \frac{q}{m\gamma^3 v_0^2} E_x(x,z,s) = 0 \tag{A.1}$$

where k_x describes the periodically varying focusing force and $\mathbf{E}$ follows from Poisson's equation

$$\nabla \cdot \mathbf{E} = \frac{q}{\varepsilon_0}\, n(x,z,s) \tag{A.2}$$

and n is the density obtained by projecting a 4-D phase space distribution into real space:

$$n = \int\int f\,(x,z,x',z',s)\,dx'\;dz' \tag{A.3}$$

The distribution function f is subject to Vlasov's equation ($v_0 \equiv dx/ds$).

$$\frac{\partial f}{\partial s} + (\mathbf{x}' \cdot \nabla)\; f - \left(\mathbf{k} - \frac{q}{m\gamma^3 v_0^2}\mathbf{E}\right) \cdot \nabla'_x f = 0 \tag{A.4}$$

Solution of equations (A.1)-(A.4) is fraught with the usual difficulties in solving a partial integro-differential equation. Keeping in mind that the r.m.s. emittance is defined through second-order moments of the distribution function, we follow the procedure introduced by Sacherer [13] and convert Eq. (A.4) into ordinary differential equations involving second-order moments only. We thus define the moments, where we assume that N is the total number of particles per unit length of beam, according to

$$\overline{x^2} = N^1 \int \ldots \int x^2 f\; dx \ldots dz', \tag{A.5}$$

similar with x'^2, xx' and likewise in z. The r.m.s. emittance is then defined as (note that some authors drop the factor 4)

$$\varepsilon_x \equiv 4\left(\overline{x^2 x'^2}\; -\; \overline{xx'}^2\right)^{1/2} \tag{A.6}$$

and similar in z. We then obtain from Eq. (A.4) by multiplying it with x^2 and integrating over all phase space

$$\frac{d}{ds}\overline{x^2} - 2\overline{xx'} = 0 \tag{A.7}$$

$$\frac{d}{ds}\overline{xx'}\; -\; \overline{x'^2} + k_x\overline{x^2} - \frac{q}{m\gamma^3 v_0^2}\overline{xE}_x = 0 \tag{A.8}$$

$$\frac{d}{ds}\overline{x'^2} + 2k_x\overline{xx'} - \frac{2q}{m\gamma^3 v_0^2}\;\overline{x'E}_x = 0 \tag{A.9}$$

By applying Sacherer's procedure we readily obtain the equation

$$\frac{d^2}{ds^2}\tilde{x} + k_x(s)\tilde{x} - \frac{\varepsilon_x^2(s)}{16\tilde{x}^3} - \frac{q}{m\gamma^3 v_0^2}\,\frac{\overline{xE_x}}{\tilde{x}} = 0 \tag{A.10}$$

(similar in z) with $\tilde{x} \equiv \overline{x^2}^{1/2}$ the r.m.s. envelope. This requires, however, knowledge of ε_x to be of real use. Hence, for constant ε_x, a straightforward integration of the r.m.s. envelope equation is possible, if $\overline{xE}_x$ can be calculated explicitly. This is indeed possible for a uniform charge distribution within an elliptical boundary, in which case E_x is a linear function of x (see section 3).

Hence we obtain the r.m.s. envelope equation

$$\frac{d^2}{ds^2}\tilde{x} + k_x(s)\tilde{x} - \frac{\varepsilon_x^2}{16\tilde{x}^3} - \frac{qI}{4\pi\varepsilon_0 m\gamma^3 v_0^2(\tilde{x}+\tilde{z})} = 0 \tag{A.11}$$

and similar in z. Note that $\tilde{x} = a/2$ and $\tilde{z} = b/2$, where a, b are the usual envelopes.

We also derive from Eq. (A.10) the differential equation

$$\frac{d}{ds}\varepsilon_x^2 = \frac{32q}{m\gamma^3 v_0^2}\left(\overline{x^2}\;\overline{x'E_x} \;-\; \overline{xx'}\;\overline{xE_x}\right) \tag{A.12}$$

and similar for z, where we had to introduce the moments $\overline{xE}_x, \overline{x'E_x}$. For E_x a linear function of x, one readily sees that ε_x must be constant, hence the r.m.s. envelope equations are a closed set of ordinary differential equations and can be solved explicitly. For E_x other than a linear function of x (or a constant) these moments are of higher order, hence the above differential equations are not a closed set. By going to higher order moments we would obtain an infinite set of coupled equations, in general.

A.2 Generalized Emittance Equation

In the following we show that it is possible to transform the terms involving the electric field in such a way that only the energy of the field appears explicitly. The latter still includes higher-order moments, but we benefit from the fact that it is a quantity of direct physical meaning and amenable to estimates.

For this purpose we re-write

$$\overline{x'E_x} = N^{-1}\int \ldots \int x'E_x f\; dx \ldots dz' = N^{-1}\int\int E_x n v_x\; dx\; dz \tag{A.13}$$

where we have introduced v_x as local averaged velocity of beam particles (in a frame, where the beam is at rest).

With the local current

$$\mathrm{j} = qnv_0\mathrm{v} \tag{A.14}$$

we obtain

$$\overline{x'E_x} = (Nqv_0)^{-1}\int\int E_x j_x dx\; dz \tag{A.15}$$

and similar for z. We thus find, using $\mathbf{E} = -\nabla\phi$):

$$\int\int \mathbf{E}\cdot\mathbf{j}\; dx\; dz = \int\int \phi\nabla\cdot\mathbf{j}\; dx\; dz = -qv_0\int\int \phi\frac{\delta n}{\delta s}\; dx\; dz \tag{A.16}$$

where we have used the continuity equation

$$\frac{\delta n}{\delta s} + (qv_0)^{-1}\,\nabla\cdot\mathbf{j} = 0 \tag{A.17}$$

derived from Eq. (A.4) by integration. The integration in Eq. (A.16) is performed over the cross section F, which contains the beam in its interior, hence we may neglect a boundary integral. Using Poisson's equation we obtain by partial integration:

$$\int\int \mathbf{E}\cdot\mathbf{j}\; dx\; dz = -v_0\frac{d}{ds}W - \varepsilon_0 v_0\int \phi\frac{\partial}{\partial s}E_n d\sigma \tag{A.18}$$

with E_n the normal component of $\mathbf{E}$ on the boundary of F (surface element $d\sigma$) and

$$W = \frac{\varepsilon_0}{2}\int\int \mathbf{E}^2 dx\; dz \tag{A.19}$$

the electric field energy within F.

The next step is to add the three Eqs. (A.12) and express $\overline{x'E_x}$ by the electric field energy according to Eqs. (A.15, A.18). We thus find the relationship

$$\frac{1}{\overline{x^2}}\frac{d}{ds}\varepsilon_x^2 + \frac{1}{\overline{z^2}}\frac{d}{ds}\varepsilon_z^2 = \frac{32q}{m\gamma^3 v_0^2}\left[-\frac{1}{Nq}\frac{dW}{ds} - \frac{\varepsilon_0}{Nq}\int \phi\frac{\delta}{\delta s}E_n d\sigma - I\right] \tag{A.20}$$

with

$$I = \frac{1}{2}\left(\frac{1}{\overline{x^2}}\frac{d\overline{x^2}}{ds}\overline{xE_x} + \frac{1}{\overline{z^2}}\frac{d\overline{z^2}}{ds}\;\overline{zE_z}\right). \tag{A.21}$$

Equation (A.20) holds exactly, but we need some approximation to evaluate the term I on the r.h.s. and obtain a practically useful equation.

Firstly, it can be shown that for a uniform beam with elliptic cross section the field energy per unit length calculated within a large circle of radius R is given by

$$W_u = \frac{N^2q^2}{16\pi\varepsilon_0}\left(1 + 4\; ln\frac{2R}{a+b}\right) \tag{A.22}$$

where a and b are the semi-axi in x and z. We then also find for the uniform beam that I is related to W_u according to

$$I = -\frac{1}{Nq}\frac{d}{ds}W_u \tag{A.23}$$

Next we use the result found by Sacherer that $\overline{xE_x}$ and $\overline{zE_z}$ are independent of the density profile as long as elliptical symmetry is given:

$$n(x,z,s) = n\left(\frac{x^2}{a^2} + \frac{z^2}{b^2}, s\right) \tag{A.24}$$

Hence, Eq. (A.23) is true for all beams satisfying equation (A.24) and we re-write Eq. (A.20) as

$$\frac{1}{\overline{x^2}}\frac{d}{ds}\varepsilon_x^2 + \frac{1}{\overline{z^2}}\frac{d}{ds}\varepsilon_z^2 = -4K\frac{d}{ds}\frac{W - W_u}{w_0}. \tag{A.25}$$

Here we have neglected the boundary integral (which is justified for large R) and introduced the generalized perveance

$$K \equiv \frac{Nq^2}{2\pi\varepsilon_0 m\gamma^3 v_0^2} \tag{A.26}$$

and the field energy normalization constant $w_0 \equiv (N^2q^2)/(16\pi\varepsilon_0)$, which gives the field energy of a uniform beam within the actual beam boundary. We note that the generalized emittance Eq. (A.25) derived here for 2-D beams can be derived for 3-D bunched beams as well [7].

RADIO-FREQUENCY QUADRUPOLE

M. Weiss
CERN, Geneva, Switzerland

ABSTRACT
The principle of operation of the Radio-Frequency Quadrupole (RFQ) is shown. Rather than describe all the details of this accelerator, basic equations are given and the approach which led to the invention of the RFQ is explained. Similar ideas may be applied to the solution of other problems, of which one example is given.

1. INTRODUCTION

The Radio-Frequency Quadrupole (RFQ) is a linear accelerator of a relatively recent date. In spite of its rather simple principle, it was invented only in the late 1960s by Kapchinsky [1, 2] of the Institute of Theoretical and Experimental Physics (ITEP), Moscow. Teplyakov [3] of the same Institute constructed the cavity, in which a high-frequency electromagnetic field is created that is of such a type and mode as to permit the RFQ to operate properly. Important contributions to the RFQ have also been made by the Los Alamos National Laboratory (LANL) [4], where a model called 'Proof of Principle' (POP) was constructed and put into operation in 1980 [5]. Since then, the RFQ has become so popular that in just a few years several tens of RFQs have been constructed and put into operation in many laboratories and universities all over the world.

The salient features of the RFQ are that it bunches, focuses, and accelerates charged particles by using RF fields only. The RFQ is particularly suited to accelerating ions in the low-energy range (up to 1 or 2 MeV per nucleon). For this purpose it is certainly the best machine existing today.

In order to grasp fully the originality of the RFQ, we shall start this paper by recalling some properties of the electromagnetic fields in resonant cavities and show the conditions which such cavities must satisfy to be used as accelerators. After that, an extensive analysis of the RFQ will follow, comprising beam dynamics, electrodynamics, and some information concerning its construction. We shall end by briefly showing how the same approach as that which led to the invention of the RFQ could eventually be used to solve other tasks.

2. WAVES AND CAVITIES

2.1 Waves in empty cavities

Electromagnetic waves in free space are of the 'transverse electromagnetic' type (TEM): both the electric and the magnetic fields are transverse, i.e. perpendicular to the direction of propagation. In empty cavities, such waves are not possible and one of the fields, either the electric (TM wave) or the magnetic (TE wave), has to be in the longitudinal direction.

In cavities, each wave type can be present in various modes; the modes are indicated by subscripts and they describe the spatial periodicity of the wave. For example, when working in rectangular coordinates, TM_{mnp} indicates a transverse magnetic wave type with m, n, and p half periods along the x, y, and z direction, respectively. (In cylindrical coordinates the meaning of subscripts is somewhat different: m indicates the number of full circumferential periods, n is the number of radial variations, and p is, as before, the number of half periods along z.)

The wave equation of the electromagnetic field in vacuum is

$$\nabla^2 \boldsymbol{E} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0$$

$$\nabla^2 \boldsymbol{B} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{B}}{\partial t^2} = 0 \ ,$$

where the operator ∇ (nabla) stands for spatial derivatives $\left[\nabla = (\partial / \partial x)\bar{i} + (\partial / \partial y)\bar{j} + (\partial / \partial z)\bar{k}\right]$ and c is the wave velocity (the velocity of light). In cavities one sometimes prefers to deal with currents and voltages rather than with magnetic and electric fields; the wave equation applies then also to the former quantities.

In a bounded medium one usually tries to express the solution of the wave equation as a product of functions depending only on one variable

$$E_i(x,y,z,t) = X_i(x)Y_i(y)Z_i(z)T_i(t) \ ,$$

where the subscript i indicates the i^{th} component of the field.

Dealing with harmonic oscillations, the time dependence can be expressed by $e^{j\omega t}$; assuming z to be the direction of wave propagation, the function $Z(z)$ is of the form $\exp\left(-jk_z z\right)$. Hence the wave equation can be rewritten as:

$$\left[\nabla_{xy}^2 + \left(-k_z^2 + k^2\right)\right]E = 0$$

where $\partial^2 E / \partial z^2$ is replaced by $-k_z^2 E$ and $\left(-1/c^2\right)\left(\partial^2 E / \partial t^2\right)$ by $k^2 E\left(k^2 = \omega^2 / c^2\right)$. The constants k, which are called wave numbers, can be grouped as

$$-k_z^2 + k^2 = k_t^2 \ ,$$

and for real k_t one has:

$$k_z = \pm\sqrt{k^2 - k_t^2} \ ,$$

which implies $|k_z| < |k|$.

Writing the wave equation solution as

$$E_i(x,y,z,t) = X_i(x)Y_i(y)\ \exp\left[j(\omega t - k_z z)\right]$$

one recognizes a wave with a 'phase velocity':

$$v_{ph} = \omega / k_z > c \quad (\text{note } c = \omega / k) \ .$$

Usually one represents the relation between k (= ω/c) and k_z by a so-called dispersion diagram; for a given mode (k_t = const) one has the graph presented in Fig. 1. The curve is symmetric around the origin: negative k_z means propagation in the opposite direction, which is equally possible. The slope of the radius vector to a certain point P on the curve is equal to the phase velocity of that wave; the points on the curve are above the asymptote, indicating that $v_{ph} > c$.

In empty cavities the phase velocity is always higher than c and therefore such a cavity cannot be used directly to accelerate particles, because synchronism between wave and particle is required. One has to find a means of slowing down the wave. This can be done by 'loading' the cavity periodically, i.e. introducing some periodic obstacles.

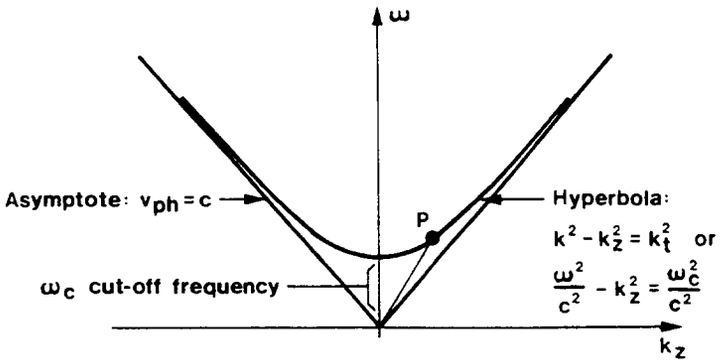


Fig. 1 Dispersion diagram (empty cavity)

2.2 Waves in loaded cavities

If one proceeds as shown in Fig. 2, the wave equation solution $Z(z)$ has to exhibit the periodicity in z; according to Floquet's theorem for periodic structures, the solution may differ, as one goes from a period to the next, only by a constant factor such as exp $(-jk_z d)$, where d is the period:

$$Z(z) \propto \exp(-jk_z z) \sum^{n} A_n \exp\left(-j\frac{2\pi n}{d} z\right).$$

In contrast to the empty cavity, the loaded cavity has a solution containing an infinite sum; only the correct spectrum of the various 'space harmonics' A_n will satisfy the now complicated boundary conditions.

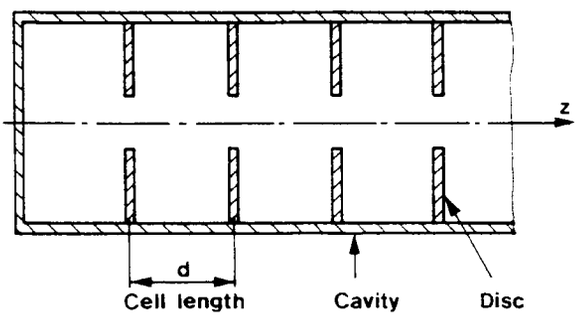


Fig. 2 Periodically disc-loaded cavity

The n^{th} harmonic of the electric field is:

$$E_n(x,y,z,t) = A_n F(x,y) \exp -\left\{ j\left[\omega t - \left(k_z + \frac{2\pi n}{d} \right) z \right] \right\}$$

and the phase velocity is slowed down:

$$v_{ph} = \frac{\omega}{k_z + (2\pi n / d)} = \frac{\omega}{k_n} .$$

One can draw a dispersion diagram also for the loaded cavity. It will be very different from that of an empty cavity: the 'disturbance' created by the periodic discs is shown in the new dispersion curve $\omega = f(k_n)$, Fig. 3. A synchronism between wave and particle is now possible, but with a much more complicated cavity. Owing to complex boundary conditions, the fields are usually determined numerically.

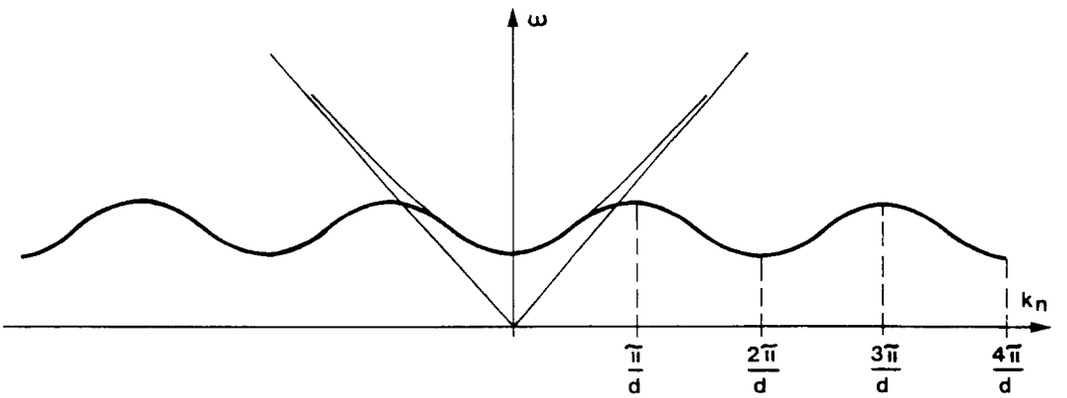


Fig. 3 Dispersion diagram of a periodic structure

For completeness we mention that the energy propagation in a cavity is linked with a velocity called the group velocity. This velocity is given by the tangent of the dispersion curve:

$$v_g = \frac{d\omega}{dk_n} .$$

2.3 Stability of motion of particles

There is one more thing which merits our attention: when accelerating a bunch of particles, one has to keep them together, both longitudinally and transversally: the particles perform synchrotron and betatron oscillations around a stable or 'synchronous' particle. In a frame which moves synchronously with the wave, the situation is quasi-static and the equation

$$\nabla \boldsymbol{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0$$

applies. For centripetal forces, all the terms must be negative, which is impossible. If we make $\delta E_z / \delta z < 0$ in order to have longitudinal stability, then at least one of the other two terms is positive and the motion is unstable transversally. An additional, 'external' focusing has therefore to be applied in such cases, and this is usually done with quadrupoles or solenoids.

3. PRINCIPLE OF OPERATION OF THE RFQ

The principle of operation of the RFQ can easily be explained by looking at some simple drawings. We start with Fig. 4, which represents an electric quadrupole having alternating

voltages on its electrodes. If particles move along the z-axis and they are slow enough (or the quadrupole is long enough), so that they stay inside the quadrupole for several periods of the alternating voltage, they will be exposed to an alternating electric field and undergo an alternating gradient focusing. Hence, instead of the usual sequence of 'spatially' distributed quadrupoles with alternating polarities, the same effect can be produced by a long, single quadrupole, with 'time' varying fields; see Fig. 5.

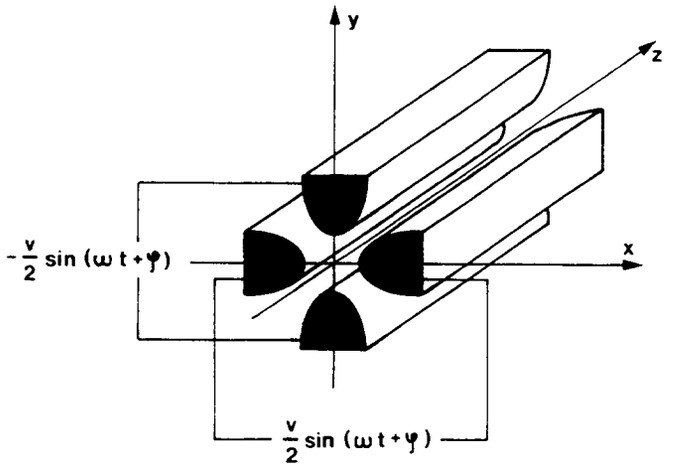


Fig. 4 Electric quadrupole with a.c. voltage

What about the longitudinal fields? They can be created by 'perturbing' the electrodes of the a.c. quadrupole of Fig. 4 in a periodic way: a part of the originally transverse field is deviated into the longitudinal direction. A convenient perturbation is a sinusoidal modulation and the sinusoids of the horizontal and vertical electrodes have to be displaced by half a period, see Fig. 6. This is the essence of the RFQ: the period of the perturbation is such as to create synchronism between 'wave' and particle, necessary for acceleration; in addition, the transverse a.c. field provides an alternating gradient focusing (note that the equation $\nabla \boldsymbol{E} = 0$, in the moving frame, has always one of the transverse terms > 0; the stability of the betatron motion is obtained by alternating these terms).

The RFQ operates at high frequencies, typically from several tens to several hundreds of MHz. The electrodes, which are called 'vanes', are therefore placed in a cavity (resonator) to prevent the RF fields from radiating. The problems of the cavity (electrodynamics) are distinct from the problems of beam propagation (beam dynamics). The latter are concentrated around the z-axis, and inside a radius much smaller than the cavity radius which is proportional to the wavelength λ. Owing to the symmetry, the magnetic field on the z-axis of the RFQ is zero; this is approximately true also for the region $r << \lambda$. The consequences for the beam dynamics are important:

i) The wave equations can, in this region, be replaced by the simpler Laplace equation.

ii) The vanes present well-defined boundaries, with a potential from which one can analytically derive the fields.

iii) It is possible to ask for specific fields and then determine the corresponding boundaries.

One can see that there are many possibilities, a fact recognized by Kapchinsky. We shall essentially follow his procedure by starting with the Laplace equation in cylindrical coordinates:

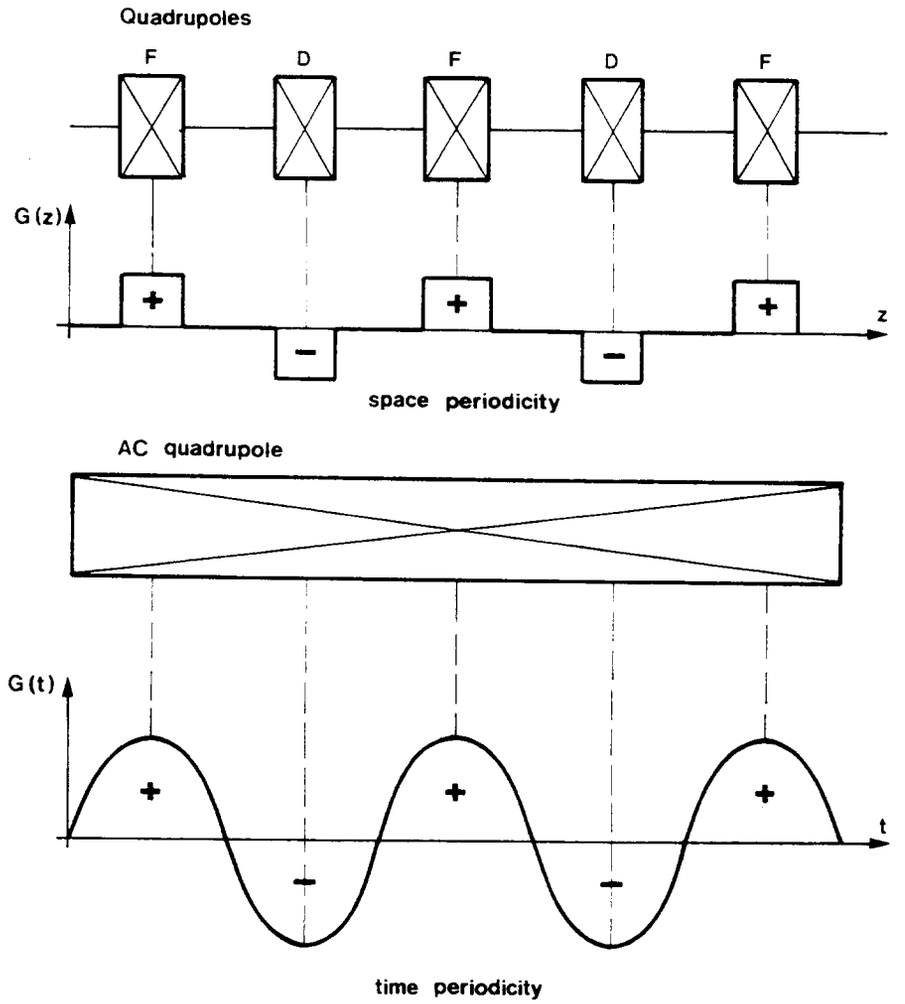


Fig. 5 Alternating gradient focusing with 'space' and 'time' periodicity

$$\nabla^2 U(r,\theta,z) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial U}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 U}{\partial \theta^2} + \frac{\partial^2 U}{\partial z^2} = 0 \ , \tag{1}$$

with $U(r,\theta,z)$ being the electric field potential, which has to be considered as multiplied by the factor sin $(\omega t + \varphi)$, owing to its a.c. character.

A general solution of this equation has been derived in Appendix A; the result is

$$U(r,\theta,z) = \frac{V}{2}\left[\sum^{n} A_{0n} r^{2n} \cos 2n\theta + \sum^{n}\sum^{m} A_{mn} I_{2n}(mkr) \cos mkz\right] , \tag{2}$$

where

$m + n = 2p + 1$; $p = 0, 1, 2, \ldots$;
$\pm V/2$ is the vane potential (see Fig. 4);
$I_{2n}(x)$ is the modified Besssel function of order $2n$;
$k = 2\pi/\beta\lambda$ with β being the relativistic factor and λ the wavelength.

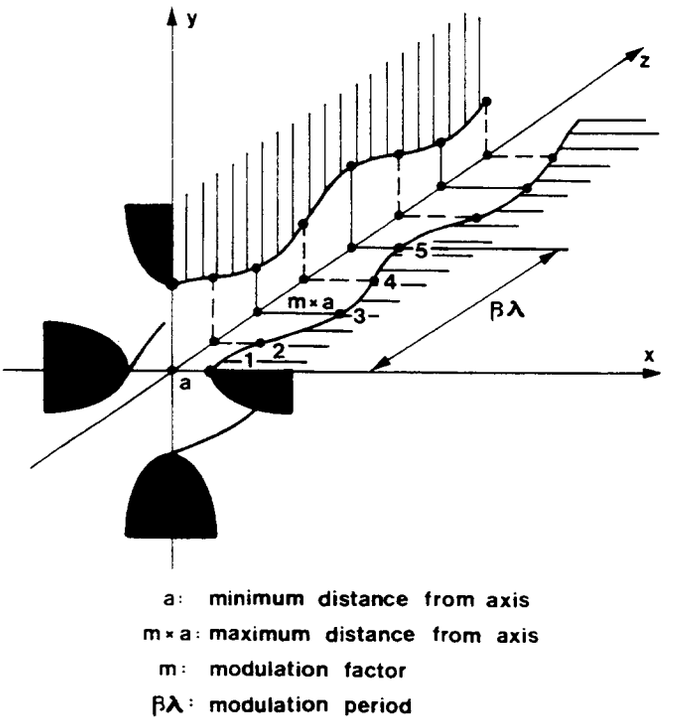


Fig. 6 Modulation of RFQ electrodes to create longitudinal fields

The general solution of the potential function $U(r,\theta,z)$ contains all the harmonics in infinite series; a real RFQ is usually sufficiently well described with only a few harmonics.

To facilitate the analysis, we shall consider the lowest-order solution, which contains only two terms, one out of each infinite series:

$$U(r,\theta,z)=\frac{V}{2}\left[A_{01}r^2\cos 2\theta + A_{10}I_0(kr)\cos kz\right] . \tag{3}$$

In the first term one recognizes the potential of an electric quadrupole; the second gives the accelerating field.

Imposing the condition that $U(r,\theta,z)$ is constant along a vane (see Fig. 6),

$$U(a,0,0)=U\left(ma,0,\frac{\beta\lambda}{2}\right)=\frac{V}{2} ,$$

one gets for the constants

$$\text{acceleration parameter } A_{10} = \frac{m^2 - 1}{m^2 I_0(ka) + I_0(mka)}$$

$$\text{focusing parameter } A_{01} = \frac{1}{a^2}\left[1 - A_{10} I_0(ka)\right] \tag{4}$$

$$= \frac{1}{a^2} \frac{I_0(ka) + I_0(mka)}{m^2 I_0(ka) + I_0(mka)} = \frac{\chi}{a^2} \ .$$

The constants are expressed by geometric parameters (a and m, see Fig. 6) which may change from period to period. For example, increasing m we get more acceleration, while decreasing a we get more focusing. When particles are accelerated, their velocity $v = \beta c$ increases and so the period $\beta\lambda$ must also increase (synchronism between wave and particle). One sees that linear accelerators are in fact only quasi-periodic structures.

From the above relation

$$1 - A_{10} I_0(ka) = \chi$$

we get, after multiplying by V:

$$\chi V + A_{10} I_0(ka) V = V \ .$$

This formula is very interesting: it tells us that the intervane voltage V is composed of a part required for focusing (χV) and another required for acceleration $\left[A_{10} I_0(ka)\right]$, see Eqs. (4).

For completeness, we also present the field components, in cylindrical coordinates, derived from the lowest-order potential function (3):

$$E_r = -\frac{\partial U}{\partial r} = -\frac{V}{2}\left[2 \ A_{01} r \cos 2\theta + k A_{10} I_1(kr) \cos kz\right]$$

$$E_\theta = -\frac{1}{r}\frac{\partial U}{\partial \theta} + V A_{01} r \sin 2\theta$$

$$E_z = -\frac{\partial U}{\partial z} = \frac{V}{2} k A_{10} I_0(kr) \sin kz \ .$$

The vanes have to be shaped so as to produce fields required by the beam dynamics. The vane surface equations are obtained from Eq. (3) putting $U(r, \theta, z) = \pm V / 2$:

$$S(r, \theta, z) = A_{01} r^2 \cos 2\theta + A_{10} I_0(kr) \cos kz = \pm 1 \ .$$

4. BEAM DYNAMICS IN THE RFQ

It was shown how the fields in the RFQ can be derived analytically from a well-defined potential function, a solution of the Laplace equation. In fact, one would like to impose conditions on these fields, according to beam dynamics requirements, and then find the corresponding potential and vane shapes.

The RFQ is used as an accelerator in the low-energy range; therefore non-relativistic equations of motion can be applied ($\gamma = 1$). We shall first consider the transverse dynamics

(betatron motion) and then the longitudinal dynamics (synchrotron motion). The Cartesian coordinate system will be used and for simplicity, but without losing sight of the essential points, fields belonging to the lowest-order potential function will be analysed.

4.1 Transverse beam dynamics

The equation of motion in one of the transverse planes (x or y) is

$$\frac{d^2x}{dt^2} = \frac{e}{m} E_x \ .$$

From the potential (3) with the time variation included,

$$U(x,y,z,t) = \frac{V}{2}\left[A_{01}\left(x^2 - y^2\right) + A_{10} I_0(kr) \cos kz\right] \sin\left(\omega t + \varphi\right) \ , \tag{5}$$

one gets the field

$$E_x = -\frac{\partial U}{dx} = -\frac{V}{2}\left[2A_{01}x + A_{10} I_1(kr) k \frac{dr}{dx} \cos kz\right] \sin\left(\omega t + \varphi\right) \ . \tag{6}$$

The first term gives the alternating (in time) gradient focusing; the second is the perturbation of the quadrupole field due to vane modulation. Making substitutions in Eq. (6):

$$I_1(kr) \sim kr/2 \qquad \text{(linearizing for small arguments)} \ ,$$

$$dr/dx = x/r \qquad \text{(differentiating } r^2 = x^2 + y^2) \ ,$$

$$\cos(kz) \sin\left(\omega t + \varphi\right) = \frac{1}{2}\left[\sin\left(2kz + \varphi\right) + \sin\varphi\right]$$

(replacing ωt by $\omega z / v = kz$) ,

one gets the linearized field

$$E_x = -\frac{V}{2}\left\{2A_{01} \sin\left(kz + \varphi\right) + \frac{1}{4} A_{10} k^2 \left[\sin\left(2kz + \varphi\right) + \sin\varphi\right]\right\} x \ .$$

Note that φ is the RF phase when the particle enters a cell.

It is convenient to introduce a normalized independent variable τ, which changes by one over a structure period $\beta\lambda$ (cell):

$$kz + \varphi = \omega t + \varphi = 2\pi\tau \ .$$

The equation of motion in the new variable is

$$\frac{d^2x}{d\tau^2} + \left[B \sin 2\pi\tau + \Delta_{rf} + \frac{\Delta_{rf}}{\sin\varphi} \sin 4\pi\tau\right] x = 0$$

where

$$B = \frac{\lambda^2}{c^2}\frac{e}{m}\frac{\chi V}{a^2} \qquad \text{is the focusing factor} \tag{7}$$

$$\Delta_{rf} = \frac{e}{m}\frac{\pi^2 \sin\varphi}{2c^2\beta^2}A_{10}V \qquad \text{is the RF defocusing factor .} \tag{8}$$

This is Hill's equation with a coefficient periodic in τ. There are two harmonics of the alternating gradient focusing in the periodic coefficient: $\sin 2\pi\tau$ and $\sin 4\pi\tau$. The first harmonic is preponderant, so we neglect the second one and get:

$$\frac{d^2x}{d\tau^2} + \left[B \sin 2\pi\tau + \Delta_{rf}\right]x = 0 \ . \tag{9}$$

This is the Mathieu equation, which is typical for beam dynamics in periodic structures. We look for solutions of the form

$$x(\tau) = e^{j\sigma_T\tau}\sum^{m} c_m e^{j2\pi m\tau} \ , \qquad m = 0, \pm 1, \pm 2, \ ...$$

where exp ($j\sigma_T\tau$) is the slowly varying part (smooth part), oscillating with the betatron frequency σ_T and depending on the average focusing, while exp ($j2\pi m\tau$) is the wiggle, containing the periodicity of the structure. In Appendix B, the Mathieu equation has been analysed and the solution, to first order in m, obtained as

$$x(\tau) = C_0\, e^{j\sigma_T\tau}(1 + C \sin 2\pi\tau)$$

with

$$\sigma_T^2 \sim \frac{B^2}{8\pi^2} + \Delta_{rf} \tag{10}$$

and

$$C \sim \frac{B}{4\pi^2} \ . \tag{11}$$

The smooth frequency σ_T and the wiggle are expressed with the parameters of the Mathieu equation. Note that with τ as independent variable, the frequency σ_T is identical with the betatron phase advance per period.

4.2 Longitudinal beam dynamics

In the longitudinal plane (synchrotron motion), the equations are somewhat different. We consider first the energy gain of the particle in a cell $\beta\lambda$.

From Eq. (5) we get the longitudinal field

$$E_z(x,y,z) = -\frac{\partial U}{\partial z} = \frac{1}{2}kA_{10}I_0(kr)\sin kz \sin\left(\omega t + \varphi\right) \ ,$$

and integrating it over the n^{th} cell (note that the contribution of each half cell is the same) we obtain the energy gain in that cell:

$$\begin{aligned} W_n - W_{n-1} &= 2\int_0^{\beta\lambda/2} eE_z(0,0,z)dz \\ &= 2\left[\frac{1}{2}ekA_{10}V\int_0^{\beta\lambda/2}\sin kz\,\sin(kz+\varphi)dz\right] \\ &= 2\left(\frac{\pi}{4}eA_{10}V\cos\varphi\right) = 2\Delta W_{1/2} \end{aligned} \tag{12}$$

where

$\Delta W_{1/2}$ is the energy gain in a half cell
W_n is the particle energy after the n^{th} cell.

The energy gain in a cell is independent of the cell length. This is because the gain is determined by the intervane voltage V, which is fixed. In contrast, in most other accelerators it is the mean field $\overline{E}$ which is fixed and the energy gain is proportional to $\overline{E}\,\beta\lambda$. The RFQ is therefore not suitable for higher β. It can, however, operate advantageously at low β's which is impossible with other accelerators (see Fig. 7). An Alvarez accelerator needs to be preceded by a high-voltage electrostatic preaccelerator (several hundred kV) to bring particles to an acceptable β.

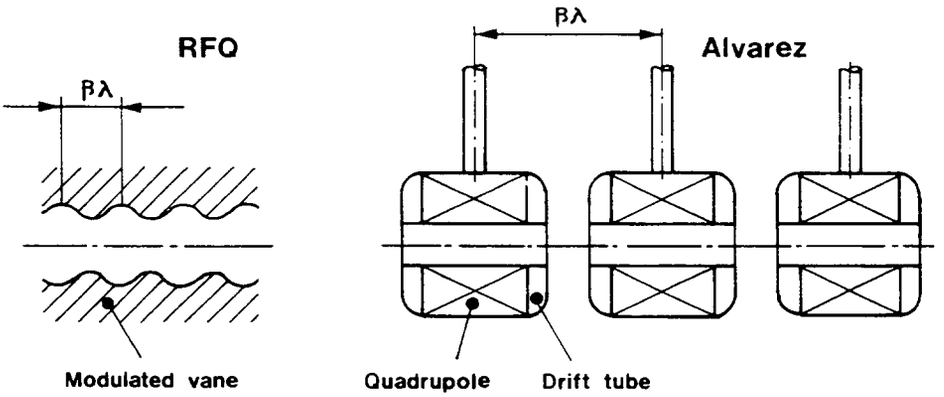


Fig. 7 RFQ and Alvarez linear accelerators

The energy gain in a cell depends on the phase φ. For stable motion, particles must oscillate in phase and energy around the stable, synchronous particle. The stability of the synchrotron motion is analysed with the usual canonic variables which, for ω = const are

$$\begin{aligned} \Delta\varphi &= \varphi - \varphi_s \\ \Delta W &= W - W_s \quad , \end{aligned}$$

indicating the phase and energy differences between a test particle and the synchronous one. (We consider only test particles on the axis to avoid complicated equations due to the coupling between synchrotron and betatron oscillations.)

From

$$W = W_s + (dW/d\beta)_{\beta_s}(\beta - \beta_s) ,$$

where

W is the energy of the test particle
W_s is the energy of the synchronous particle

one gets

$$\Delta\beta = \beta - \beta_s = (W - W_s)/mc^2\beta_s$$

and can derive the differential equation for $\Delta\varphi$

$$d\Delta\varphi/dt = -\omega\Delta\beta/\beta_s = -\omega\Delta W/mc^2\beta_s^2 ,$$

or in the variable τ

$$d\Delta\varphi/d\tau = -2\pi\Delta W/mc^2\beta_s^2 . \tag{13}$$

The corresponding equation for the energy is

$$d\Delta W/d\tau = (\pi/2)eA_{10}V[\cos\varphi - \cos\varphi_s] . \tag{14}$$

This is the basic set of equations for the synchrotron motion. We shall, essentially, analyse two cases: i) linearized equations with slowly varying β; ii) non-linear equations in the adiabatic approximation (β = const).

i) To get a linear set of equations, we linearize Eq. (14) for small $\Delta\varphi$:

$$\cos(\varphi_s + \Delta\varphi) - \cos\varphi_s \sim -\Delta\varphi \sin\varphi_s ,$$

and obtain

$$d\Delta W/d\tau = a^2(\tau)\Delta\varphi \tag{15}$$

and from Eq. (13)

$$d\Delta\varphi/d\tau = -b^2(\tau)\Delta W \tag{16}$$

where $a^2(\tau)$ and $b^2(\tau)$ are slowly varying functions of τ:

$$a^2(\tau) = -(\pi/2)eA_{10}V\sin\varphi_s ,$$

(note that $\varphi_s < 0$ for stability of synchrotron motion) and

$$b^2(\tau) = 2\pi/mc^2\beta_s^2 .$$

At a given moment τ, Eqs. (15) and (16) specify the smooth motion in the $(\Delta\varphi, \Delta W)$ phase plane. Eliminating $d\tau$ we get:

$$\frac{a^2}{b^2} = -\frac{\Delta W d\Delta W}{\Delta\varphi d\Delta\varphi} ,$$

which gives after integration:

$$\frac{\Delta\varphi^2}{b^2} + \frac{\Delta W^2}{a^2} = C .$$

The small oscillations describe an ellipse in the phase plane with the ratio of axes

$$\frac{\Delta\hat{W}}{\Delta\hat{\varphi}} = \frac{a}{b} .$$

This ratio of axes changes with the energy of the particles; it is proportional to β, which is easily verified. [For comparison, in an Alvarez accelerator this ratio is, in relativistic form, $\Delta\hat{W} / \Delta\hat{\varphi} \approx (\beta\gamma)^{3/2}$.]

Differentiating Eq. (16) and substituting $d\Delta W / d\tau$ from Eq. (15), we get the second-order differential equation for $\Delta\varphi$:

$$\frac{d}{d\tau}\left(\beta_s^2 \frac{d}{d\tau}\Delta\varphi\right) + \frac{\pi^2 e A_{10} V}{mc^2} |\sin\varphi_s| \Delta\varphi = 0 . \tag{17}$$

Such an equation is solved by multiplying it first by β_s^2 and then replacing $\beta_s^2 (d / d\tau)$ by d/du, which leads to an equation, in the new variable u, of the familiar form

$$\frac{d^2}{du^2}\Delta\varphi + K(u)\Delta\varphi = 0 .$$

The coefficient $K(u)$ is varying slowly, so the equation can be solved by the BKW (Brillouin-Kramer-Wenzel) method. Transforming the solution back to the variable τ we have

$$\Delta\varphi(\tau) = \Delta\hat{\varphi}(\tau) \sin\left[\int_0^\tau \sigma_L(\tau')d\tau' + \varphi_0\right] , \tag{18}$$

with

$$\Delta\hat{\varphi}(\tau) = \text{ const}\left[\frac{\pi^2 e A_{10} V \beta_s^2 |\sin\varphi_s|}{mc^2}\right]^{-1/4}$$

and

$$\sigma_L(\tau) = \left[\frac{\pi^2 e A_{10} V |\sin \varphi_s|}{mc^2 \beta_s^2} \right]^{1/2} . \tag{19}$$

Both the amplitude $\Delta\hat{\varphi}$ and the 'frequency' σ_L (phase advance per period) decrease with energy:

$$\Delta\hat{\varphi} \propto \beta^{-1/2} \quad \text{(phase damping)}$$
$$\sigma_L \propto \beta^{-1} .$$

[The corresponding, but relativistic, values in an Alvarez accelerator are $\Delta\hat{\varphi} \propto (\beta\gamma)^{-3/4}$, $\sigma_L \propto \beta^{-1/2}\gamma^{-3/2}$].

ii) So far we have analysed the synchrotron motion in the linear approximation. The linearization of the problem enabled us to obtain and solve the differential equation (17). When dealing with large oscillations, where the amplitudes approach the limits of stability, linearization is no longer valid. To solve the non-linear system (13), (14), we have to make other approximations: we assume adiabaticity (β = const) and get differential equations with constant coefficients. From Eqs. (13) and (14) we derive the second-order differential equation:

$$\frac{d^2\Delta\varphi}{d\tau^2} + \sigma_L^2 \frac{\cos(\varphi_s + \Delta\varphi) - \cos\varphi_s}{|\sin\varphi_s|} = 0 . \tag{20}$$

The first or energy integral is:

$$\frac{d}{d\tau}\left[\frac{1}{2}\left(\frac{d\Delta\varphi}{d\tau} \right)^2 + \sigma_L^2 V(\Delta\varphi) \right] = 0 ,$$

with $V(\Delta\varphi)$ being the potential function

$$V(\Delta\varphi) = \frac{\sin(\varphi_s + \Delta\varphi) - \Delta\varphi\cos\varphi_s - \sin\varphi_s}{|\sin\varphi_s|} , \tag{21}$$

and satisfying $V(0) = 0$. From the energy integral one has

$$\frac{1}{2}\left(\frac{d}{d\tau}\Delta\varphi \right)^2 + \sigma_L^2 V(\Delta\varphi) = \text{ const } = H , \tag{22}$$

where H indicates the oscillating energy of a particle, which is composed of the kinetic and the potential part. From this expression interesting plots can be drawn, as shown in Fig. 8: The first plot shows the potential $V(\varphi)$ connected with phase oscillations. One recognizes the stable phase φ_s and the stability limits φ_1 and φ_2 belonging to the highest allowed oscillating energy $H_{\max}$. The second plot shows particle trajectories in the $(\Delta\varphi, \Delta W)$ phase plane. To obtain them, $d\Delta\varphi / d\tau$ in Eq. (22) has been replaced by ΔW from Eq. (13). For each value of H we get another curve in the $(\Delta\varphi, \Delta W)$ phase plane. For $H = H_{\max}$ we get the separatrix, limiting the stable region and passing, as is known, through the unstable stationary (or fixed)

point. The extension of the separatrix along the $\Delta\varphi$ axis depends on the stable (synchronous) phase φ_s: ψ is the extension $\Delta\varphi_1 - \Delta\varphi_2$. The function $\psi(\varphi_s)$ is given in the implicit form

$$\text{tg}\,\varphi_s = \frac{\sin\psi - \psi}{1 - \cos\psi} \,. \tag{23}$$

It can be derived from Eq. (21). For $\Delta\varphi = 0 (\varphi = \varphi_s)$ we have the maximum allowed energy spread:

$$\Delta W = \sqrt{mc^2\beta_s^2 eA_{10}V(\varphi_s \cos\varphi_s - \sin\varphi_s)} \,. \tag{24}$$

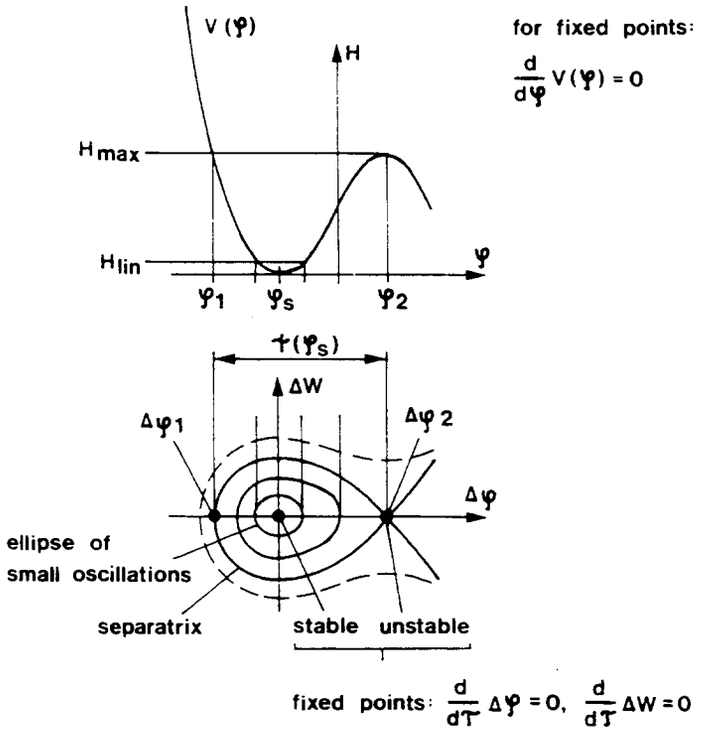


Fig. 8 Energy plot and trajectories in the $(\Delta\varphi, \Delta W)$ phase plane

Note that for φ_s = -90°, the extension of the separatrix ψ is over 360°. In this case we have only bunching, but not acceleration, of particles [cf. Eq. (12)].

This is essentially the beam dynamics of the RFQ. Many of the expressions derived have, however, a quite general character and apply to other linear accelerators working in the low-β region.

4.3 Conditions put on parameters of the RFQ

What conditions do we put on the beam dynamics in the RFQ?

Since the RFQ operates at low energies and is used for intense beams, the space-charge forces are quite important and require careful handling. Kapchinsky proposed a 'sound' condition: to try to keep the external and the space-charge forces constant. Constant external

forces mean constant σ_T and σ_L; constant space-charge forces mean constant beam radius (smooth) and bunch length [in Δz not $\Delta\varphi$; $(\beta\lambda / 2\pi)\Delta\varphi$].

Applying Kapchinsky's condition, we get from Eqs. (10), (7), (8), (19), and (23):

$$\begin{aligned} \frac{\chi}{a^2} &= \frac{1 - A_{10} I_0(ka)}{a^2} = \text{ const} \\ \frac{A_{10} \sin \varphi_s}{\beta_s^2} &= \text{ const} \\ \psi(\varphi_s)\beta_s\lambda &= \text{ const}\,. \end{aligned} \tag{25}$$

Equations (25) give us the functional relation between RFQ parameters, when quasi-stationary bunches are to be accelerated. The RFQ parameters are thereby determined in a unique way.

There is, however, a small 'drawback' in the above procedure: analysing Eqs. (25) we see that the beam must enter the RFQ already in bunches in order that the equations hold strictly. If we inject a continuous beam, there must be a kind of 'transition' region over which the beam gets slightly bunched and then Kapchinsky's conditions can be applied. Without going into details here, we just mention that transition sections can be designed in several ways (as, for example, at Los Alamos, Berkeley, and Saclay) without impairing the beam quality.

There is still another, rather important point which Kapchinsky has not solved: how to match a 'time-independent' focusing structure (a normal transport system with focusing elements) to a 'time-dependent' focusing structure in the RFQ. In fact, the RFQ acceptance varies with the phase of the RF voltage and, at different moments in the RF cycle, it can have shapes as different as shown in Fig. 9a. Crandall, from LANL, has made an important contribution by proposing to move the vanes apart at the RFQ input, so that there is very little focusing. The vanes would then come progressively to their nominal distance, with a focusing that would slowly (adiabatically) increase. The procedure was unbelievably efficient: the RFQ acceptances at the input of the 'radial matching section' (RMS) became nearly time independent, see Figs. 9b and 10.

Today the shape of the vanes in the RMS is so refined (see Appendix C) that the RFQ acceptance is practically a single ellipse.

4.4 Vane machining

The vanes are usually machined on computer controlled milling machines. The cutter follows the path defined by the vane surface equation, which is, in turn, derived from the potential function. In practice, one approximates the hyperbola-like vane cross-section by an arc of a circle and straight lines; also one often tries to keep the transverse radius of curvature of the vane tip constant along the vane. Such a practice requires the introduction of higher-order multipoles in the potential function. The number of multipoles corresponds to the number of additional conditions we impose on the vane shape. The multipoles appear, of course, also in the fields and one has to keep a good compromise between the simplicity of vane profiles and the presence of higher-order terms in the fields.

5. RFQ AS RESONATOR

5.1 Wave type and mode

When dealing with beam dynamics, we concentrated on a small region around the longitudinal RFQ axis, where the beam propagates. The Laplace equation was sufficient to

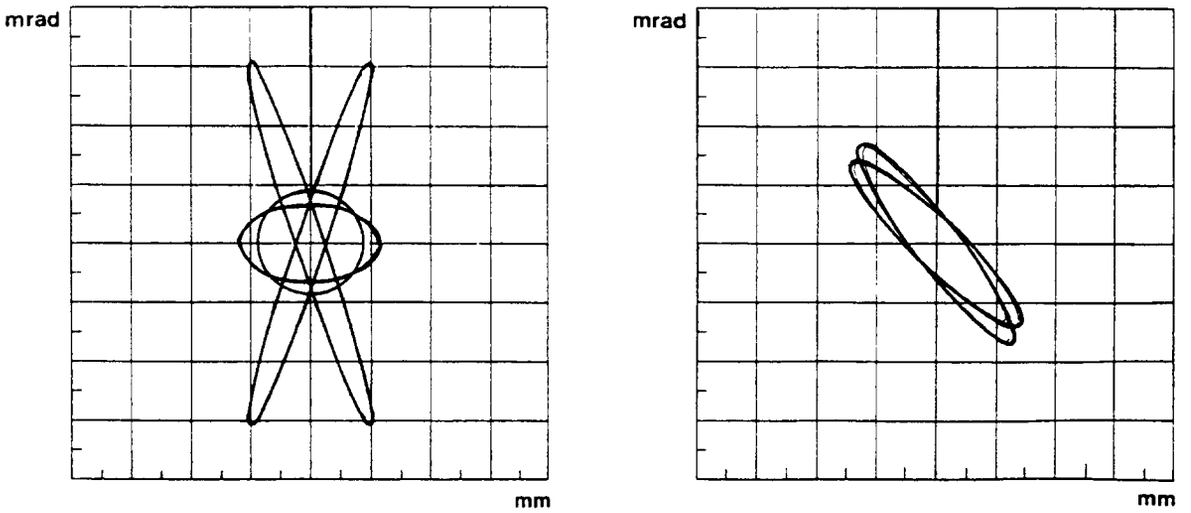


Fig. 9 RFQ acceptance without (a) and with (b) RMS

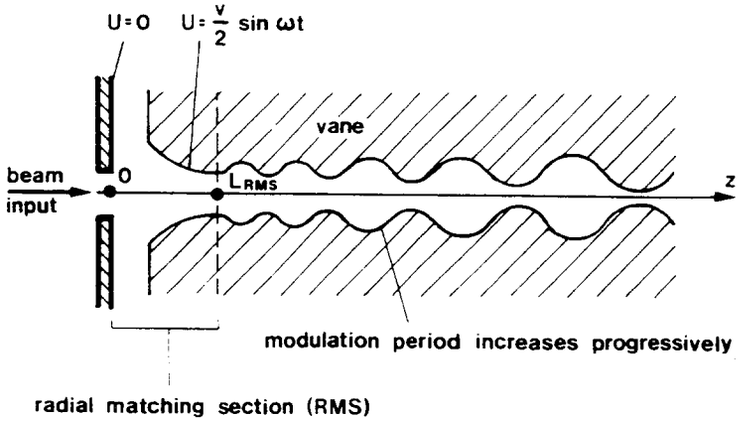


Fig. 10 Typical vane shape in an RFQ

describe the fields existing in this small region. On the contrary, when dealing with electrodynamics, i.e. with conditions necessary to establish certain wave types and modes at a given frequency, we have to consider the whole resonant cavity (resonator). In this case, the physics is described by the wave equation, as already mentioned before.

There are several types of resonators used for the RFQ, the most frequent one being the 'four-vane resonator' introduced at LANL. It is shown in Fig. 11, together with an empty cavity, where the 'pure' TE_{210} type and mode is indicated. This is the mode used with the

RFQ. In the RFQ the same mode is modified owing to the presence of the vanes; nevertheless, it is still considered a TE_{210} mode. Note that the TE_{210} mode has transverse electric fields and a longitudinal magnetic field. To create a longitudinal electric field, but only in the beam region, the vane tips are modulated; for the resonator as a whole, this modulation is only a minor perturbation.

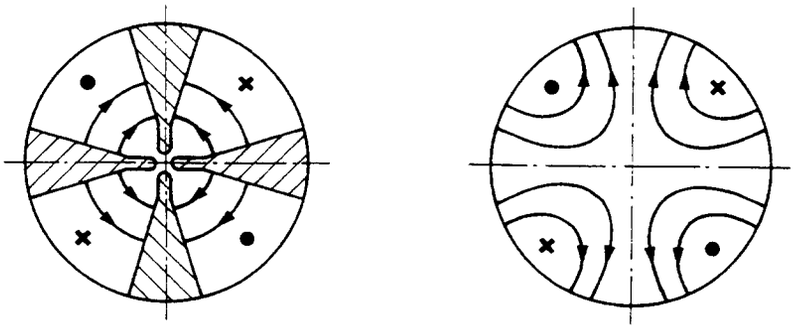


Fig. 11 TE_{210} mode in a four-vane cavity and in an empty cavity

To find the RFQ cavity dimension corresponding to a given resonance frequency and to the TE_{210} mode, one uses computer programs. This is true, in general, for all types of loaded cavities. However, apart from the TE_{210} mode, there are other modes on which the RFQ can resonate, of course at different frequencies. The difference in frequency is sometimes small and other modes can 'mix in', also when working at the TE_{210} frequency. Such a situation occurs when the boundary conditions for the TE_{210} mode are not perfect, which is always due to mechanical tolerances. The theory of perturbation tells us that the bigger the error the more the neighbouring modes mix in. There are methods to 'stabilize' the cavity, i.e. to make it less sensitive to errors. Without going into details, we must mention that stabilization means essentially increasing the group velocity of the desired wave which is equivalent to increasing the distance (in frequency) to the neighbouring modes.

The mode which is the nearest to the operating mode is TE_{110}, the dipole mode, shown in Fig. 12. This is the most dangerous mode. Generally speaking, disturbing modes affect either the transverse quadrupole symmetry or the longitudinally constant voltage of the vanes. For a constant vane voltage, the longitudinal magnetic field must be constant along z:

$$B_z(x,y,z) \neq f(z) \rightarrow \nabla \times \boldsymbol{E} \neq f(z) \rightarrow U(x,y,z) \neq f(z) \ .$$

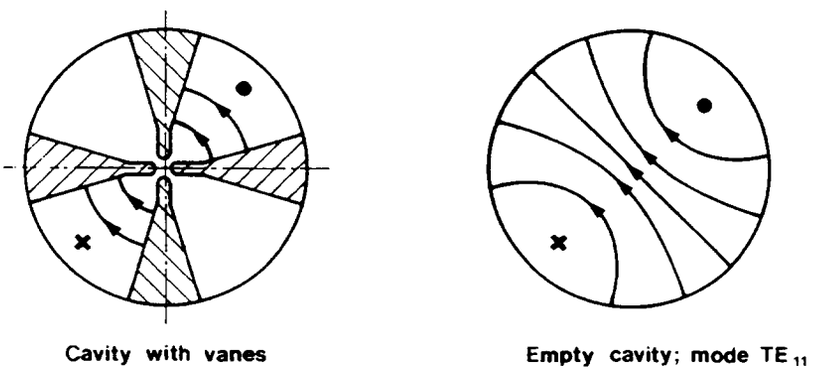


Fig. 12 TE_{110} mode in a four-vane cavity and in an empty cavity

It is instructive to draw simple equivalent circuits of the RFQ resonator and analyse them by Kirchhoff's laws. As mentioned, vane modulation is neglected and the resonator is considered, longitudinally, a uniform structure. A simple equivalent circuit for a perfectly symmetric RFQ is shown in Fig. 13. The 'capacitance' is concentrated mostly between the vane tips, the 'inductance' in the outer intervane space. The capacitances C_0 are the 'coupling' capacitors, which are very important for the mode spacing. The equivalent circuit, owing to its form, will have four modes; the dipole (TE_{110}) and quadrupole (TE_{210}) frequencies are related through

$$f_d = \frac{f_q}{\sqrt{1+k}}$$

where k is the 'coupling' defined as $k = C_0/C$, and the quadrupole frequency

$$f_q = \frac{1}{2\pi\sqrt{LC}} ,$$

i.e. the frequency of a quadrant.

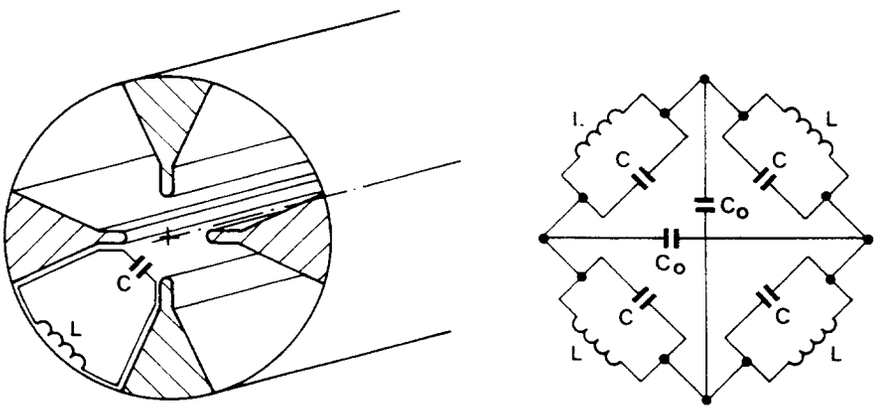


Fig. 13 Equivalent circuit of the RFQ (simplified)

5.2 The end region

The boundary conditions imposed on the electromagnetic field in a cavity are

$$E_p = 0 \ , \quad B_n = 0 \quad \text{(Dirichlet)} \ ,$$

$$\frac{\partial E_n}{\partial n} = 0 \ , \quad \frac{\partial B_p}{\partial n} = 0 \quad \text{(Neumann)} \ ,$$

where p and n indicate the directions parallel and perpendicular to the boundary.

The RFQ resonator has, of course, to be closed at both ends. What about the longitudinal magnetic field? It cannot be perpendicular to the end cover, it must be parallel: the magnetic field must therefore turn round and this change of direction must not influence the constancy of the vane potential. The solution applied for the RFQ extremities is shown in Fig. 14. The vanes do not extend right up to the end covers and, in addition, are cut out to facilitate the turning of the magnetic field. The 'end region' can also be represented by an equivalent circuit: it must resonate at the quadrupole frequency, i.e.

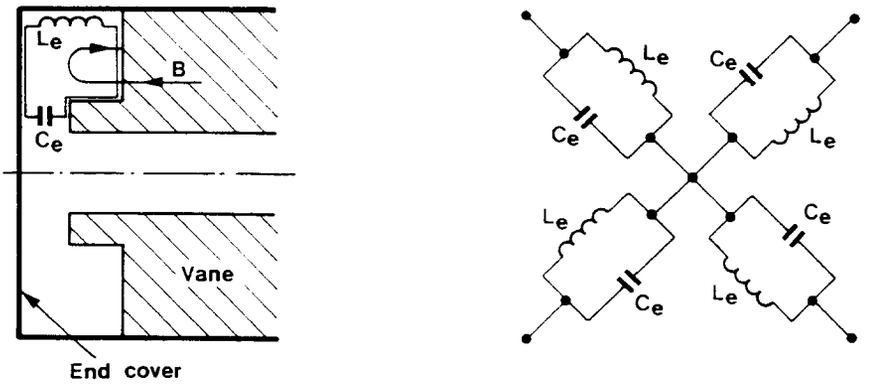


Fig. 14 Equivalent circuit of RFQ extremity

$$L_e C_e = LC \ ,$$

a condition necessary to keep the constancy of the vane potential.

5.3 Feeding of RF power

The electromagnetic energy is brought into the cavity via a loop, as schematically shown in Fig. 15. Sometimes two or four loops are used (in two opposite or in all the four quadrants) to force the RFQ to resonate in the wanted mode. The loop represents a magnetic coupling of the external RF generator to the RFQ; this coupling must be such as to match the impedance of the RFQ (under beam-loading conditions) to the outer circuit (usually the characteristic impedance of the feeder line); see Fig. 16.

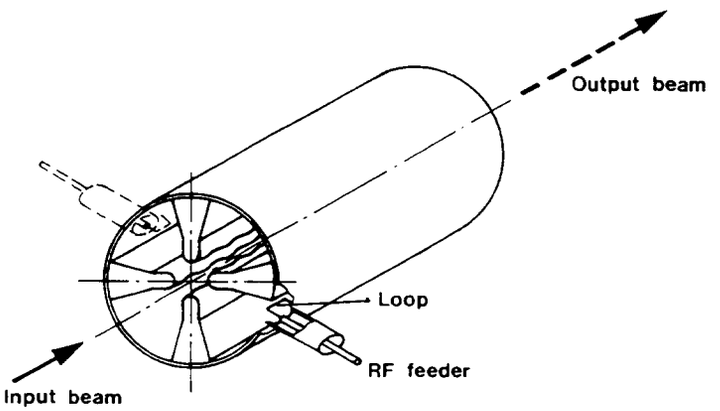


Fig. 15 RFQ with feeder loops

5.4 Other types of resonators for the RFQ

It has been mentioned that the four-vane resonator is not the only one applied to the RFQ. In fact, at low frequencies this resonator becomes quite large (the cavity diameter is proportional to the RF wavelength). For such cases, another type developed at the University of Frankfurt-on-Main is preferred; see Fig. 17. The resonant conditions are obtained with the capacity of the electrodes and the inductance of the holding bars. The electrodes are simpler,

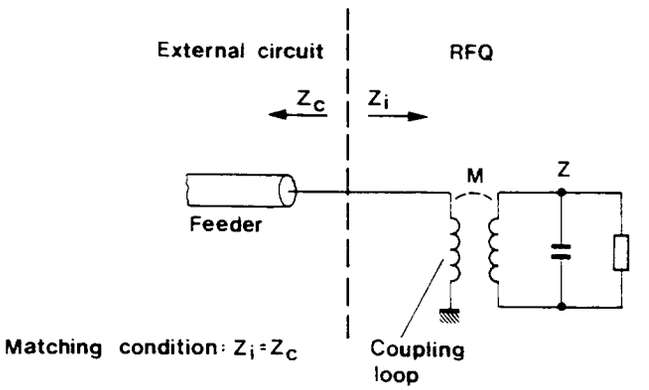


Fig. 16 Matching of the RFQ to the external circuit

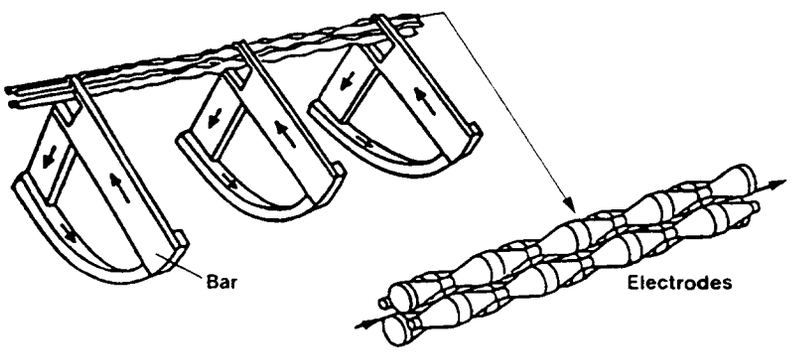


Fig. 17 Frankfurt University RFQ resonator and electrodes

rotationally symmetric, and can be machined on a lathe. However, owing to the heavy currents concentrated in the bars, the losses are higher with such circuits.

5.5 Summary of RFQ characteristics

To end the analysis of the RFQ, we shall briefly summarize its main characteristics:

- it is a linear ion accelerator operating in the low-energy range, typically (for protons) from $\geq$ 20 keV to 2 MeV;
- the low input energy avoids expensive high tension (HT) installations as used with Alvarez linear accelerators;
- it bunches, focuses, and accelerates particles with RF fields;
- typical length: 1 to 2 m;
- operating frequency: ~ 50 to 400 MHz;

- intervane voltage: ~100 kV;
- minimum distance vane-axis: 3 to 4 mm;
- Q factor: ~10,000;
- intervane capacity: $\sim 10^{10}$ F/m;
- impedance: if we define the impedance as $Z = U^2/2P$, where P is the power delivered to the structure to establish an intervane voltage U, we have approximately, for 1 m long structures,

$$Z \cong 5 \times 10^4 \Omega \ .$$

6. 'RFQ APPROACH' FOR OTHER TASKS

In analysing the beam dynamics in the RFQ, we have seen how the well-defined boundaries and the validity of the Laplace equation allowed us to shape fields according to the action we wished to perform on the beam. This action was matching, focusing, and acceleration. There are certainly other actions we could perform by applying the same approach. One of them will be analysed in what follows.

To accelerate very intense beams, we must overcome space-charge forces, which are particularly strong at low energies (at higher energies, the focusing action of the self magnetic field compensates the electric repulsion). The idea is to accelerate two (or more) separate beams at a lower frequency (higher space-charge limit) and then to bring them together (funnel) and to continue to accelerate them at a higher frequency. Figure 18 shows schematically the funnelling of two bunched beams.

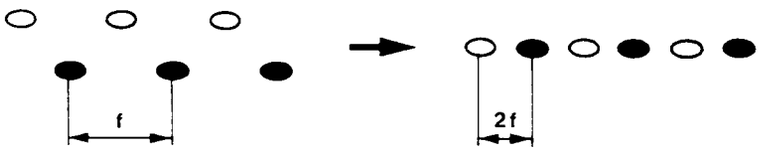


Fig. 18 Funnelling of two beams

How can one funnel the beams? We shall follow, essentially, the approach applied by Stokes of LANL.

Consider an alternating focusing system with regularly displaced quadrupoles (see Fig. 19). The analysis shows that such a system has a dipole and a quadrupole effect on the particles: there is a new, displaced neutral axis, around which the particles will oscillate (see Fig. 20). The displacement of the neutral axis d is much larger than the displacement of the individual quadrupoles a and is always in the direction of the displaced focusing (F) quadrupole.

What happens if, instead of space periodicity, we have a time periodic system? The displacement will depend on the RF phase at which the particle arrives: for two particles 180° (or π) out of phase, the displacements will be in opposite directions, as shown in Fig. 21a; if, however, these particles come in already along their displaced neutral axes, they will stay on them. If we reduce now, in a controlled way, the displacement of the electrodes, the neutral axes (and the particles) will be brought progressively to the z-axis (see Fig. 21b). We have thus achieved a funnelling action. The procedure to follow is:

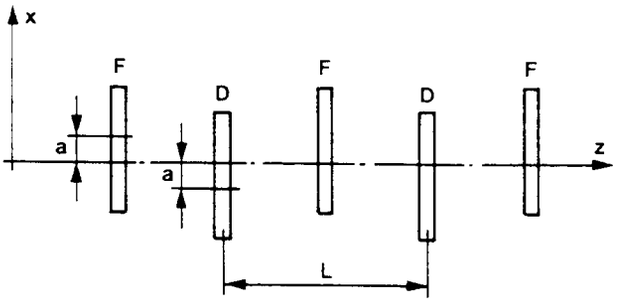


Fig. 19 Periodically displaced quadrupole system

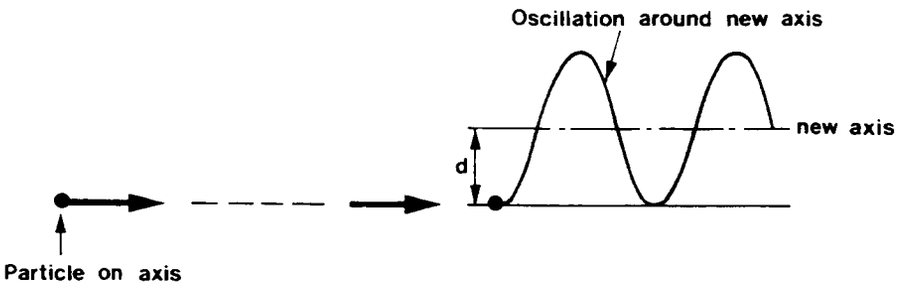


Fig. 20 Oscillation around displaced neutral axis

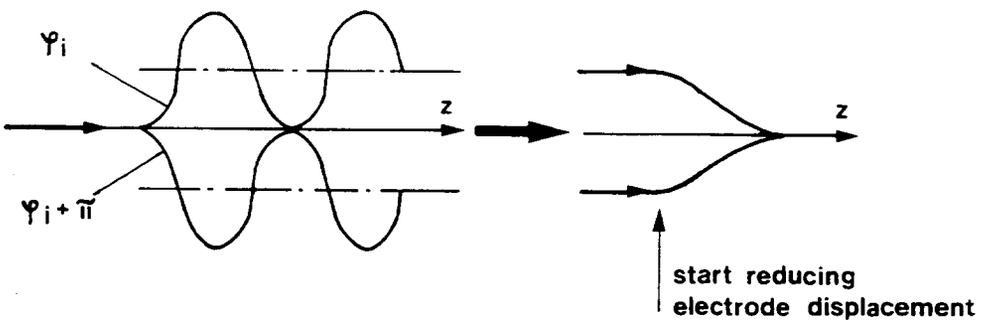


Fig. 21 Principle of funnelling with the RFQ

- find solutions of the Laplace equation, which give fields necessary for funnelling,
- determine corresponding shapes of vanes (electrodes).

The Laplace equation, with conditions for funnelling, has been solved in Appendix D. We rewrite the lowest-order solution and compare it with the previously obtained one for the RFQ accelerator:

$$U(r,\theta,z)=\frac{V}{2}\left[Cr^2\cos 2\theta + D\sinh(kx)\cos kz\right] \quad \text{for the funnel RFQ}$$

$$U(r,\theta,z)=\frac{V}{2}\left[A_{01}r^2\cos 2\theta + A_{10}\mathrm{I}_0(kr)\cos kz\right] \quad \text{for the RFQ accelerator .}$$

The electric fields in the funnel RFQ are

$$\left.\begin{aligned} E_x &= -\frac{V}{2}\left[2Cx + kD\cosh(kx)\cos kz\right] \\ E_y &= \frac{V}{2}2Cy \\ E_z &= \frac{V}{2}kD\sinh(kx)\sin kz \end{aligned}\right\} \sin(\omega t + \varphi)\ .$$

The deflecting force is produced by the second term in the expression for E_x. Averaging over $\beta\lambda$ (neglecting the wiggle) we get

$$\begin{aligned} \overline{E}_{\bar{x}} &= -\frac{V}{2}\frac{1}{2}kD\cosh(kx)\sin\varphi \\ \overline{E}_y &= 0 \\ \overline{E}_z &= -\frac{V}{2}\frac{1}{2}kD\sinh(kx)\cos\varphi\ . \end{aligned}$$

The force $\overline{E}_x$ is an even function of x; it depends on the RF phase and it is in the opposite directions for particles which are 180° out of phase. This is what we wanted.

The force $\overline{E}_z$ also has interesting properties: it is an odd function of x and, depending on the phase φ, can accelerate, decelerate, bunch, or debunch off-axis particles. To get only a bunching action, the bunches must have phase angles as follows:

$$\begin{aligned} &\text{for } +x\text{:}\quad \varphi = -90° \\ &\text{for } -x\text{:}\quad \varphi = +90°\ ; \end{aligned}$$

with these phase angles we achieve also maximum displacement forces.

A tunnel RFQ not only deflects particles to funnel them, but at the same time also focuses and bunches them. Figure 22 compares the vane shapes of a funnel RFQ (as determined by Stokes) and an RFQ accelerator.

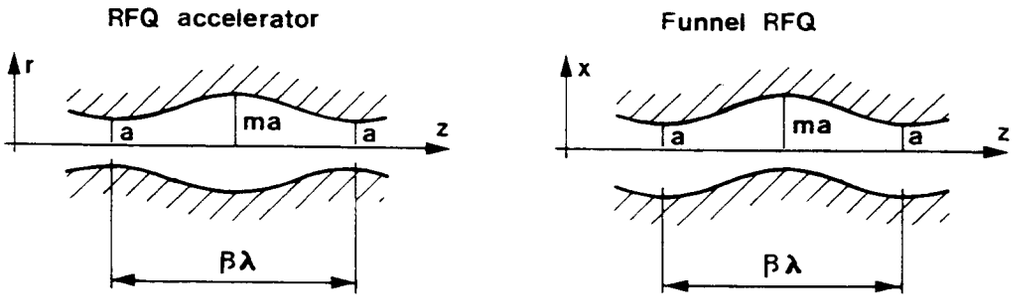


Fig. 22 Vane shapes of the RFQ

7. CONCLUSIONS

In this paper the RFQ has been presented emphasizing its basic principles; the description of practical solutions has not been dealt with. The basic equations of the RFQ and the beam dynamics have been treated in considerable detail. Most of the derived formulae are of a quite general validity in linear periodic structures with low β. Although the expressions are non-relativistic (typical for RFQs), they can relatively easily be brought into the relativistic form and be applied also for βs approaching unity.

To complete the information about the RFQs, some photographs are appropriate. Figure 23 shows the cavity of the CERN RFQ during assembly. This RFQ (designated RFQ1) operates at 200 MHz and accelerates protons from 50 to 520 keV. Figure 24 shows the vanes seen from the input end of RFQ1. One recognizes the radial matching section (RMS) with its characteristic 'opening' of vanes. Also visible are the vane cuts, required for the magnetic flux to turn round. The cylinders which are seen in the cavity walls are tuners which help to symmetrize and tune the RFQ to the right frequency.

Acknowledgements

I acknowledge the valuable help of P. Lapostolle, for reading the manuscript and commenting on it, of R. Stora, for explaining the derivation of the solution of the Mathieu equation, and of A. Lombardi, for assisting in the analysis of potential functions.

* * *

REFERENCES

[1] I.M. Kapchinsky, Ionen-Linearbeschleuniger mit starker Hochfrequenzfokussierung, Teil 1 GSI-Tr-14/76, Darmstadt (1976) (translated from the Russian).

[2] I.M. Kapchinsky and N.V. Lazarev, Linear accelerator structures with space-uniform quadruple focusing, IEEE Trans. Nucl. Sci. NS-26, No. 3, (1979), p. 3462.

[3] I.M. Kapchinsky and V.A. Teplyakov, Linear ion accelerator with spatially homogeneous strong focusing, Prib. Tekh. Eksp., No. 2 (1970) (in Russian).

[4] K.R. Crandall et al., RF quadrupole beam dynamics design studies, Proc. Linear Accelerator Conference, Montauk, NY, 1979 (BNL 51134, Upton, NY, 1980), p. 205.

[5] R.H. Stokes et al., The radio-frequency quadrupole: general properties and specific applications, Proc. 11th Int. Conf. on High-Energy Accelerators, Geneva, 1980 (Birkhauser, Basle, 1980), p. 399.

[6] K.R. Crandall, Proposal for a new radial matching section for RFQ linacs, Los Alamos memorandum AT-1: 83-3 (1983).

[7] H. Klein, Development of different RFQ accelerating structures and operating experience, IEEE Trans. Nucl. Sci. NS-30, No. 4 (1983), p. 3313.

[8] R.H. Stokes and G.N. Minerbo, Proposed use of the radio-frequency quadrupole structure to funnel high-current ion beams, IEEE Trans. Nucl. Sci. NS-32, No. 5 (1985), p. 2593.

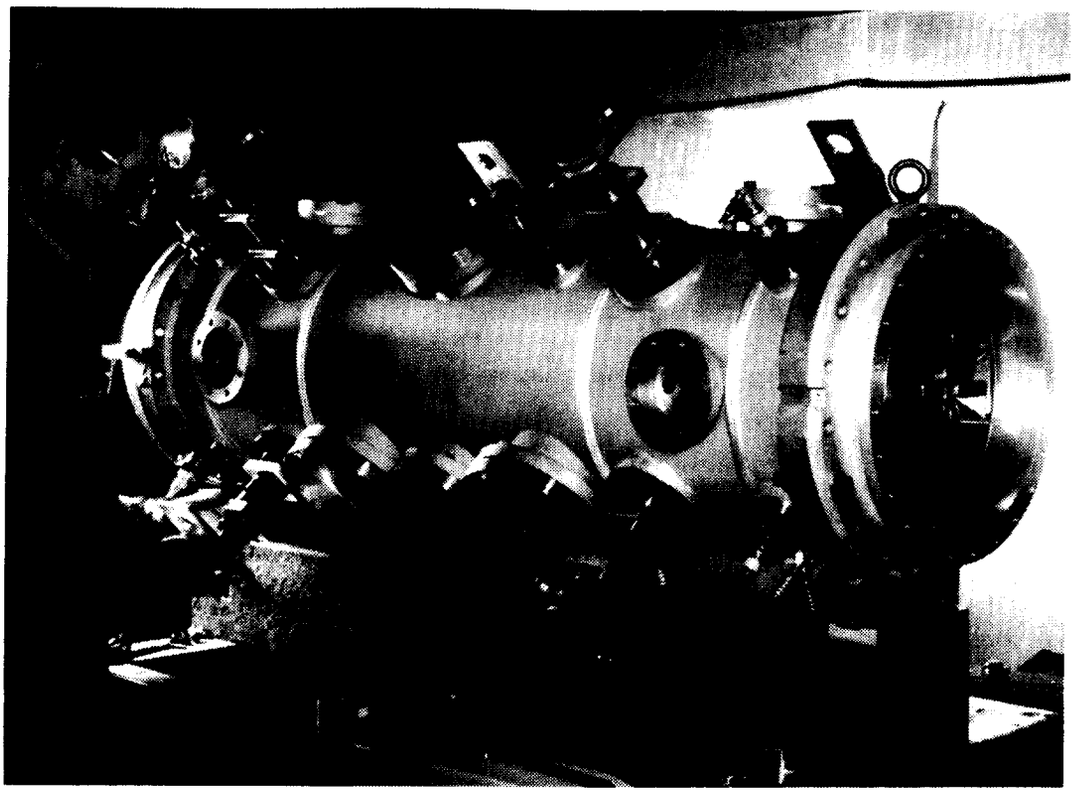


Fig. 23 Cavity of the CERN RFQ (assembling, 1982)

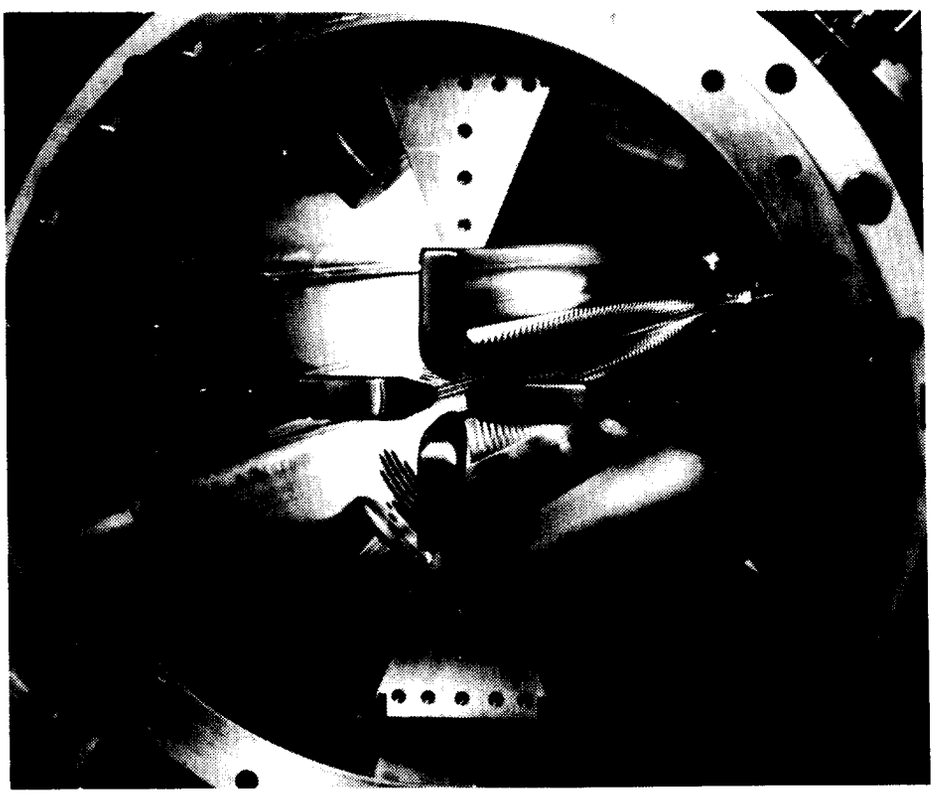


Fig. 24 CERN RFQ: vanes at input end

APPENDIX A

General solution of the Laplace equation applied to the RFQ

The Laplace equation in cylindrical coordinates

$$\frac{\partial^2 U}{\partial r^2}+\frac{1}{r}\frac{\partial U}{\partial r}+\frac{1}{r^2}\frac{\partial^2 U}{\partial \theta^2}+\frac{\partial^2 U}{\partial z^2}=0\ , \tag{A.1}$$

has a solution of the form

$$U(r,\theta,z)=R(r)\ \Theta(\theta)\ Z(z)\ , \tag{A.2}$$

which must satisfy basic conditions imposed by the RFQ:

$$U(r,\theta,-z)=U(r,\theta,z) \tag{A.3}$$

$$U(r,\theta,z+\beta\lambda)=U(r,\theta,z) \tag{A.4}$$

$$U(r,-\theta,z)=U(r,\theta,z) \tag{A.5}$$

$$U(r,\pi\pm\theta,z)=U(r,\theta,z) \tag{A.6}$$

$$-U\left(r,\theta\pm\frac{\pi}{2},\frac{\beta\lambda}{2}-z\right)=U(r,\theta,z)\ . \tag{A.7}$$

Introducing Eq. (A.2) into (A.1) one obtains a set of equations for $R(r)$, $\Theta(\theta)$ and $Z(z)$

$$\frac{1}{Z}\frac{d^2 Z}{dz^2}+k_z^2=0 \tag{A.8}$$

$$\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2}+k_\theta^2=0 \tag{A.9}$$

$$\frac{r^2}{R}\left(\frac{d^2R}{dr^2}+\frac{1}{r}\frac{dR}{dr}\right)-\left(k_z^2r^2+k_\theta^2\right)=0\ , \tag{A.10}$$

where k_z and k_θ are, still undefined, real constants.

The solution of Eq. (A.8) with conditions (A.3) and (A.4) is

$$Z(z)=\sum^{m} a_m \cos\left(m\frac{2\pi}{\beta\lambda}z\right)\ .$$

Note that k_z has been expressed with

$$k_z=m\frac{2\pi}{\beta\lambda}\ ,\qquad m=0,\ 1,\ 2,\ \ldots\ .$$

The solution of Eq. (A.9) with conditions (A.5) and (A.6) is

$$\Theta(\theta) = \sum^{n} b_n \cos 2n\theta$$

with $k\theta$ given by

$$k\theta = 2n \ , \qquad n = 0,\ 1,\ 2,\ \ldots \ .$$

The condition (A.7) links $Z(z)$ and $\Theta(\theta)$:

$$-\cos\,(m\pi)\cos\,(n\pi) = 1 \ ,$$

giving the relation between m and n:

$$m + n = 2p + 1 \ , \qquad p = 0,\ 1,\ 2,\ \ldots \ .$$

If m is 0 or even, n must be odd and vice versa.

To solve Eq. (A.10), we rewrite it by replacing k_z and $k\theta$ with the expressions found above:

$$\frac{r^2}{R}\left(\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr}\right) - \left(m^2k^2r^2 + 4n^2\right) = 0 \ , \tag{A.11}$$

where $k = (2\pi)/\beta\lambda$. For $m \neq 0$, Eq. (A.11) is the modified Bessel equation, which has the modified Bessel functions as solutions: $I_{2n}(mkr)$. For m = 0, Eq. (A.11) becomes

$$r^2\frac{d^2R}{dr^2} + r\frac{dR}{dr} - 4n^2R = 0 \ , \tag{A.12}$$

with solutions of the form

$$C_n r^{2n} \ .$$

The general solution of the Laplace equation with conditions for the RFQ is

$$U(r,\theta,z) = \sum^{n} A_{0n} r^{2n} \cos 2n\theta + \sum^{n}\sum^{m} A_{mn} I_{2n}(mkr) \cos 2n\theta \cos mkz \ ,$$

$$m + n = 2p + 1 \ , \quad p = 0,1,2,\ \ldots$$

Appendix B

Solution of the Mathieu equation

We are interested in solutions of the Mathieu equation

$$\frac{d^2x}{d\tau^2}+\left[A+\frac{B}{2j}\left(e^{j2\pi\tau}-e^{-j2\pi\tau}\right)\right]x=0\ , \tag{B.1}$$

which have the form

$$x(\tau)=e^{j\sigma_T\tau}\sum^{m}c_m e^{j2\pi m\tau}\ ,\quad m=0,\ \pm1,\ \pm2,\ \ldots \tag{B.2}$$

The frequency σ_T is assumed to be much smaller than 2π.

Inserting Eq. (B.2) into (B.1) we get a general expression linking the coefficients C_m:

$$E_m\equiv\left(-\sigma_T^2-4\pi^2m^2+A\right)C_m+\frac{B}{2j}C_{m-1}-\frac{B}{2j}C_{m+1}=0\ .$$

Going to the order n, we have a system of $2n + 1$ equations, from which we determine the smooth frequency σ_T and the coefficients C_m, $|m|\le n$ (all C_m are given in terms of C_0). The system is obtained by keeping the first $2n + 1$ equations, E_0, $E_{\pm1}$, $E_{\pm2}$, ..., $E_{\pm n}=0$ and putting $C_{n+1}=C_{-(n+1)}=0$. The resulting matrix is thin around the diagonal and Hermitian.

If we stop at first order, we have the system:

$$\begin{aligned}
E_{-1}&=\left(-\sigma_T^2-4\pi^2+A\right)C_{-1}-\frac{B}{2j}C_0=0\ ,\\
E_0&=\frac{B}{2j}C_{-1}+\left(-\sigma_T^2+A\right)C_0-\frac{B}{2j}C_1=0\ ,\\
E_1&=\frac{B}{2j}C_0+\left(-\sigma_T^2-4\pi^2+A\right)C_1=0\ .
\end{aligned} \tag{B.3}$$

The determinant of the system must be equal to zero; this determines the frequency σ_T:

$$\begin{vmatrix}
-\sigma_T^2-4\pi^2+A & -\dfrac{B}{2j} & 0\\
\dfrac{B}{2j} & -\sigma_T^2+A & -\dfrac{B}{2j}\\
0 & \dfrac{B}{2j} & -\sigma_T^2-4\pi^2+A
\end{vmatrix}=0$$

$$\sigma_T^2-A=-2\pi^2+\sqrt{4\pi^4+B^2/2}=\frac{B^2/2}{2\pi^2\left[1+\sqrt{1+B^2/\left(8\pi^4\right)}\right]}$$

For $B^2 << 8\pi^4$, we get the simple relation

$$\sigma_T^2 \sim \frac{B^2}{8\pi^2} + A \ . \tag{B.4}$$

Having determined σ_T, Eqs. (B.3) give us the coefficients C_m to first order:

$$C_1 = C_0 \frac{B}{2j} \frac{1}{\sigma_T^2 + 4\pi^2 - A} \sim C_0 \frac{B}{j8\pi^2} \ ,$$

$$C_{-1} = C_0 \frac{B}{2j} \frac{1}{\sigma_T^2 + 4\pi^2 - A} \sim -C_0 \frac{B}{j8\pi^2} \ ,$$

where we have used the fact $\sigma_T^2 - A << 4\pi^2$. Inserting into Eq. (B.2), we obtain the first-order solution:

$$x(\tau) = C_0 \ e^{j\sigma_T \tau} \left[1 + C_0 \frac{B}{j8\pi^2} \left(e^{j2\pi\tau} - e^{-j2\pi\tau} \right) \right]$$

$$= C_0 \ e^{j\sigma_T \tau} \left(1 + \frac{B}{4\pi^2} \sin 2\pi\tau \right) .$$

This solution, valid in principle for $\sigma_T << 2\pi$, is still good enough even for σ_T becoming a larger fraction of 2π. The reason is that the higher-order coefficients decrease rapidly with n:

$$|C_n| < \left(\frac{B}{8\pi^2} \right)^n \frac{1}{(n!)^2} |C_0| \ .$$

APPENDIX C

General solution of the Laplace equation applied to the radial matching section (RMS)

The RMS has vanes modulated in a special way (see Fig. 10).

According to Crandall and similar to the procedure in Appendix A, the solution of the Laplace equation is of the form

$$I_n(mkr)\left[a_n \cos n\theta + b_n \sin n\theta\right]\left[c_m \cos mkz + d_m \sin mkz\right] , \tag{C.1}$$

and the conditions it has to satisfy are

$$U(r,-\theta,z) = U(r,\theta,z) \tag{C.2}$$

$$U(r,\pi \pm \theta,z) = U(r,\theta,z) \tag{C.3}$$

$$-U\left(r,\theta \pm \frac{\pi}{2},z\right) = U(r,\theta,z) \tag{C.4}$$

$$U(r,\theta,0) = 0 \tag{C.5}$$

$$\frac{\partial}{\partial z} U(r,\theta,z)_{z=L_{RMS}} = 0 \ . \tag{C.6}$$

Note that L_{RMS} is the length of the radial matching section. Applying conditions (C.2), (C.3), and (C.4) simplifies the function $\Theta(\theta)$:

$$b_n = 0 , \quad n = 2(2p+1) , \quad \text{with } p = 0,\ 1,\ 2,\ \ldots \ .$$

The conditions (C.5) and (C.6) apply to $Z(z)$:

$$c_m = 0 , \quad k = \frac{\pi}{2L_{RMS}} , \quad m = 2p+1 , \quad \text{with } p = 0,\ 1,\ 2,\ \ldots \ .$$

The general solution for the RMS is

$$U(r,\theta,z) = \sum^{n}\sum^{m} A_{mn} I_n(mkr) \cos n\theta \sin mkz ,$$

where

$$n = 2(2p+1)$$
$$m = 2p+1 , \quad \text{with } p = 0,\ 1,\ 2,\ \ldots$$

APPENDIX D

Solution of the Laplace equation applied to a funnel RFQ

For the funnel RFQ, it is preferable to work in rectangular coordinates. We shall derive the lowest-order solution only. The equation

$$\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}+\frac{\partial^2 U}{\partial z^2}=0 \tag{D.1}$$

has solutions of the form

$$U(x,y,z)=X(x)\ Y(y)\ Z(z)\ ,$$

which lead to a system of equations

$$\begin{aligned}\frac{d^2X}{dx^2}+k_x^2X&=0\\ \frac{d^2Y}{dy^2}+k_y^2Y&=0\\ \frac{d^2Z}{dz^2}+k_z^2Z&=0\end{aligned} \tag{D.2}$$

with

$$k_x^2+k_y^2+k_z^2=0\ . \tag{D.3}$$

The solutions of (D.2) are series of exponential functions with real or imaginary exponents. Applying, as in Appendix A, conditions to $Z(z)$ we get

$$Z(z)=\sum^{n} a_n \cos\ (nkz)\ , \tag{D.4}$$

with

$$k_z=nk=n\frac{2\pi}{\beta\lambda}\ . \tag{D.5}$$

Only two terms of (D.4) will be used. For the first one, $n = 0$, and then Eq. (D.1) reduces to the elliptical equation

$$\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}=0\ ,$$

which is conveniently solved in the complex plane, because analytic functions of the complex variable

$$x+jy$$

are solutions of the Laplace equation in two dimensions. Choosing the real part of the lowest-order solution

$$(x + jy)^2 ,$$

one obtains

$$U_0(x,y,z) = b\left(x^2 - y^2\right) .$$

The modulation of the vanes in the funnel RFQ is chosen so as to deflect the beam in the x-direction. To get the second term of the lowest-order solution of (D.1), we put n = 1 and impose now

$$k_y^2 = 0 .$$

From Eq. (D.3) we get

$$k_x = \pm jk_z ,$$

which gives, with n =1 and Eq. (D.4),

$$k_x = \pm jk ,$$

leading to the following solution for $X(x)$:

$$X(x) = c \cosh (kx) + d \sinh (kx) .$$

To deflect the beam, $X(x)$ must be an odd function:

$$c = 0 , \quad X(x) = d \sinh (kx) .$$

Now the lowest-order solution for the funnel RFQ can be written in the form

$$U(x,y,z) = \frac{V}{2}\left[C\left(x^2 - y^2\right) + D \sinh (kx) \cos kz\right] .$$

The constants C and D are obtained by imposing $U(x,y,z) = V/2$ along the vane, see Fig. 22:

$$C = \frac{1}{a^2}\left[1 - D \sinh (ka)\right] \quad \text{and} \quad D = \frac{m^2 - 1}{m^2 \sinh (ka) + \sinh (mka)} .$$

The expressions for C and D are formally very similar to those for A_{01} and A_{10} of the RFQ accelerator.

SUPERCONDUCTING ACCELERATOR MAGNETS

P. Schmüser
II. Institut für Experimentalphysik, Universität Hamburg, Germany

Abstract
All new large hadron storage rings and colliders are equipped with superconducting magnets which allow to achieve significantly higher particle energies at much reduced operational costs. The basic properties and design criteria of superconducting accelerator magnets are reviewed with special emphasis on the following topics: field calculation and multipole expansion; layout of coils; influence of iron yoke; mechanical tolerances; magnetic forces and stress in the coils; persistent currents in the superconductor filaments and resulting field distortions; prospects for future applications.

1 INTRODUCTION

The vanishing electrical resistance of superconducting coils as well as their ability to provide magnetic fields far beyond those of saturated iron is the main motivation to use superconductor technology in all new big hadron accelerators. The first machine of this kind, the Tevatron at the Fermi National Accelerator Laboratory near Chicago, USA, has been operating as a proton-antiproton collider for many years providing centre-of-mass energies of 1800 GeV with excellent luminosity. The successful dipole and quadrupole magnets developed at FNAL have strongly influenced most later designs of superconducting accelerator magnets. The first dipole prototypes built at DESY as well as similar magnets constructed at Saclay for the Russian UNK project were basically copies of the Fermilab magnet. Essential features like the methods to wind, cure and clamp the coils have been retained in more recent designs such as in the magnets for the proton-electron collider HERA, the meanwhile terminated Superconducting Super Collider SSC in the USA, the large Hadron Collider LHC at CERN and the Relativistic Heavy Ion Collider RHIC at Brookhaven. A high field quality is needed in storage rings. The relative deviation from the ideal dipole or quadrupole field should not exceed a few parts in 10000 to guarantee a reasonable beam intensity for storage times of 24 hours or more. This poses no particular problem with normal magnets whose field distribution is determined by accurately shaped iron yokes. In a superconducting coil, however, the field pattern is governed by the arrangement of the current conductors and a precise coil geometry is of utmost importance. The typical accuracy required in any cross section of the coil is 10 to 20 μm. This precision must be maintained in spite of the huge Lorentz forces acting on the current conductors: the two halves of a dipole coil repel each other with a force of about 10^6 N (100 tons) per meter length at a field of 5 T. The coils are confined by strong clamps ('collars') which take up the Lorentz forces and define the exact geometry. Superconducting magnets have a number of properties which are not found in normal magnets and require careful attention.

1.1 Quenches, degradation, training

A quench is the transition from the superconducting to the normal state. Such a transition will invariably occur if one of the three parameters: temperature, magnetic field or current density exceeds a critical value. The reason may be a conductor motion under the influence of Lorentz forces resulting in a heating of the cable by frictional energy. At high currents a motion of a few μm may be sufficient since only a tiny energy deposition, typically 1 mJ/g, is needed to heat the conductor beyond the critical temperature. If a quench happens in a large dipole or quadrupole the current in the coil must be reduced to zero in a short time interval (typically in less than a second at 5 T) to avoid overheating and possible destruction of the normal conducting part of the coil. The quench protection of a single magnet is straightforward: when a quench is detected the power supply is switched off and the stored magnetic energy is dissipated in a dump resistor. For a chain of magnets connected in series, however, the large inductivity does not allow the current to be reduced to zero in less than a second because then dangerously high induced voltages of many tens of kV would arise. A possible solution is to bypass each magnet in the chain with a diode; if a magnet quenches, the current in the chain is decreased slowly but it is guided around the quenched coil by means of the diode. A reliable quench detection and protection system is one of the most important safety features of a superconducting accelerator. It is equally important to construct the magnets in such a way that they have a high inherent stability against quenches. Many superconducting magnets have shown a phenomenon called *degradation*: the magnet could not be excited up to the critical current of the conductor but quenched at significantly lower values. The reason may be insufficient clamping of the windings or insufficient cooling. If the windings can move slightly under the influence of Lorentz forces the magnet may exhibit *training*: the first quench occurs when a certain part of a winding starts to move; if this part is, after the motion, in a stable position the magnet can be excited to a higher current in the next try. The second quench will then be caused by the motion of another part of the windings. It is quite common that magnets can be 'trained' this way and finally reach the critical current of the conductor after a certain number of training steps. A well-built magnet will not train again when it is cooled down and excited for a second time. The large accelerator dipole and quadrupole magnets can fortunately be built so well that they show little if any training and can be excited to the critical current of the superconducting cable almost in the first step. The essential criteria for such a good performance are a sufficiently high prestress in the coil preventing conductor motion and an optimum cooling by making the coils permeable to helium.

1.2 Persistent magnetisation currents

The advantage of superconducting coils turns into a drawback at low excitations since the eddy currents — induced in any electromagnet when the field is changed — do not decay exponentially but continue to flow as long as the coil is in the superconducting state. These bipolar currents generate higher-order multipole fields which may become intolerably large. A well-known example is the sextupole component measured in all superconducting dipoles but in a machine with very low injection energy like HERA even higher multipoles come in and require a compensation by correction coils. The persistent currents are not constant but exhibit a nearly logarithmic time dependence, quite different from the exponential decay of eddy currents in normal conducting circuits. This unusual

behaviour is partly due to 'flux creep' in the superconductor: some of the magnetic flux bundles trapped at pinning centres may be released by thermal activation. The current in the correction coils has to be adjusted to compensate the drift.

1.3 Iron yoke

Superconducting accelerator magnets are generally equipped with an iron yoke which, however, differs considerably from the yoke of a normal electromagnet. The yoke is a hollow cylinder mounted concentrically around the coil. It serves three purposes:

(1) The inner dipole field is increased by 10 to 40%, depending on the proximity between coil and yoke.

(2) The yoke shields the surroundings against the high inner field.

(3) The stored magnetic energy is reduced, which is an advantage when the magnet quenches.

There has been a long debate on the relative virtues of 'warm' and 'cold' iron yokes, i.e. whether the yoke should be outside or inside the liquid-helium cryostat. The presently favoured solution is a yoke inside the cryostat which surrounds the coil clamped with non-magnetic collars. Magnets of this type have been first constructed for the HERA proton storage ring and are foreseen for UNK and LHC. In the RHIC magnets the iron yoke itself is used to clamp the coils.

This report presents the contents of a one-hour seminar at the CERN Accelerator School and is an abbreviated version of a previous publication [1]. An important topic, stability and quench protection, has been omitted here but can be found in reference [1].

2 FIELD CALCULATIONS FOR SUPERCONDUCTING MAGNETS

2.1 Multipole expansion for a single current conductor

A schematic view of a superconducting dipole for a large accelerator is given in Fig. 1. The length of such a magnet is much larger than its aperture. The current conductors run parallel to the beam on one side and are guided across to the other side in the relatively short coil heads. The dipole magnets are usually not straight but follow the beam orbit; however, the deviation from a straight line is only 18 mm for the 9 m long HERA dipole. Under these circumstances one can consider the magnetic field essentially as two-dimensional and perform a multipole expansion. The magnet axis is chosen as the z direction of a cylindrical coordinate system (r, θ, z), shown in Fig. 2a. In the almost straight section of the magnet all current conductors are parallel to the z axis and can be considered as infinitely long since the transverse dimensions are small. Consider a line current in the positive z direction which is located at the position $(r = a, \theta = \phi)$ (Fig. 2b). The vector potential generated by this current has only a z component:

$$A_z(r,\theta) = -\frac{\mu_0 I}{2\pi} \ln\left(\frac{R}{a}\right) . \qquad (1)$$

Here $R = [a^2+r^2-2ar\cos(\theta-\phi)]^{-1/2}$ is the distance between an arbitrary point $P = (r,\theta)$ and the location of the current. For $r < a$ it is useful to write

$$R^2 = a^2[1-(r/a)\exp(i(\theta-\phi))][1-(r/a)\exp(-i(\theta-\phi))]$$
$$\ln(R/a) = \frac{1}{2}\ln[1-(r/a)\exp(i(\theta-\phi))] + \frac{1}{2}\ln[1-(r/a)\exp(-i(\theta-\phi))] \quad .$$

Now we use the Taylor expansion of the logarithm

$$\ln(1-\xi) = -\xi - \frac{1}{2}\xi^2 - \frac{1}{3}\xi^3 - \ldots - \frac{1}{n}\xi^n - \ldots$$

which converges for arbitrary complex numbers ξ with $|\xi| < 1$. The vector potential and field components are therefore for $r < a$

$$A_z(r,\theta) = \frac{\mu_0 I}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{r}{a}\right)^n \cos[n(\theta-\phi)] \tag{2}$$

$$B_\theta = -\frac{\partial A_z}{\partial r} = -\frac{\mu_0 I}{2\pi a}\sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^{n-1}\cos[n(\theta-\phi)]$$

$$B_r = \frac{1}{r}\frac{\partial A_z}{\partial \theta} = -\frac{\mu_0 I}{2\pi a}\sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^{n-1}\sin[n(\theta-\phi)] \tag{3}$$

$$B_z = 0$$

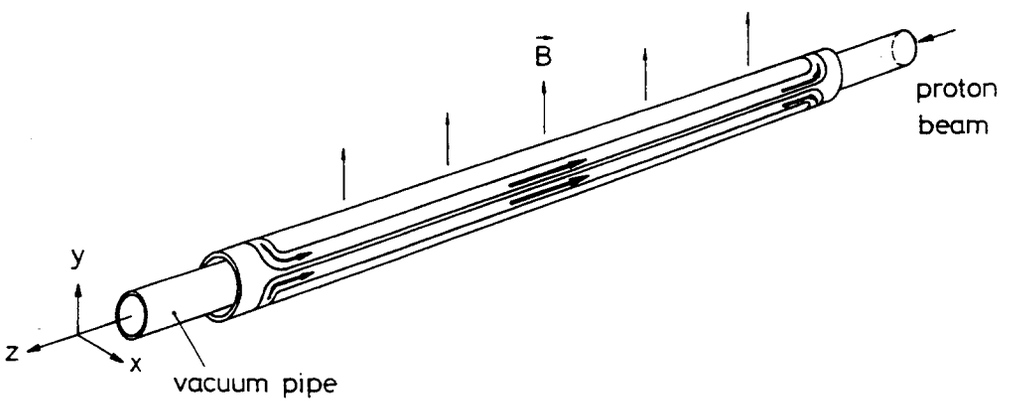


Figure 1: Schematic view of a superconducting dipole coil

2.2 Generation of pure multipole fields

Equation (2) shows that a single line current produces multipole fields of any order n. To obtain a more useful field we consider an arrangement of currents on a cylinder of radius a. A pure multipole field of order $n = m$ is obtained inside the cylinder if the current density as a function of the azimuthal angle ϕ is given by

$$I(\phi) = I_0\cos(m\phi) \tag{4}$$

The statement is easily proved by computing the vector potential resulting from the current distribution (4):

$$A_z(r,\theta) = \frac{\mu_0 I_0}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{r}{a}\right)^n \int_0^{2\pi}\cos(m\phi)\cos[n(\theta-\phi)]d\phi \quad .$$

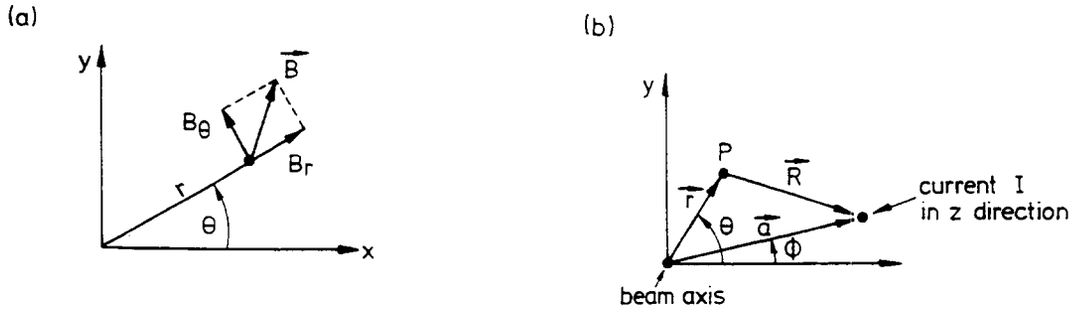


Figure 2: (a) Coordinate system for the multipole expansion (b) Field calculation for a line current.

Using

$$\cos[n(\theta-\phi)] = \cos(n\theta)\cos(n\phi) + \sin(n\theta)\sin(n\phi)$$

and the orthogonality of the trigonometric functions one immediately sees that the integral is zero unless $n = m$, so only a single term in the sum remains:

$$\begin{aligned} A_z(r,\theta) &= \frac{\mu_0 I_0}{2} \cdot \frac{1}{m}\left(\frac{r}{a}\right)^m \cos(m\theta) \\ B_\theta(r,\theta) &= -\frac{\mu_0 I_0}{2a}\left(\frac{r}{a}\right)^{m-1} \cos(m\theta) \\ B_r(r,\theta) &= -\frac{\mu_0 I_0}{2a}\left(\frac{r}{a}\right)^{m-1} \sin(m\theta) \quad . \end{aligned} \tag{5}$$

For $m = 1, 2, 3$ we obtain dipole, quadrupole and sextupole fields. These are shown in Fig. 3, together with the iron pole shoes of the corresponding normal magnets. The fields (5) are the so-called 'normal' multipole fields. If we rotate the current distribution (4) by an angle of $\pi/(2m)$, we obtain a $\sin(m\phi)$ distribution leading to 'skew' multipole fields. A skew dipole, for instance, has a horizontal field. Such magnets are positioned close to the vertically focusing quadrupoles to correct the orbit in the vertical plane. All other skew multipoles are quite undesirable in a circular accelerator. Skew quadrupole fields arise from an angular misalignment of the normal quadrupoles. They have the unpleasant feature of coupling horizontal and vertical betatron oscillations. A few correction quadrupoles, rotated by 45° around their axis, are usually needed to eliminate the coupling.

2.3 Approximation of pure multipole coils by current shells

Current distributions with a $\cos(m\phi)$ dependence are difficult to fabricate with a superconducting cable of constant cross section. In this section we discuss how they can be approximated with sufficient accuracy by current shells. The quality of such an approximation can be judged from the general multipole expansion

$$\begin{aligned} B_\theta(r,\theta) &= B_{ref}\sum_{n=1}^{\infty}\left(\frac{r}{r_0}\right)^{n-1}[b_n\cos(n\theta) + a_n\sin(n\theta)] \\ B_r(r,\theta) &= B_{ref}\sum_{n=1}^{\infty}\left(\frac{r}{r_0}\right)^{n-1}[-a_n\cos(n\theta) + b_n\sin(n\theta)] \end{aligned} \tag{6}$$

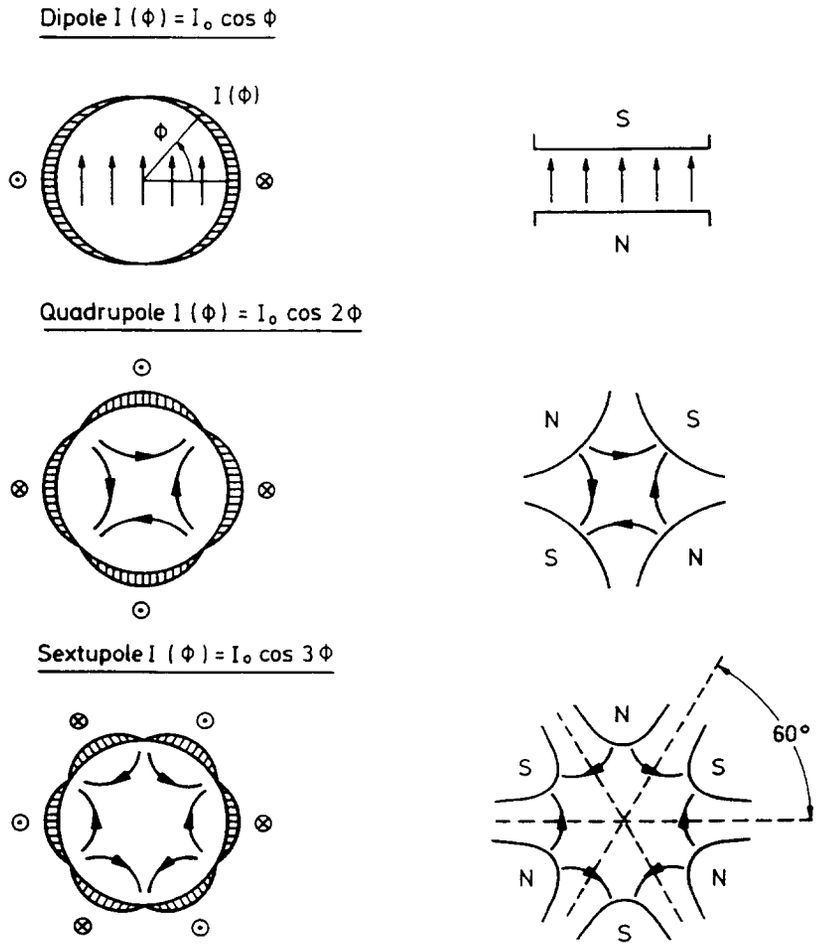


Figure 3: Generation of pure dipole, quadrupole and sextupole fields by $\cos(m\phi)$ current distributions and by conventional magnets with iron pole shoes.

Here r_0 is a reference radius which should be in the same order as the maximum deviation of the protons from the centre axis of the magnet. A reasonable choice is about 2/3 of the inner-bore radius of the coil. B_{ref} is a reference field, for instance the magnitude of the main field at the reference radius, so $B_{ref} = B_1$ in a dipole and $B_{ref} = B_2(r_0) = g \cdot r_0$ in a quadrupole (g is the gradient). The b_n are called the *normal* multipole coefficients, the a_n are the *skew* coefficients. With the above choice of B_{ref} the main coefficient is normalized to unity: $b_1 = 1$ in a dipole, $b_2 = 1$ in a quadrupole. The remaining coefficients should be very small for a good magnet, typically $|a_n|$, $|b_n| < 1 \cdot 10^{-4}$.

We observe that the ideal multipole coils of Fig. 3 have well defined symmetries. In a dipole coil, for any line current $+I$ at an angle ϕ, there are three more currents: $+I$ at $-\phi$ and $-I$ at $\pi - \phi$ and $\pi + \phi$ (see Fig. 4a). The vector potential of these four currents, using Eq. (2), is

$$A_z(r,\theta) = \frac{2\mu_0 I}{\pi} \sum_{n=1,3,5,\ldots} \frac{1}{n} \left(\frac{r}{a}\right)^n \cos(n\phi)\cos(n\theta) \quad . \tag{7}$$

The important consequence is: a coil with dipole symmetry possesses only *normal multipoles* (b_1, b_3, b_5, ...) and no skew multipoles and only *odd* values of n appear.

A coil with quadrupole symmetry (Fig. 5a) has again only normal multipoles and the allowed orders are $n = 2, 6, 10, 14, \ldots$.

The simplest current shell arrangement with dipole symmetry is shown in Fig. 4b. We assume a constant current density J and compute the vector potential inside the coil using Eq. (7)

$$A_z(r,\theta) = \frac{2\mu_0 J}{\pi} \sum_{n=1,3,5,\ldots} \frac{1}{n} \int_{a_1}^{a_2} \left(\frac{r}{a}\right)^n a\, da \int_0^{\phi_l} \cos(n\phi)\, d\phi\, \cos(n\theta) \quad . \tag{8}$$

Here ϕ_l is the limiting angle of the current shell and a_1, a_2 are its radii.
Assuming

$$\Delta a = a_2 - a_1 \ll a = \frac{1}{2}(a_1 + a_2)$$

we get

$$A_z(r,\theta) = \frac{2\mu_0 J}{\pi} a \Delta a \sum_{n=1,3,5,\ldots} \frac{1}{n^2} \left(\frac{r}{a}\right)^n \sin(n\phi_l) \cos(n\theta) \quad . \tag{9}$$

The magnitude of the field of multipole order n is

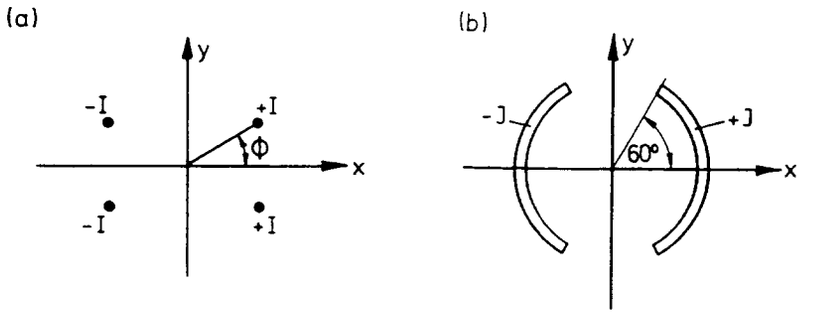


Figure 4: (a) Four line currents with dipole symmetry. (b) Simplest current shell arrangement for a dipole coil.

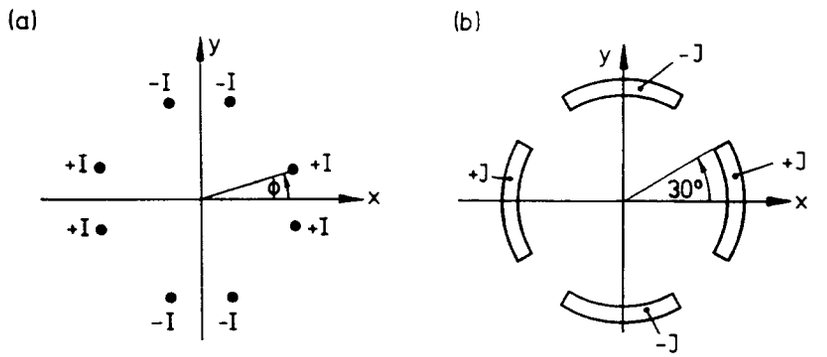


Figure 5: (a) Line current arrangement with quadrupole symmetry. (b) Simplest current shell arrangement for a quadrupole coil

$$B_n = \frac{2\mu_0 J}{\pi} \Delta a \frac{1}{n} \left(\frac{r}{a}\right)^{n-1} \sin(n\phi_l) \tag{10}$$

Choosing a limiting angle $\phi_l = 60°$ the sextupole term $n = 3$ vanishes. Then the first non-vanishing higher multipole is the decapole $n = 5$. For typical coil dimensions the ratio

$$\frac{B_5}{B_1} = \frac{1}{5}\left(\frac{r}{a}\right)^4 \frac{\sin 300°}{\sin 60°}$$

is a few percent, two orders of magnitude larger than is tolerable. A single-layer current shell arrangement is therefore too rough an approximation for a dipole coil. With two current shells, the sextupole and decapole can both be made to vanish by taking a limiting angle of about 72° in the inner and 36° in the outer layer. The Fermilab dipoles are built this way. There remain higher coefficients (b_7, b_9) which are in the order of 10^{-3}. A further reduction of all higher multipoles below the 10^{-4} level is possible by introducing longitudinal wedges into the inner and outer coil layer. The HERA, SSC and LHC magnets are constructed in this manner (Fig. 6a).

A single-layer quadrupole coil with a limiting angle of 45° (Fig. 5b) has a vanishing dodecapole ($n = 6$) but a b_{10} of about 2%. In the Fermilab and HERA quadrupoles two shells with additional wedges are used (Fig. 6b), and then most of the higher multipoles are in the 10^{-4} range or lower.

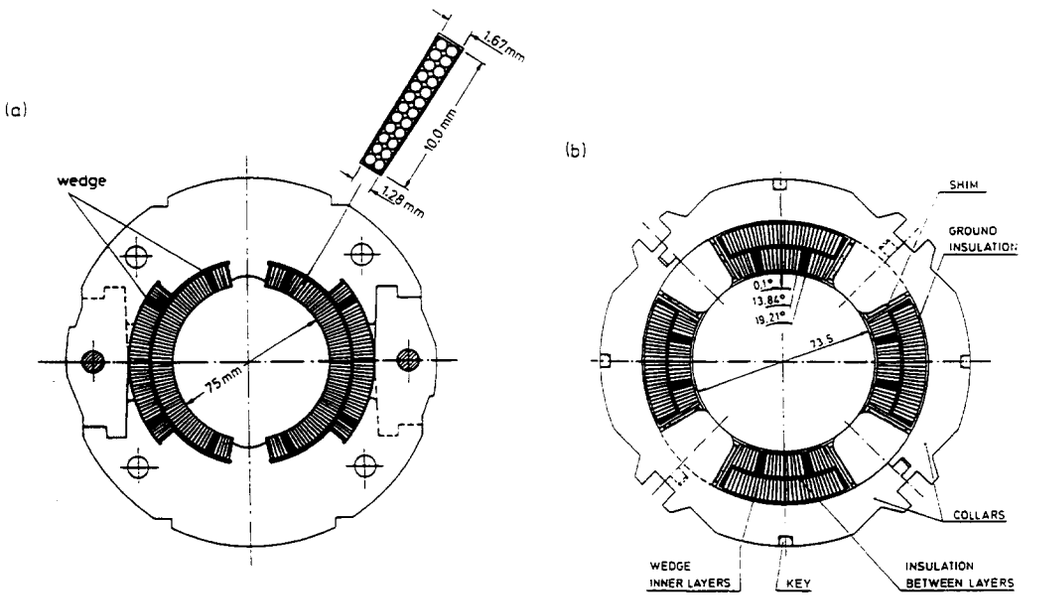


Figure 6: (a) A two-shell dipole coil with longitudinal wedges in the inner and outer layer for improved field homogeneity. Shown in an enlarged view is the superconducting cable which has a trapezoidal cross section. The coil is confined by nonmagnetic collars. This type of coil is used in the HERA, LHC and SSC magnets. (b) Cross section of HERA quadrupole coil [2].

2.4 Influence of the iron yoke

The dipole and quadrupole magnets of an accelerator like the Tevatron or HERA are equipped with an iron yoke with a cylindrical inner bore which confines the magnetic

field. Its influence on the field at the proton beam can be analyzed with the method of image currents provided the iron is not saturated and the permeability μ is uniform. Consider a current I inside a hollow iron yoke whose inner surface is a cylinder of radius R_y. The effect of the iron on the inner field is equivalent to that of an image current I', located at the radius a':

$$I' = \frac{\mu - 1}{\mu + 1} \cdot I \,, \quad a' = \frac{R_y^2}{a} \quad . \tag{11}$$

The image current I' is parallel to the real current I and enhances the inner field. Figure 7 shows the images of a single line current and of a current shell. In the latter case the image current density is lower due to the increased area

$$J' = \frac{\mu - 1}{\mu + 1} \cdot J \cdot \left(\frac{a}{R_y}\right)^4 \quad . \tag{12}$$

For a single-layer dipole coil with concentric iron yoke the nth order multipole field is

$$B_n(r,\theta) = \frac{2\mu_0}{\pi} \sin(n\phi_l) \frac{1}{n} \left[J\Delta a \left(\frac{r}{a}\right)^{n-1} + J'\Delta a' \left(\frac{r}{a'}\right)^{n-1} \right] \quad .$$

Here the first term in the bracket is the coil contribution and the second term is the iron contribution. Now

$$J'\Delta a' = \frac{\mu - 1}{\mu + 1} \cdot J\Delta a \cdot \left(\frac{a}{R_y}\right)^2 \quad .$$

So for $n = 1$ and $\mu \gg 1$

$$(B_1)_{iron}/(B_1)_{coil} = (a/R_y)^2 \quad . \tag{13}$$

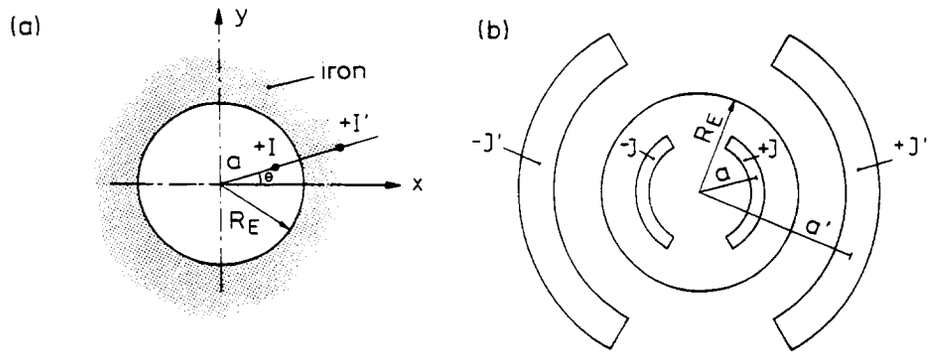


Figure 7: (a) Image of a line current inside a hollow iron yoke (b) Image of a single-shell dipole coil.

As a simple example we consider just the inner coil shell in the HERA dipole whose average radius is $a = 42.5$ mm. The iron yoke radius is $R_y = 88$ mm. In this case the relative iron contribution to the total dipole field on the axis is 19%. For higher multipole orders n the iron contribution is much smaller:

$$(B_n)_{iron}/(B_n)_{coil} = (a^2/R_y^2)^n \quad . \tag{14}$$

For the sextupole field B_3 this amounts to about 1.3% in the above example. The normalized sextupole coefficient $b_3 = B_3/B_1$, however, is reduced by about 18% because of the 19% iron contribution to the dipole field. An important observation is that an unsaturated iron yoke does not create any new multipoles. In a two-layer coil, however, the sextupole and higher poles are modified by the yoke because the mirror image inverts the inner and outer coils. The limiting angles of the coil shells are adjusted in such a way that the sextupole vanishes with the yoke present. Then the collared coil without yoke acquires a non-vanishing sextupole ($b_3 = 13 \cdot 10^{-4}$ in the HERA dipole coil).

2.5 Saturation of iron yoke

The image current method fails when the yoke saturates since the permeability μ is then dependent on position. Finite element programs are needed to compute the field pattern. With iron saturation the dipole field B_1 increases less than linearly with the current I and non-vanishing sextupole and decapole coefficients arise which show a current dependence. The saturation effects depend strongly on the proximity between coil and yoke, and three typical cases shall be considered.

2.5.1 'Warm-iron' dipole

In the Tevatron magnets (Fig. 8) the yoke is outside the cryostat and thus fairly far away from the coil. In this type of magnet saturation is almost negligible up to the critical current of the conductor. The iron contribution to the dipole field is about 10%; the field depends linearly on the current and no higher multipoles are observed.

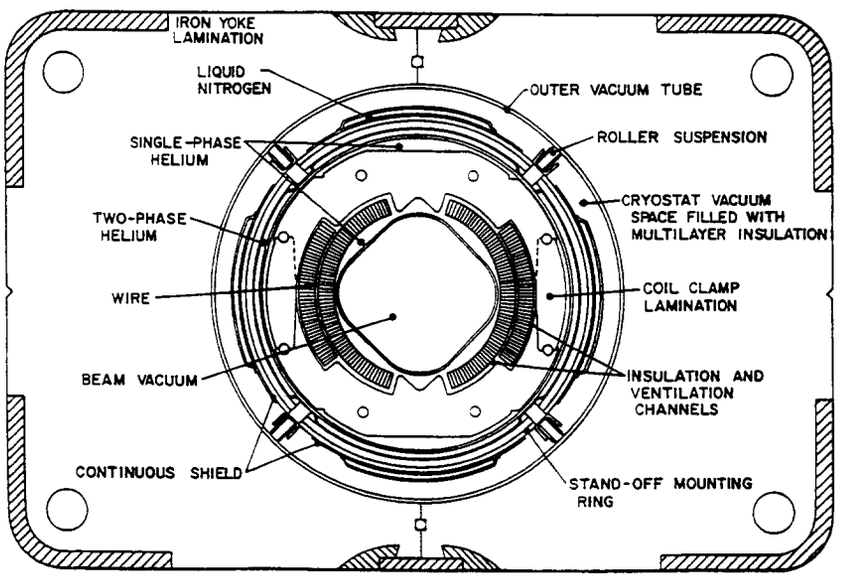


Figure 8: The Tevatron warm-iron dipole [3].

2.5.2 'Cold-iron' dipole

For the Colliding Beam Accelerator project CBA and later the RHIC project, the Brookhaven laboratory has developed a dipole type whose coil is surrounded by a soft-

iron yoke that is contained in the liquid-helium cryostat. The yoke contributes about 40% to the central field, so a substantial saving in superconductor is possible. However, the relation between B and I is quite nonlinear and strongly current-dependent sextupoles and decapoles are present (Fig. 9). Storage ring operation requires a compensation by correction coils, but even then this magnet type appears to be limited to fields below 5 T because the magnet-to-magnet fluctuations of the multipoles would probably exceed the tolerable level even when the average value is corrected for. The operating field in RHIC is only 3.5 T, so the field distortions from yoke saturation are no problem.

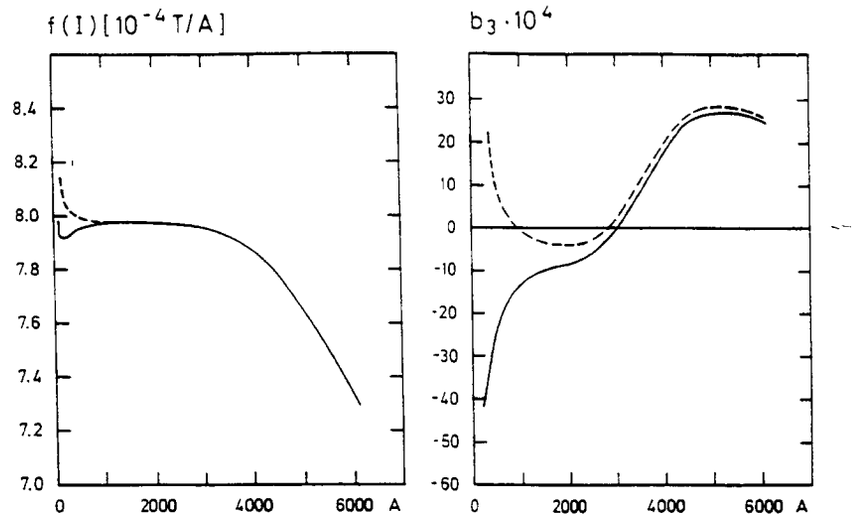


Figure 9: Current dependence of the 'transfer function' $f(I) = B(I)/I$ and of the sextupole coefficient b_3 in the RHIC cold-iron dipole [5]. At low currents the data split into two curves owing to persistent currents in the superconductor (see Sec. 4).

2.5.3 'HERA-type' dipole

A third type, devised at DESY [4], combines the coil of the warm-iron design, clamped by nonmagnetic collars, with an iron yoke inside the cryostat (Fig. 10). Here, the non-linearity in $B(I)$ is quite moderate (0.2% at 5 Tesla) and the sextupole remains small for fields up to 6 Tesla. The LHC dipole combines two coils of opposite polarity in a common iron yoke (Fig. 11). Iron saturation at high field generates a normal quadrupole b_2 that can be corrected by the tuning quadrupoles in the LHC ring.

3 MECHANICAL ACCURACIES AND MAGNETIC FORCES

3.1 Mechanical tolerances

We have seen that in a single-shell coil only the sextupole b_3 can be reduced to zero by adjusting the coil angle to 60°. Other allowed harmonics exceed the 10^{-4} level by far. A more uniform field is accomplished with a two-shell coil like the one used in the Tevatron dipoles. Here two coil angles are available to eliminate both the sextupole b_3 and the decapole b_5, but the 14-pole (b_7) and 18-pole (b_9) components are still in the 10^{-3} range. By introducing wedges into the inner and outer coil layers also these harmonics can be avoided, a method first used in the CBA magnets at Brookhaven. Although one is not completely free in the choice of the coil angles because of the fixed cable thickness

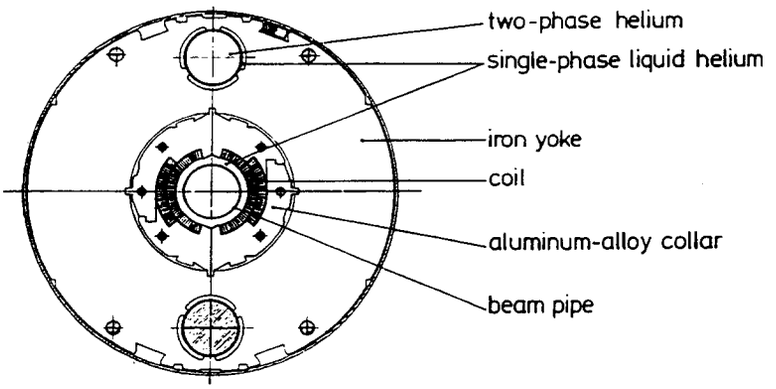


Figure 10: Cross section of the cryogenic part of the superconducting HERA dipole magnet [6,7]. The coil is clamped by an aluminium collar and then surrounded with a cold-iron yoke.

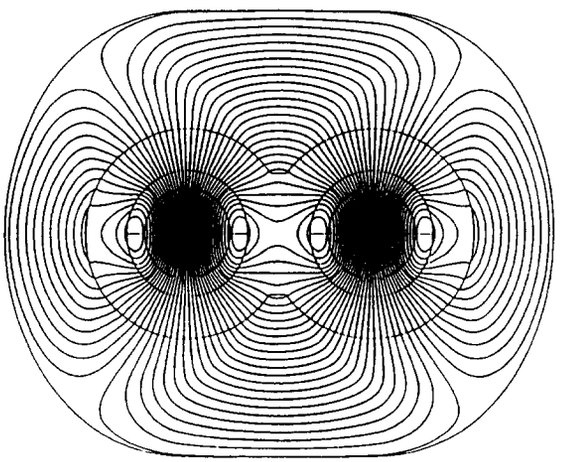


Figure 11: An early version of the 'two-in-one' dipole, foreseen for the LHC proton-proton colllider. Also shown are the computed field lines [8].

an excellent field homogeneity can still be achieved. An example for such an optimised structure is the HERA dipole coil (Fig. 6a) whose calculated harmonics in the straight section are all below 10^{-4}.

In real magnets, the accuracy of the coil angles and radii is limited due to manufacturing errors which are usually in the order of a few hundredths of a millimeter. The influence of a typical geometrical error on the field quality will be discussed for a single-layer dipole coil since the field distortion can be computed analytically. If the limiting angle ϕ_l of the current shell differs from 60° the sextupole coefficient is no longer zero but according to Eq. (10)

$$b_3 = \frac{1}{3}\left(\frac{r_0}{a}\right)^2 \frac{\sin(180° + 3\delta\phi)}{\sin(60° + \delta\phi)}$$

where $\delta\phi$ is the angular error. The condition that $|b_3| \leq 1 \cdot 10^{-4}$ requires $\delta\phi \leq 0.25$ mrad which implies that the arc length of a half coil must be accurate to 10 μm.

Apart from the coil angles the position of the mid-plane between the top and bottom half coils of a magnet may move as a consequence of fabrication tolerances. This is actually quite likely due to the fact that there is no mechanical element in the coil clamping structure which keeps the mid-plane in place, so basically this plane is floating. During collaring the half coils are compressed like two springs and, depending on the relative size of the upper and lower half coil and on their elastic moduli, the mid-plane adjusts itself. A displacement of only 0.1 mm produces already a skew quadrupole (a_2) of $6.5 \cdot 10^{-4}$. Similarly, a tilt of the mid-plane generates a skew dipole $a_1 = 8.4 \cdot 10^{-4}$ and a skew sextupole $a_3 = 2.4 \cdot 10^{-4}$. A method to reduce these distortions is to measure the arc length and the modulus of elasticity on each side of all half coils and combine only half coils that match in arc length and elastic modulus. This tedious procedure was applied for the HERA magnets and for prototype SSC magnets. In spite of this effort mid-plane shifts of up to 50μm remained, resulting in harmonic distortions, especially skew quadrupoles a_2 and skew sextupoles a_3. A detailed analysis [9] revealed a good correlation between the measured skew sextupoles in SSC dipoles and those calculated from the measured coil dimensions.

From the above considerations we conclude that the mechanical accuracies needed to satisfy the field homogeneity requirements are 10 to 20 μm typically. Such narrow tolerances are difficult to accomplish by conventional machining, in particular for 9 m or even 15 m long magnets. Using precision-stamped laminations to assemble the tooling for coil winding and baking and also for the collars which clamp the finished coil one can achieve the required precision at any cross section of the coil.

Another type of distortion is the radial displacement of the coil within the cylindrical yoke. The field of a dipole acquires then a quadrupole component. A horizontal displacement of 0.1 mm leads to a normal quadrupole coefficient of $1 - 2 \cdot 10^{-4}$, for a vertical shift one gets a skew quadrupole of similar size. In a cold-iron magnet, a centring of better than 0.1 mm is no problem since the collars and the yoke can be interlocked by stamped notches and grooves (see Fig. 10) but in a warm-iron dipole many supports and a precise alignment are needed. An interesting aspect is the observation made at the SSC laboratory that an asymmetric arrangement of the dipole magnet inside the steel vacuum vessel of the cryostat may lead to a skew quadrupole at high excitation where the iron yoke is insufficient to confine the entire magnetic flux.

3.2 Multipole measurements as a means of quality control

The quality control measurements which were performed on all HERA dipole and quadrupole magnets yielded a wealth of information on the accuracies which can be achieved in a large-scale industrial production of superconducting magnets. One of the most stringent tests on the precision of the coil cross section and of the placement of the current conductors is provided by the multipole measurements. Figure 12 shows the normal and skew multipole coefficients of the HERA dipoles at the nominal current of 5000 A, corresponding to a field of 4.66 Tesla. Most of the coefficients are very small and well within the limits of $\pm 0.5 \cdot 10^{-4}$, which were used in the particle tracking program to determine the dynamic aperture (maximum stable beam size) of the HERA storage

ring. Two coefficients show a larger scattering: the normal sextupole b_3, which is particularly sensitive to slight changes in the limiting angles of the coil shells, and the skew quadrupole a_2, which may arise from an up-down asymmetry of the coil. The dipoles have been sorted in the ring to minimize the sextupole effects. The general conclusion from the data in Fig. 12 is that a very high mechanical precision has been achieved. In the companies the collared coils were measured at room temperature before their installation in the iron yoke and cryostat. If the sextupole turned out too large, the collars were opened, shims were added and then the collars were closed again. In order to minimize top-bottom asymmetries the two half coils in a dipole were matched with respect to their elastic moduli.

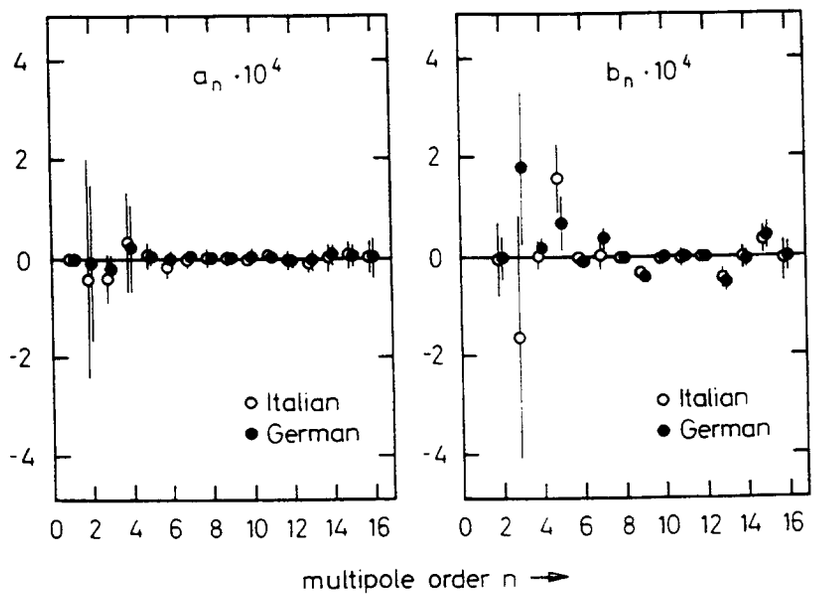


Figure 12: The skew (a_n) and normal (b_n) multipole coefficients of the HERA dipoles at a current of 5000 A, corresponding to a proton energy of about 800 GeV. Plotted are the average values of 200 magnets with their RMS standard deviations. The data have been averaged over the whole length of the magnets, including the coil heads [10]. Half of the dipoles were produced in Italy, the other half in Germany.

Since cryogenic measurements are very time consuming there is considerable interest whether the measurements in the normal-conducting coils at room temperature are sufficiently accurate for determining the field quality. In Fig. 13 I show the correlation between the harmonic coefficients measured in dipole coils with an ac current of 11 Hz and 5–10 A in the normal state and those in complete magnets at 5000 A in the superconducting state.

For most of the coefficients, a good correlation is observed and from the spread of the data one can conclude that room-temperature measurements are adequate to ensure the proper geometry of the coils. This is very important for quality control during production since field measurements without the need of cooling can be performed at the companies.

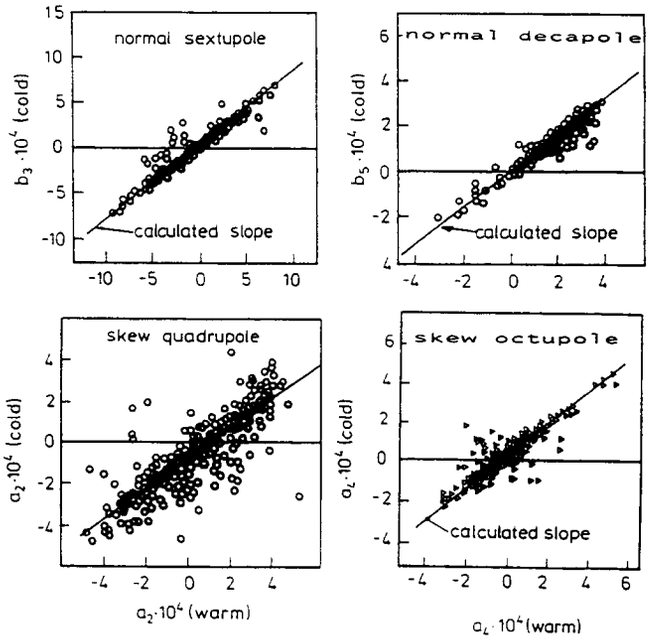


Figure 13: Correlation between room-temperature and cryogenic measurements of multipole coefficients.

3.3 Magnetic forces

We consider as an example the HERA dipole at a field of 5 T and neglect for simplicity the longitudinal wedges inside the coils. The radial and azimuthal components of the Lorentz forces acting on the superconducting cable are plotted in Fig. 14 as functions of the conductor number. In the horizontal plane the radial forces dominate. A conductor in the inner coil shell is pushed outwards with a force of 13000 N per metre, a conductor of the outer shell is pulled inwards with 6000 N/m. Summing over all windings we obtain a horizontal force of 10^6 N for a 1 m long coil section.

Close to the limiting angles of the coil shells the azimuthal forces dominate. They are directed towards the median plane and have the tendency to move the coil package away from its end stop. Such a movement has to be inhibited to avoid frictional heat and quenching.

The azimuthal motion in a half coil will be studied using a simplified model proposed by Tollestrup [11]. The coil package is considered as a system of compressed springs which are confined between two end stops (Fig. 15). The azimuthal magnetic force acts between any two springs and is roughly proportional to the conductor number i (see Fig. 14): $F_i = \alpha \cdot i$. Let k be the spring constant and δX_i the displacement of spring number i. Then we have the following system of difference equations:

$$k(\delta X_{i+1} - \delta X_i) - k(\delta X_i - \delta X_{i-1}) = -F_i = -\alpha \cdot i \quad . \tag{15}$$

This system can be solved by assuming for the displacements X_i a third-order polynomial

$$\delta X_i = ai^3 + bi^2 + ci + d \quad . \tag{16}$$

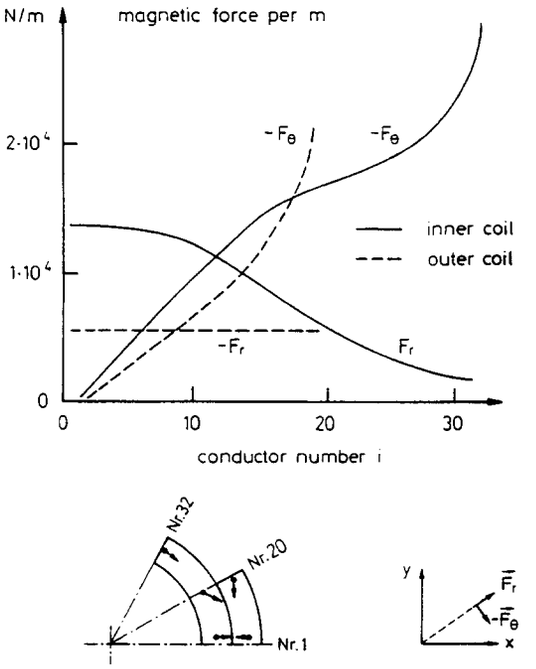


Figure 14: The azimuthal and radial magnetic forces per metre length of conductor at a field of about 5T.

Substituting into (15) and using the boundary condition $\delta X_0 = \delta X_{N+1} = 0$ (we require that the package does not lift off at the end stops) one gets

$$a = -\frac{\alpha}{6k}, \qquad b = 0, \qquad c = \frac{\alpha}{6k}(N+1)^2, \qquad d = 0.$$

The displacement of conductor i is therefore

$$\delta X_i = \frac{\alpha}{6k} i[(N+1)^2 - i^2] \quad . \tag{17}$$

The constant α is about 840 N for the inner shell. The spring constant k is related to the elastic modulus of the coil package. From the series measurements on the HERA dipole coils one obtained $E = (27 \pm 4)$ GPa. This corresponds to a spring constant $k = 1.6 \cdot 10^{11}$ N/m for a 1 m coil section.

The maximum displacement of the conductors under the influence of the Lorentz forces amounts to about 11 μm and is obtained for conductor 20. The resulting magnetic force F_{end} which tries to pull the coil away from the end stop at the limiting angle (Fig. 16) is given by

$$F_{end} = k\delta X_N = \frac{\alpha}{6} N(2N+1) = 2.7 \cdot 10^5 \mathrm{Nm}^{-1} \quad . \tag{18}$$

The coil is 10 mm wide, so this corresponds to a (negative) pressure of 27 MPa at the limiting angle. Choosing a mechanical pre-tension S well in excess of this value one can avoid a coil motion. The mechanical pre-compression in the HERA magnets is equivalent to a positive pressure of about 60 MPa (600 bar). One should bear in mind that the calculation is over-simplified since any friction between the coil and clamp has been neglected and the coil package has been treated as an elastic spring which is far from reality.

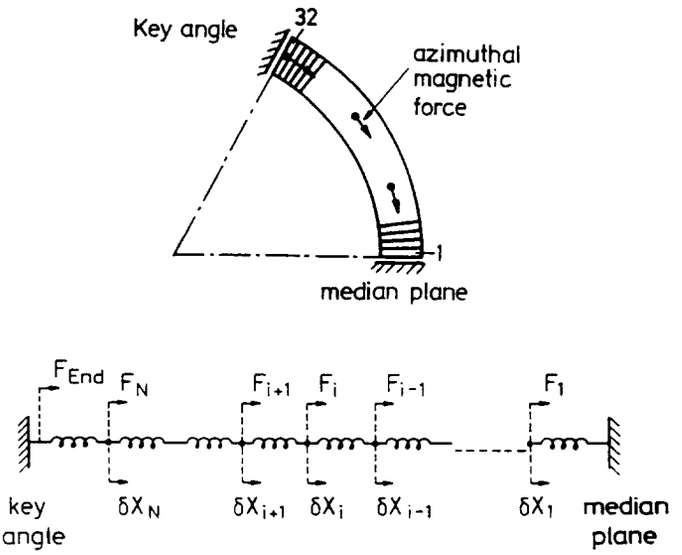


Figure 15: Pre-compressed coil section and equivalent spring model.

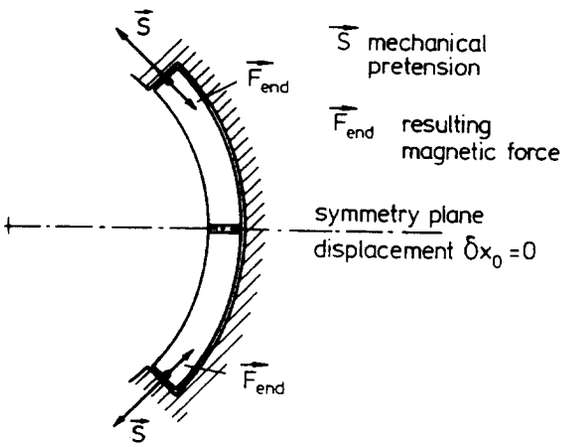


Figure 16: Simplified picture of pre-compressed coil. S: mechanical pre-tension, F_{end}: resulting magnetic force.

3.4 Pre-compression of coil and measurement of internal forces

The collars which confine the coils have to be strong enough to apply the required pre-stress on the coils and to take up the huge Lorentz forces. Figure 17 shows the collared HERA dipole coil and the computed deformation at a field of 6 Tesla. Although the elliptical deformation exceeds by far the limits given in Sec. 3.1 it causes only a small sextupole (less than $1 \cdot 10^{-4}$) since the effects of the increased horizontal radius and the decreased limiting angle cancel each other almost perfectly. For fields above 6 T the deformed collar touches the very stiff iron yoke and no further deflection occurs.

We have seen that the field quality requires a high accuracy of 10 to 20 μm, the quench safety a large internal pre-stress of about 50 MPa. To some approximation, the coil package can be considered as a compressed spring but it is far from being an ideal spring: one observes friction, plastic flow and hysteresis. To ensure that the collars provide the

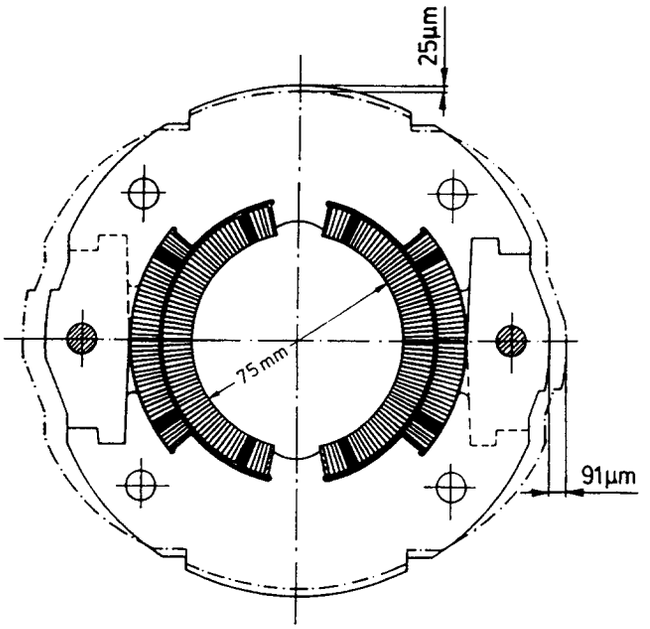


Figure 17: Calculated deformation of the collared HERA coil at 6 T [12]. The collar material is aluminium alloy AlMg4.5Mn(G35) with σ_{02}= 270 MPa and a yield strength of 350 MPa. The maximum calculated stress in the collar is 150 MPa [12]. In a test setup with a mechanical load applied to a short stack of collar laminations the measured and computed deflections agreed to within 5%.

correct geometry and pre-stress after assembly the coils have to be manufactured with a well-controlled oversize which can be determined only experimentally. A further complication arises from the differential shrinkage of the various materials during cooldown. Between room and liquid-helium temperature, the relative shrinkage is

coil package $\cong 3.3 - 3.9 \cdot 10^{-3}$

stainless steel $\cong 3.0 \cdot 10^{-3}$

aluminium alloy $\cong 4 \cdot 10^{-3}$

soft iron $\cong 2 \cdot 10^{-3}$.

If the coil is clamped with aluminium the pre-stress should therefore increase slightly upon cool down whereas with stainless-steel collars it should decrease. Soft iron does not appear very adequate as a collaring material since an enormous room temperature pre-stress would be necessary. This may be dangerous for the Kapton insulation which starts to yield at about 70 MPa at room temperature.

The magnet group at Brookhaven has developed a strain-gauge system to measure forces inside the coil during the assembly in the collaring press and also in cold magnets [13]. As an example I show in Fig. 18 the stress in a 1.8 m long SSC model dipole, measured as a function of current. Starting from 45 MPa at $I = 0$ the pre-stress in the inner coil shell drops more rapidly than in the outer shell due to the larger magnetic forces. But even at 8000 A, far beyond the nominal current of 6600 A, the inner coil shell

keeps a pre-stress of 10 MPa and is thus safe against motion of windings due to Lorentz forces.

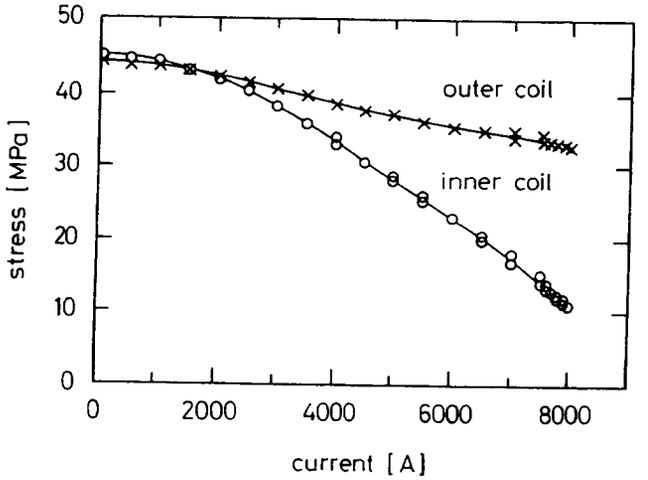


Figure 18: Measured pre-stress in the inner and outer coil shell of a SSC model dipole. I thank C. Goodzeit for providing this figure.

3.5 Forces between coil and yoke

The coil has to be well centred in the yoke to avoid not only field distortions, but also asymmetry forces between coil and yoke. The right half of the dipole coil is attracted by the image currents on the right, the left half is pulled to the left. The two forces balance each other in the case of symmetry but the equilibrium is unstable. If the coil is shifted to one side the force in this direction increases whereas the force in the opposite direction decreases. For a warm-iron magnet the net force between coil and yoke is about 2500 N/m at a field of 4.5 T and a displacement of 0.5 mm. Many supports are needed between coil and yoke to provide a good centring in spite of the asymmetry forces. These supports lead to a substantial heat flux from a warm yoke into the liquid-helium vessel. The cold-iron magnet has a big advantage in this respect since, firstly, the asymmetry forces are small and, secondly, yoke and coil are at the same temperature.

3.6 Longitudinal forces

In the coil heads the Lorentz forces act in the longitudinal direction and tend to lengthen the coil. For the HERA dipole the forces add up to about 15 tons at 5 Tesla. The coil itself can bear these forces; it would elongate elastically by about 3 mm. Such an elongation is undesirable because it changes the field integral and, more dangerously, may lead to frictional heating or slip-stick motion between coil and collars or collars and yoke which might trigger quenches. In some of the early 17 m-long SSC prototype magnets premature quenches were observed which could be traced back to a motion induced by the longitudinal forces. The best solution is to confine the coil heads by stainless-steel end plates which are welded to a longitudinal support structure like the stainless-steel tube serving as the liquid-helium container.

4 PERSISTENT CURRENT EFFECTS IN SUPERCONDUCTING MAGNETS

4.1 Superconductors in a magnetic field

At low excitation superconducting magnets suffer from severe field distortions which are caused by persistent magnetisation currents in the superconductor. These bipolar currents generate all multipoles which are allowed by coil symmetry: b_1, b_3, b_5, b_7, ... in a dipole, b_2, b_6, b_{10}, b_{14}, ... in a quadrupole. A distinct hysteresis behaviour is observed: the multipole fields have opposite signs for increasing and decreasing main field, respectively. Before going into a detailed discussion of these phenomena I will give a brief overview of the magnetic properties of superconductors.

It has been known since the famous Meissner-Ochsenfeld experiment that superconductors show a peculiar behaviour in external magnetic fields. The bulk of a superconducting specimen is shielded from any magnetic field by surface currents provided the field strength stays below a critical value $B_c(\mathrm{T})$. For larger fields superconductivity breaks down. The genuine 'Meissner' effect is observed in type-I superconductors which are in general pure elements like lead or mercury. A wide range of superconductors, mostly alloys but also the important superconducting element niobium, show a more complicated behaviour. These so-called 'type-II' superconductors are characterised by two critical fields, $B_{c1}(\mathrm{T})$ and $B_{c2}(\mathrm{T})$. The upper critical field $B_{c2}(\mathrm{T})$ may exceed by far the critical field $B_c(\mathrm{T})$ of a type-I superconductor. For magnetic fields below $B_{c1}(\mathrm{T})$ a type-II superconductor stays in the Meissner phase with complete field exclusion. For fields between $B_{c1}(\mathrm{T})$, and $B_{c2}(\mathrm{T})$, however, the material enters the mixed phase: magnetic field lines penetrate the bulk material in the form of flux tubes, each containing a single elementary flux quantum $\phi = h/2e$. The flux tubes form a regular hexagonal lattice which can be observed in an electron microscope by spraying fine paramagnetic particles on the specimen which stick to the surface where the flux tubes leave the material (see e.g. [14]). A more advanced method is electron holography.

When the external field is increased, more and more flux tubes move in until near the upper critical field B_{c2} the density becomes so high that the current vortices surrounding the flux tubes start touching each other. If the external field is lowered again, flux tubes move out of the specimen. The response of an ideal type–II superconductor to an increasing or decreasing magnetic field is thus completely reversible. These superconductors, however, are not suitable for electromagnets. If a current flows a Lorentz force is exerted upon the flux tubes which then begin to move through the specimen in a direction perpendicular to the field and to the current. This motion generates heat, so effectively a type-II superconductor acts like an ohmic resistor. A 'flux flow resistance' has indeed been observed [15]. The decisive step towards constructing resistance-free cables for magnets is the introduction of pinning centres which inhibit the motion of the flux tubes. The most important pinning centres are normal-conducting precipitates, mainly titanium in NbTi, and lattice defects. Their properties depend strongly on the various cold-drawing and heat-treatment steps during the wire production and here is where the art of making a good superconductor comes in. A so-called 'hard' superconductor with good pinning centres exhibits no flux flow and consequently no apparent ohmic resistance when currents are flowing in the presence of a magnetic field.

The pinning centres have a strong impact on the magnetic properties. If the external field is raised beyond B_{c1}, the flux tubes entering the specimen are captured. When the field is lowered again, many of the flux tubes remain bound to the pinning centres, so the superconductor retains a frozen-in magnetisation. The result is a hysteresis behaviour similar to that found in iron but with opposite polarity (it is diamagnetic instead of paramagnetic).

Although flux flow does not occur in current-carrying hard superconductors some 'flux creep' effects remain: at non-vanishing temperatures (even at 4.2 K) the thermal energy is sufficient to release a few flux tubes from their pinning centres which then move through the specimen. During their motion they generate an apparent resistance which reduces the critical current density. The first experiments on flux creep, performed by Kim and co-workers [16], showed that the trapped flux in a superconducting cylinder decreases proportionally to the logarithm of time. A simple theoretical model of thermally-activated flux creep [17] is able to account for the logarithmic time dependence of the superconductor magnetisation.

4.2 Theoretical model of persistent magnetisation currents in superconductors

When the field in a magnet is changed, eddy currents are induced in all conducting materials exposed to the time-dependent magnetic flux, in particular of course in the superconducting cable. Most superconducting accelerator magnets are made from the so-called 'Rutherford type' cable shown in Fig. 19. The HERA dipole cable consists of 24 wires ('strands') with a diameter of 0.84 mm. Each strand contains 1230 NbTi filaments of 14 μm diameter that are embedded in a copper matrix. To reduce induction effects the strands in the cable are transposed and the filaments in the strands are twisted.

Eddy currents between different filaments of a strand have to cross the resistive copper matrix and decay exponentially with a time constant [18]

$$\tau = \frac{\mu_0}{2\rho}\left(\frac{l_{twist}}{2\pi}\right)^2 \tag{19}$$

where ρ is the resistivity of the copper and l_{twist} the twist pitch. In the HERA magnets with $l_{twist} = 25$ mm the time constant τ is less than 0.1 s.

A similarly short decay time is obtained for eddy currents between different strands of a cable. Persistent currents can therefore exist only within single filaments. According to the experimentally verified 'critical state' model [19] a hard superconductor tries to expel any external field change by generating a bipolar current distribution with the highest possible density, namely the critical current density $J_c(B,T)$ at the given local field and temperature. Let us study the response of a superconductor filament to a homogeneous external field B_{ext} which is first raised from zero and then lowered again. With increasing B_{ext}, a $\cos\phi$-like current distribution is induced (Fig. 20a) producing a homogeneous internal field B_{int} which just cancels B_{ext} in the current-free region of

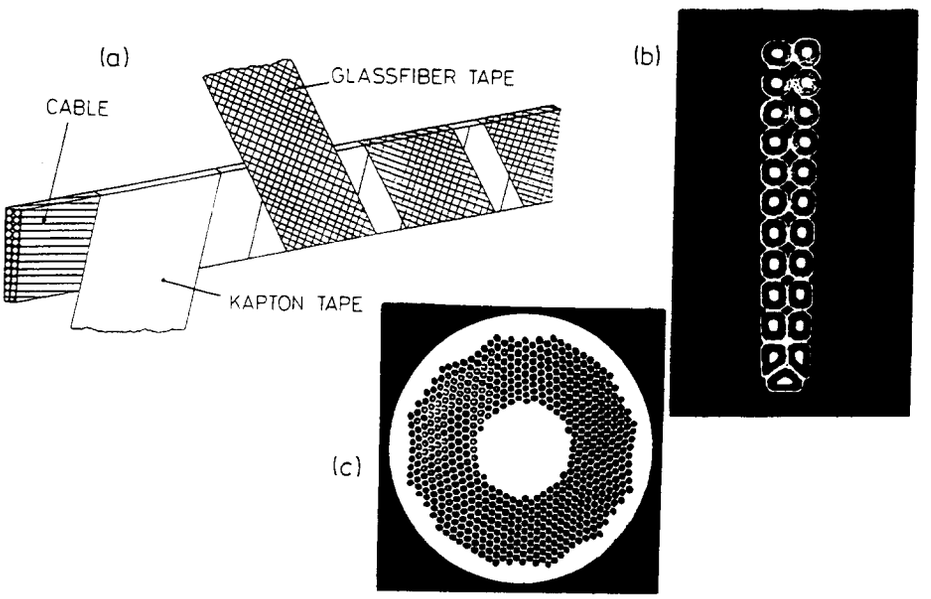


Figure 19: (a) Rutherford-type cable with 'helium-transparent' Kapton and glass-fibre insulation. (b) Cross section of the Vacuumschmelze cable used in the HERA quadrupoles. (c) Cross section of a single strand. The copper appears in white, the NbTi filaments in black.

the filament. This region can be approximated by an ellipse with large half axis $a = r_f$ (filament radius), small half axis b and eccentricity $\varepsilon = \sqrt{1-(b/a)^2}$. The field inside the ellipse, generated by the induced currents, is found by simple integration [18]

$$B_{int} = -\frac{2\mu_0 J_c r_f}{\pi}\left(1-(1-\varepsilon^2)^{1/2}\cdot\frac{\arcsin\varepsilon}{\varepsilon}\right) \quad . \tag{20}$$

The eccentricity ε as a function of the external field B_{ext} is derived from the condition $B_{int} = -B_{ext}$. The highest field which can be shielded from the interior of the filament is called the 'penetrating' field B_p and is obtained for an ellipse shrunk to a line, i.e. $\varepsilon = 1$.

$$B_p = \frac{2\mu_0 J_c r_f}{\pi} \quad . \tag{21}$$

Figure 20b shows the currents in the 'fully penetrated' filament. The applied field may be raised to much larger values than B_p which is only about 0.13 T for the HERA conductor. In that case the same current pattern is obtained as in Fig. 20b but the field inside the filament is no more zero. If now the field is decreased again, persistent currents with opposite polarity are superimposed because the superconductor tries to avoid a change of the inner field. A more complicated current pattern arises as indicated in Fig. 20c. The current loops are assumed to be closed at the coil ends. The effect of the short coil ends on the integrated multipole fields can be neglected.

The magnetisation (magnetic moment per unit volume) of the current distribution

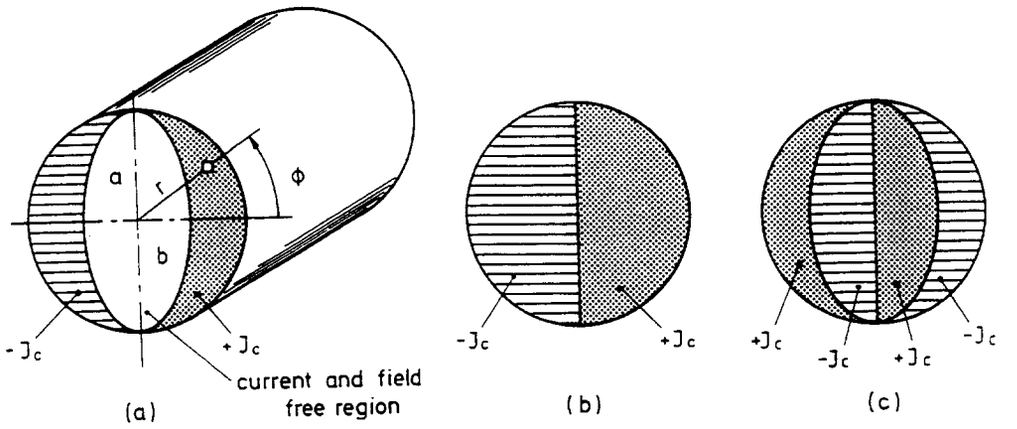


Figure 20: Schematic view of the persistent currents which are induced in a superconducting filament by a time-dependent external field. (a) The external field is raised from zero to a value B_{ext} less than the penetrating field B_p. (b) A 'fully-penetrated' filament, i.e. $B_{ext} \geq B_p$. (c) Current distribution which results when the external field is first increased from zero to a value above B_p and then decreased again.

shown in Fig. 20a is easily computed [18]

$$M = -\frac{4}{3\pi} J_c r_f \varepsilon^2 \quad . \tag{22}$$

The peak magnetization is obtained for the fully penetrated filament

$$M_p = |M|_{max} = \frac{4}{3\pi} J_c r_f \quad . \tag{23}$$

Note that the quantity M_p is not constant but decreases proportional to the critical current density $J_c(B_{ext}, T)$ when the external field is raised beyond the penetrating field.

In the presence of a transport current, the equations (20) to (23) have to be modified by a factor $(1-(J_t/J_c)^2)$. Here J_t is the average density of the transport current, defined as the transport current per filament divided by the filament area πr_f^2. In reality, also the transport current flows with the critical density J_c but is confined to an elliptical region inside the filament similar to the one sketched in Fig. 20a. The correction is negligible near the injection field where $J_t \ll J_c$ but leads to a significant reduction of the magnetisation at high excitation of the magnet.

From the equations (20) and (22) one can compute the magnetisation as a function of the external field. The result is plotted schematically in Fig. 21. We observe a hysteresis behaviour with three different states: Starting at the virgin state the magnetisation follows an initial curve (i) and reaches its peak value at $B_{ext} = B_p$. After going up to high field the ramp direction is reversed and M follows the 'down-ramp' branch (d). At a certain minimum current the field is increased again and the magnetisation follows the 'up-ramp' branch (u) which has the remarkable feature that M changes its sign from positive to negative values. This is exactly what is observed in the 6-pole and 12-pole coefficients (see Fig. 24). In Fig. 21 also the current pattern in the filament is indicated at different positions of the hysteresis loop. An experimental magnetisation curve [20] is presented in Fig. 23.

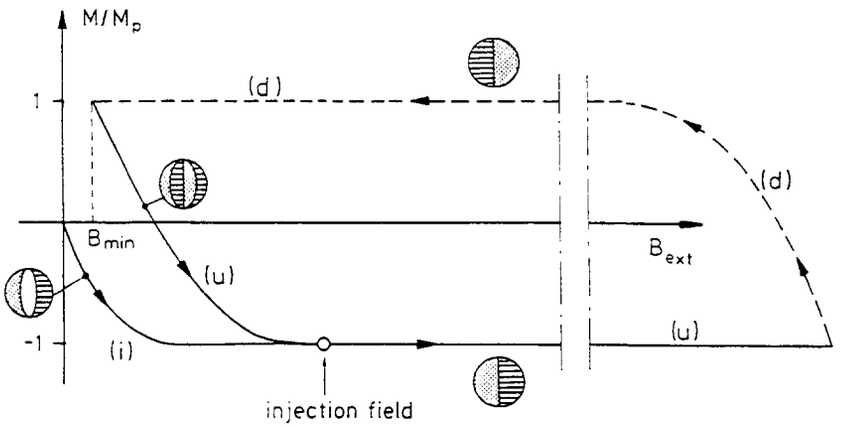


Figure 21: The normalized magnetisation M/M_p of a NbTi filament as a function of the external field. (i): initial curve, (u): up-ramp branch, (d): down-ramp branch. Also shown are the current distributions in the filament.

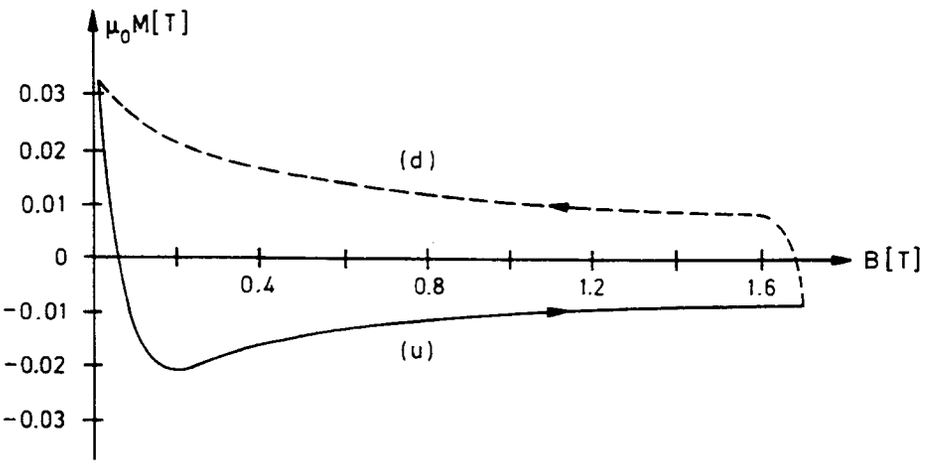


Figure 22: The measured magnetisation of the HERA dipole conductor as a function of the external field [20]. The width of the hysteresis curve decreases towards higher fields due to the field dependence of the critical current density.

4.3 Calculation of persistent-current multipole fields and comparison with data

The field distortions from persistent currents have been computed by various authors [21, 22]. In the following I briefly describe a program [23] developed at DESY which is able to reproduce the experimental results rather accurately.

In the first step the local field is calculated at any current conductor inside the coil of the magnet. Using the formulae in Sec. 4.2 the magnetisation currents are then computed from the time variation of the local field following the complete history. In the third step the fields generated by the bipolar or even more complicated current patterns are determined.

The field calculation will be described for a dipole coil. We start with four symmetrically arranged filaments in the dipole coil in which bipolar currents have been induced

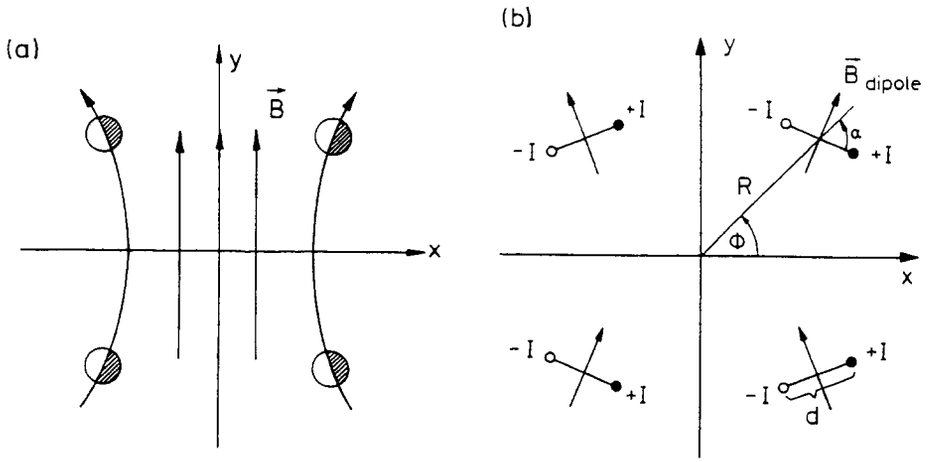


Figure 23: (a) Magnetisation currents induced by the time-dependent main field in four symmetrically arranged filaments inside the dipole coil. (b) Equivalent pairs of line currents. The separation d between the positive and negative currents is grossly exaggerated.

by the increasing main field (see Fig. 23a). The current distribution in each filament can be replaced by a pair of line currents $+I$ and $-I$ whose strength equals the integrated current density and whose separation d is chosen such that the computed filament magnetisation (22) is obtained (d is about equal to the filament radius). Since $d \ll R$, the vector potential of the four current pairs in Fig. 23b can be derived by first-order Taylor expansion from the vector potential A of four single currents (see Eq. (7)):

$$A^{pair} = \frac{\partial A}{\partial R}\Delta R + \frac{\partial A}{\partial \phi}\Delta\phi \quad . \tag{24}$$

With the relations $\Delta R = d\cos\alpha$, $\Delta\phi = -d\sin\alpha/R$ we obtain

$$A^{pair}(r,\theta) = -\frac{2\mu_0 I d}{\pi R}\sum_{n=1,3,\ldots}\left(\frac{r}{R}\right)^n \cos(n\theta)\cos(n\phi+\alpha) \quad . \tag{25}$$

The influence of the iron yoke with an inner bore radius R_y is taken into account by the image current method. The image of a current pair at a radius R and angle ϕ appears at $R' = R_y^2/R$ and $\phi' = \phi$. The separation of the image currents is $d' = d \cdot R'/R$ and the angle with respect to the position vector $\vec{R}'$ is $\alpha' = \pi - \alpha$. Replacing the quantities R, d, α in Eq. (25) by R', d' and α' one gets the iron contribution A'^{pair} to the vector potential. The resulting multipole expansion of the azimuthal field component is then given by

$$\begin{aligned} B_\theta(r,\theta) &= -\frac{\partial}{\partial r}(A^{pair} + A'^{pair}) \\ B_\theta(r,\theta) &= \frac{2\mu_0(I\cdot d)}{\pi R^2}\sum_{n=1.3,\ldots} n\cos(n\theta)\left[\left(\frac{r}{R}\right)^{n-1}\cos(n\phi+\alpha)\right. \\ &\qquad \left. -\frac{R}{R'}\left(\frac{r}{R'}\right)^{n-1}\cos(n\phi-\alpha)\right] \quad . \end{aligned} \tag{26}$$

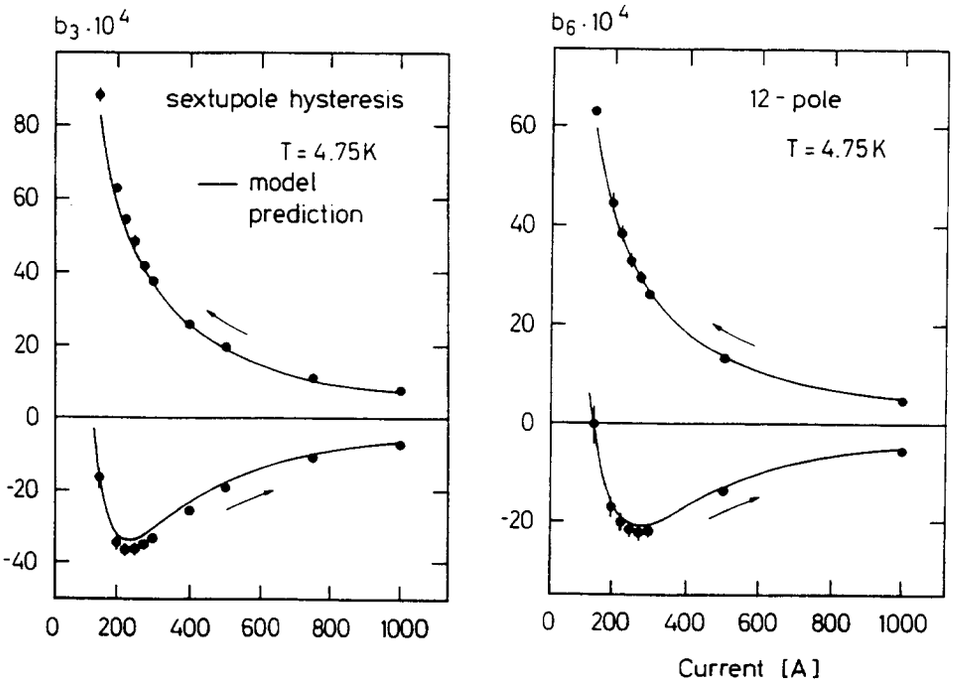


Figure 24: The averaged sextupole coefficients with RMS errors from 315 HERA dipoles (12-pole coefficients from 236 quadrupoles) as a function of the coil current. The multipole coefficients are defined as the ratio of the multipole field to the main field (dipole, resp. quadrupole) at a radius of 25 mm. The curves are model calculations. The ramp direction of the current is indicated by arrows. Before starting the measurements, a current cycle 50 A $\rightarrow$ 6000 A $\rightarrow$ 50 A was performed to establish a well-defined initial conditon for the superconductor magnetisation.

For the product $(I \cdot d)$ we insert the magnetic moment per unit length derived from Eq. (22). Expression (26) has to be summed over all NbTi filaments in one quarter of the dipole coil and divided by the main dipole field to obtain the multipole coefficients. As for the transport-current field, only normal multipoles of the orders $n = 1, 3, 5, \ldots$ occur. In a quadrupole coil, the persistent currents obey again the same symmetries as the transport current and generate therefore only the allowed normal multipoles b_2, b_6, b_{10}, b_{14}, ...

An important ingredient to the model is the critical current density $J_c(B,T)$ at low fields which is unfortunately not easily accessible. The manufacturers of superconducting cables measure critical currents usually at fields of 5 - 6 T. From magnetisation measurements at low fields (Fig. 22) one derives J_c by making use of expression (23) and correcting for the volume fraction of superconductor in the cable. The uncertainty in the critical current at low fields is estimated to be 10 - 20% and is the dominant uncertainty in the calculation of persistent current multipoles. The filament diameter is known to about 5%.

The averaged sextupole data of 315 HERA dipoles and the 12-pole data of 236 quadrupoles are shown in Fig. 24 for increasing and decreasing main fields. The predictions of the model, shown as continuous curves, are in excellent agreement with the

measurements. The persistent currents have also a significant influence on the main dipole field and quadrupole gradient. Their contribution is denoted by $\tilde{B}_1$ resp. $\tilde{g}$ and is plotted in Fig. 25.

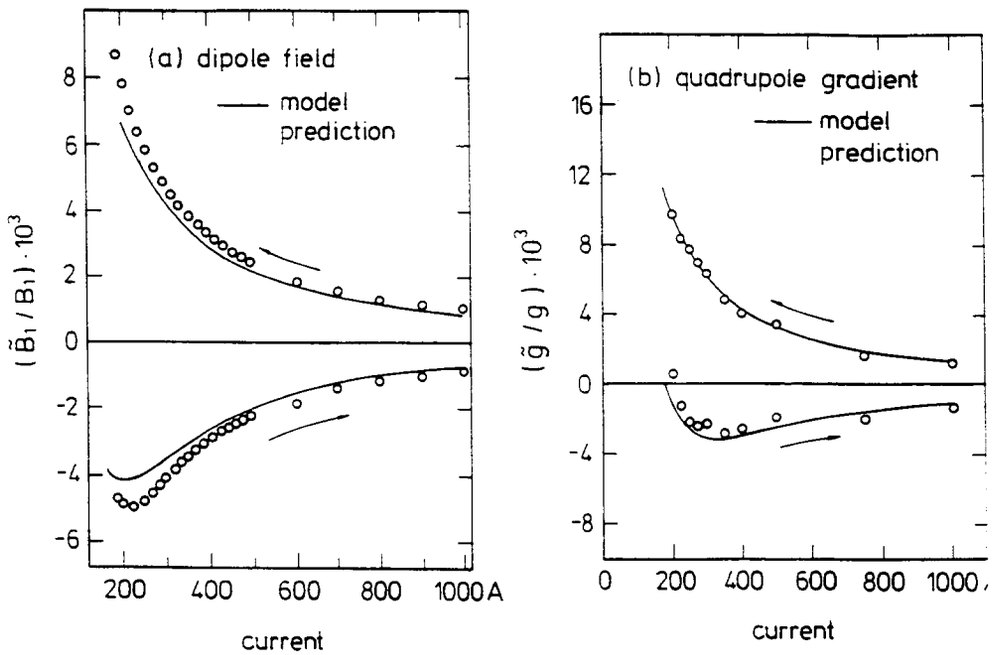


Figure 25: The contribution $\tilde{B}_1$, $\tilde{g}$ of the persistent currents to the main dipole field and the main quadrupole gradient. Solid curves: model prediction.

Again a hysteresis curve is observed and the data are in good agreement with the model prediction. It should be mentioned that the dipole field receives in addition a small contribution from the remanence of the iron yoke. At the HERA injection energy the main dipole field (quadrupole gradient) is 0.5% (0.2%) lower than the value computed from the coil current. Of course a correction is needed to match HERA to the energy of the pre-accelerator.

The superconductor magnetisation and the resulting multipoles are proportional to the filament diameter and the critical current density at low fields which is almost an order of magnitude larger than at 5 Tesla. Of course, nobody wants to sacrifice a high J_c just to reduce the undesirable persistent-current effects but a reduction in filament size is certainly advisable. The cable in the HERA dipoles has fairly thick filaments (14 - 16 μm). When the HERA superconductor was specified more than 10 years ago, a not too small filament diameter was considered important by the manufacturers to guarantee high critical currents. In the past years great progress has been made towards finer filaments. For the SSC magnets, a diameter of 6 μm had been foreseen. A similar number is envisaged for LHC. There is an interesting lower limit, however, at least for NbTi embedded in copper. With decreasing filament diameter the interfilament spacing has to be reduced too to retain the same superconductor fraction. Below 1 μm a 'proximity coupling' between neighbouring filaments has been observed, basically a tunneling of Cooper pairs through the copper, which leads to a steep increase of the multipole fields. The proximity coupling may be inhibited by adding 0.5% of manganese to the copper between the fila-

ments, thereby increasing its resistivity, but there is the danger that the quench stability might suffer when the matrix material has a poor electrical and thermal conductivity.

4.4 Time dependence of persistent current effects

A time dependence of the persistent current sextupole was first observed in the Tevatron dipoles [24]. From measurements on pre-series HERA magnets we were able to show [25] that the multipole fields decrease proportional to the logarithm of time, contrary to the exponential decay of eddy currents in circuits with inductive and resistive components. Similar observations were made at Fermilab [26, 27]. As an example I show in Fig. 26 the time dependence of the dipole and sextupole components in a dipole magnet (a similar behaviour is found for the 12-pole in the quadrupoles).

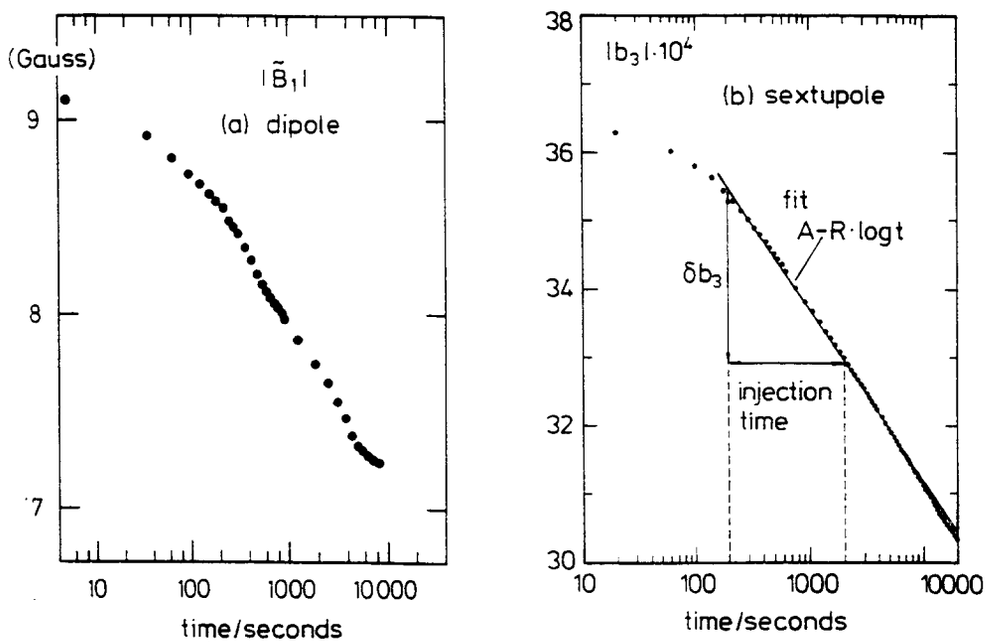


Figure 26: Time dependence of the persistent-current field distortions in a HERA dipole at a field of 0.227 T: (a) absolute value of the contribution $\tilde{B}_1$ to the dipole field (b) sextupole component.

The injection of 40 GeV protons into HERA starts about 10 minutes after the injection field of 0.227 T (corresponding to a magnet current of 245 A) has been established and may take some 20 minutes. During this time interval the drift is well represented by the form $A - R \log t$. The slope R is called 'logarithmic decay rate' and is identical with the change of the multipole fields per decade of time, for instance between $t = 200$ s and $t = 2000$s. The logarithmic decay rates R of the dipole and sextupole components are plotted against each other in Fig. 27 for more than 200 magnets. The time measurements were performed after an initial current cycle

$$0 \text{ A} \rightarrow 6000 \text{ A} \rightarrow 50 \text{ A} \rightarrow 250 \text{ A}$$

in which the maximum current I_{max} was chosen 1000 A above the nominal operating current in HERA. There is a considerable variation from magnet to magnet and in particular

between magnets made from different superconducting cables, as shown in the following table [28]:

Average change 200 → 2000s	$< \delta B_1 >$ (Gauss)	δb_3
Asea Brown Boveri (ABB) cable	0.68 ± 0.18	$(1.77 \pm 0.49) \cdot 10^{-4}$
La Metalli Industriale (LMI) cable	1.15 ± 0.23	$(3.23 \pm 0.62) \cdot 10^{-4}$

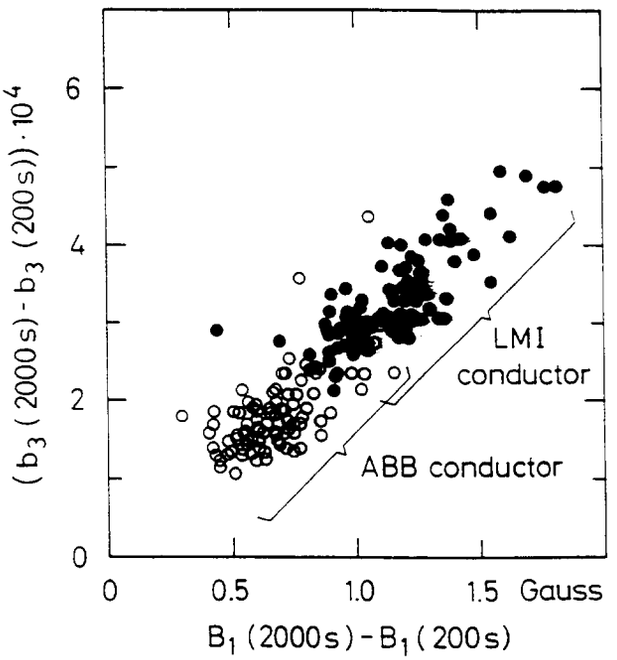


Figure 27: Correlation between the logarithmic decay rates of the dipole and sextupole components.

Secondly, the decay rates of the sextupole and dipole components are strongly correlated indicating a common origin for both decays. The data are compatible with the assumption of a logarithmically decreasing critical current density, suggesting thermally-activated flux creep as the underlying physical mechanism. There are, however, several puzzling observations which do not fit into this picture. (a) The decay rates measured in magnets are usually much larger than those in cable samples [29, 30]. (b) The data in Fig. 28 demonstrate that the persistent-current decay rate at low field is strongly influenced by the maximum excitation current I_{max} of the magnet in the preceding field cycle. In contrast to this, I_{max} has no measurable effect on the shape and width of the sextupole hysteresis curve at low fields. Also in the model discussed above, the superconductor magnetisation at low current, say 250 A, is the same for $I_{max} = 750$ A or $I_{max} = 5000$ A.

In order to clarify the situation an experimental setup was devised at DESY permitting measurement of the average magnetisation of a 5 m long sample of the insulated HERA dipole cable as well as its time dependence. (Usually magnetisation measurements are done on samples of a few mm length so possible effects of the cable pitch cannot be observed.) The results [31] are plotted in Fig. 29. The magnetisation decay at low field (actually $B = 0$ in this case) is less than 1% per decade of time but is totally independent

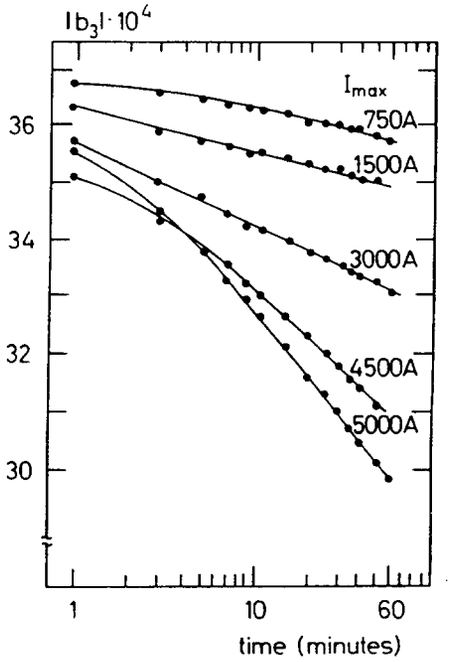


Figure 28: Decay of the sextupole in a HERA dipole at 250 A for different values of the maximum current in the initialising cycle (0 A $\rightarrow I_{max} \rightarrow$ 50 A$\rightarrow$ 250 A).

of the maximum field B_{max} in the preceding cycle. The observed decay rate agrees well with other data on flux creep in NbTi. On the contrary, the sextupole component in the superconducting dipole which was used to generate the external field in the experiment, exhibited the same strong variation of the decay rate with B_{max} as seen in Fig. 28.

Obviously, superconductor properties alone cannot account for the time dependence of persistent-current fields in magnets. Now there is an essential difference between the superconductor in the test sample and that in the coil, namely the transport current. This current is not always shared equally among the strands in the cable and it may even redistribute itself in the course of time, leading to an imbalance: some strands carry a higher current, others a lower one. Experimental evidence for such an imbalance comes from another surprising discovery at DESY [32]. The persistent-currents fields (dipole, sextupole but even 'unallowed' poles) exhibit a periodic variation along the magnet axis, see Fig. 30. The periodicity interval exactly coincides with the transposition pitch length of the strands in the cable. These findings were confirmed at BLN and SSC, and also in their magnets the period agreed with the cable pitch. Both effects: the strong magnetisation decay and the longitudinal periodicity are probably closely linked with a current redistribution in the cable but attempts to construct a quantitative model have failed so far. It is worth mentioning that the oscillating multipole fields should have a negligible effect on the proton beam emittance since their wavelength is orders of magnitude smaller that the betatron wavelength.

5 CONCLUSIONS AND OUTLOOK

The state of superconducting accelerator magnets and their future prospects have been summarized by Perin [33] in an excellent review on which part of the following dis-

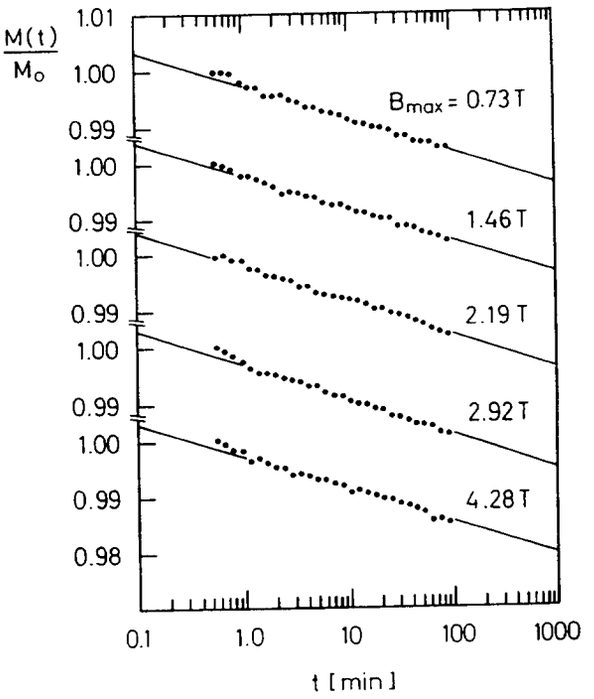


Figure 29: Magnetisation decay at zero field in a long sample of HERA cable for different maximum fields in the initial field cycle $0 \rightarrow B_{max} \rightarrow 0$ with different maximum currents in the field-generating dipole [31].

cussion is based. The last high energy proton accelerators with conventional magnets were built more than 20 years ago: the 300 GeV Proton Synchroton at Fermilab and the 300 – 400 GeV Super Proton Synchroton (SPS) at CERN. Without superconductivity, elementary particle physics might have reached an upper energy limit with these machines because of the prohibitively high electric power consumption with conventional technology in even larger accelerators. In the SPS the normal magnets have a power consumption of 52 MW at 315 GeV compared to some 6 MW which are needed to provide the cooling of the HERA proton ring with a stored beam of 800 GeV.

The pioneering Tevatron at Fermilab has opened the gate to a new generation of high energy hadron colliders. The basic design features and practical construction methods of superconducting accelerator magnets were developed for the Tevatron and in addition this machine has proven that a large superconducting system can be operated with high performance and reliability. In the HERA collider an improved version of the Tevatron magnet is used, made from a higher-performance superconductor and featuring an iron yoke directly surrounding the clamped coil. Contrary to the Fermilab magnets which were all built inside the laboratory, the HERA magnets have been produced in various industrial companies. Excellent field quality and a high degree of operational safety have been achieved in a large scale industrial production. The magnets for the 7 TeV Large Hadron Collider at CERN will also be produced by industry. The favoured superconductor is still niobium-titanium but with a helium temperature below 2 K. A prototype magnet with niobium-tin has been built [34] for the LHC project and successfully tested up to

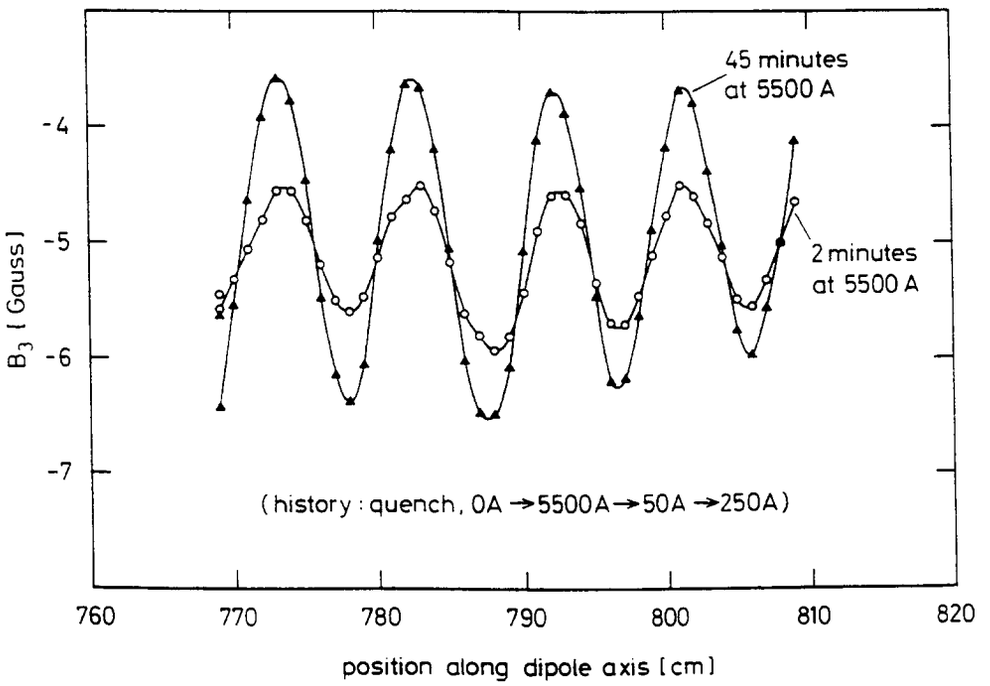


Figure 30: Periodic pattern of the persistent-current sextupole field along the axis of a HERA dipole [32].

fields of more than 9 T, but this superconductor is very expensive and the production methods are complicated because of the brittleness of Nb_3Sn which requires a lenghthy high-temperature treatment of the finished coil to recover good superconductivity. The new high–temperature superconductors are far from being applicable in any accelerator project presently under discussion.

Superconducting accelerator magnets have to be fabricated with high precision to accomplish both the required field homogeneity and the high internal pre-stress which is needed to avoid a movement of the conductors under the influence of the large Lorentz forces. Such movements are dangerous as they generate heat and may lead to a premature breakdown of superconductivity, i.e. a quench. The HERA dipoles have a design field of 4.7 T but all of them can be excited to 6.3 T (at 4.4 K) where the coil current reaches the critical current of the superconductor. The LHC magnets have a smaller inner diameter (56 mm instead of 75 mm) and a larger cable cross section so higher fields are accessible. The requirements on mechanical strength, however, rise steeply, both concerning the cable insulation and the support structure of the coils. The prototype results at CERN are very encouraging in that magnets with safe performance up to 9 T seem to be feasible.

A serious drawback of superconducting accelerator magnets are the field distortions at low excitation which are caused by persistent magnetisation currents. At the Tevatron correction schemes have been worked out which allow to operate the machine without beam loss. In the HERA machine reference magnets are used measure the time-dependent persistent current sextupoles and to control the correction coils. For future accelerators superconducting cables with finer filaments are foreseen to reduce the persistent-current

effects.

In conclusion one can say that superconducting accelerator magnets have reached a state of maturity that makes them appear as reliable and cost-effective devices to achieve proton energies in the TeV regime.

6 ACKNOWLEDGEMENTS

I wish to thank my colleagues and friends at Brookhaven, Berkeley, CERN, DESY, Fermilab, NIKHEF, Rutherford Laboratory, Saclay and SSC for many stimulating discussions. I thank D. Kahnert for his help in preparing the manuscript and S. Turner for carefully reading the text.

References

[1] P. Schmüser, Rep. Prog. Phys. **54** (1991) 683
[2] A. Auzolle, P. Le Marrec, A. Patoux, J. Perot and J.-M. Rifflet, Contribution to the ICFA Workshop on Superconducting Magnets and Cryogenics, Brookhaven, May 1986, BNL report 52006, p. 195
[3] A.V. Tollestrup in: ECFA Study of an ep Facility for Europe, DESY report79/48(1979)
[4] K. Balewski, H. Kaiser and P. Schmüser, DESY internal note, Feb. 1984
[5] E.H. Willen, ICFA Workshop on Superconducting Magnets and Cryogenics, Brookhaven, May 1986, BNL report 52006, p. 35
[6] S. Wolff, 9th International Conference on Magnet Technology, Zürich 1985 and DESY HERA 85/21 (1985)
[7] P. Schmüser, Proceedings of the Oregon APS Meeting, Aug. 1985, p. 968
[8] G. de Rijk, private communication
[9] T. Ogitsu and A. Devred, SSCL-Preprint-472, 1993
[10] H.R. Barton et al., Proceedings of the 11th International Conference on Magnet Technology MT-11, Tsukuba, Japan, 1989, p. 147
[11] A.V. Tollestrup, Fermilab report FN-331 (1980)
[12] G. Meyer, DESY, private communication
[13] C.L. Goodzeit, M.D. Anerella and G.L. Ganetis, IEEE Trans. Mag-25, No. 2 (1989) 1463
[14] W. Buckel, Supraleitung, 4. Auflage 1990, VCH Verlagsgesellschaft Weinheim
[15] Y.B. Kim, C.F. Hempstead and A.R. Strnad, Phys. Rev. 139, A1163 (1965)
[16] Y.B. Kim, C.F. Hempstead and A.R. Strnad, Phys. Rev. Letters 9, 306 (1962)
[17] P.W. Anderson, Phys. Rev. Letters 9, 309 (1962)
[18] M.N. Wilson, Superconducting Magnets, Clarendon Press, Oxford 1983
[19] C.P. Bean, Phys. Rev. Letters 8 (1962), 250
[20] A.K. Ghosh and W.B. Sampson, private communication, see also A.K. Ghosh, K.E. Robins and W.B. Sampson, IEEE Trans. on Magnetics, Vol. 21 (1985), 328
[21] M.A. Green, IEEE Trans., NS-18 No. 3 (1971) 664
[22] J.-L. Duchateau, Department SATURNE international report SEDAP/72-109 (1972)
[23] H. Brück, R. Meinke, F. Müller and P. Schmüser, Z. Phys. **C44**(1989) 385
[24] D.A. Finley et al., Proceedings of the 1987 Particle Accelerator Conference, Washington, p. 151

[25] K.-H. Mess and P. Schmüser, Proceedings of the Course on Superconductivity in Particle Accelerators, Hamburg 1988, CERN yellow report 89-04 (1989)
[26] D.A. Herrup et al., Contribution to the 1988 Applied Superconductivity Conference, IEEE Trans. Mag. 25, No. 2, 1643 (1989)
[27] R.W. Hanft et al. ibid., p. 1647
[28] H. Brück et al., Contribution to the 2nd European Accelerator Conference, Nice 1990
[29] E.W. Collings, Contribution to the Topical Workshop on Magnetic Effects of Persistent Currents in Superconductors, Fermilab March 1990
[30] A.K. Ghosh ibid.
[31] M. Halemeyer et al., Contribution to the 1992 Applied Superconductivity Conference, IEEE Trans. Applied Supercond. 3, No. 1, 168 (1993)
[32] H. Brück et al., Proceedings of the 1991 Particle Accelerator Conference, San Francisco, p. 2149
[33] R. Perin, Proceedings of the 14th International Conference on High Energy Accelerators, Tsukuba, Japan 1989, Particle Accelerators 1990, Vol. 28, p. 147
[34] A. Asner, R. Perin, S. Wenger and F. Zerobin, Proceedings of the 11th International Conference on Magnet Technology (MT-11), Tsukuba, Japan 1989, p. 36

CRYSTAL BENDING AND EXTRACTION
or 'How to build a 1000 Tesla magnet'

Søren Pape Møller
Institute for Synchrotron Radiation, University of Århus, DK-8000 Århus C

Abstract
When charged particles pass through a crystal along a crystallographic direction, the coherent scattering on the lattice atoms forces the particles to follow the lattice direction. The transmission through an aligned crystal is very different from the passage through an amorphous foil, and drastically changes the interaction with the crystal atoms. This channeling effect has many applications at low energy within both fundamental and applied physics. Also when the crystal is bent, the particles will follow the crystallographic direction. In this way, a crystal can deflect a high-energy beam corresponding to a magnetic field around 1000 Tesla, although with some losses. The lecture will be biased towards high energies (> 1 GeV), and results from recent bending and extraction experiments will be presented.

1. INTRODUCTION

A crystal is a regular arrangement of atoms sitting on so-called lattice positions. Looking on a crystal with a magnifying glass, certain directions look open, and the crystal atoms appear to be arranged as pearls on strings or planes. Intuitively, the transmission of charged particles along such directions will be very different from the passage through an amorphous foil, or along a non-aligned direction; so-called 'random' direction. As we will see in Section 2, the motion along such channeling directions is mainly governed by coherent scatterings on strings or planes of atoms, and not by the individual scatterings on single atoms. It is these strings or planes that give rise to immense electric fields over macroscopic distances; i.e. the crystal thickness.

The channeling effect was discovered in the early sixties by the anomalous energy loss of positive ions in crystalline materials [1,2,3]. A comprehensive theoretical treatment was soon given by Lindhard [4] and this work has been the basis for all subsequent work. Actually, already in 1912 Stark [5] had predicted: "In Kristallen werden demnach elektromagnetisch empfindliche Teilchen längs ausgezeichneten Achsen viele größere Schichtdicken zu durchdringen vermögen als in amorphen Körpern. .... Es wird demnach in einem Kristall nicht bloß die Absorption, sondern auch die Zerstreuung von Strahlenteilchen selektiv in verschiedenen Achsen erfolgen." The influence of the channeling effect on the penetration of charged particles through crystals has since been studied by a variety of methods, and the applications of the effect are widespread within both fundamental and applied physics. In particular the applications within semiconductor and ion implantation technology is important.

A new development started in the mid-seventies, when the investigations at high energy began [6]. One motivation was the possibility to use heavy particles of both positive and

negative charge, e.g. π^+ and π^-, as a complement to the investigations at low energy with electrons and positrons. The basic interest in channeling was further revived at both low and high energy by the prediction [7,8] and experimental verification [9,10] of channeling radiation, i.e. the electromagnetic radiation emitted by channeled electrons and positrons.

Another development, of prime interest in the present article, was the prediction [11] and observation [12,13] of beam bending by transmission of charged particles through bent crystals.

The present lecture will give an introduction to channeling with emphasis on qualitative aspects and physical insight. Some essential formulae will be given, and it is the hope that the reader after this lecture will be able to evaluate important quantities. Furthermore, we will confine ourselves to high energies (> 1 GeV) and mainly discuss the applications of channeling in beam bending and extraction, the techniques relying on channeling which are presently feasible. We shall also mainly discuss particles of positive charge, and implicitly assume the projectiles to be protons moving with a velocity close to the velocity of light, unless otherwise explicitly stated.

2. CHANNELING IN A CRYSTAL

2.1 Applicability of Classical Mechanics

Let us at this point make a small digression, since it is not at all clear how to describe the transmission of the projectile through the crystal. As was first discussed by Bohr [14], the quantity for determining whether one can apply classical mechanics for an orbital description of scattering in a Coulomb field is $\kappa \equiv 2Z_1Z_2e^2/\hbar v$. Here Z_1 and Z_2 are the projectile and target atomic numbers and v is the projectile velocity. For $\kappa > 1$ the classical picture is a good approximation. As is easily seen, for fast particles this Bohr condition is not fullfilled. This condition is even more restrictive, when screening from the atomic electrons is taken into account. Hence it may appear surprising, that channeling can be described by classical mechanics if the projectile mass is large compared to the electron rest mass [4]. Qualitatively, one may argue that in a collision with a string of atoms it is the charge of all the nuclei in the string participating in the collision that should be inserted in the Bohr condition. Strictly, one may derive a 'transverse Bohr condition', $\kappa_\perp \equiv [(\gamma M/m)Z_1Z_2^{1/3}a_0/d]^{1/2} \gg 1$, for classical mechanics to apply in the so-called transverse plane. Here $\gamma \equiv (1-v^2/c^2)^{-1/2}$, M and m the particle and electron rest mass, $a_0 \equiv \hbar^2/me^2 = 0.511$ Å the Bohr radius and d the atom spacing. As is easily seen, this condition is fullfilled for heavy particles, but maybe more surprisingly also for electrons/positrons with large relativistic mass.

2.2 The Continuum Approximation

For charged particles incident on a crystal with a small angle to an axial direction, there is a strong correlation between the successive collisions with the string atoms. In other words, the impact parameter changes slowly between subsequent binary collisions, see Fig. 1. This led Lindhard to the introduction of the continuum model, in which the atomic string is replaced by a continuum string obtained by smearing the atomic charges along the string. The channeling picture is now based on a separation between the longitudinal motion along the crystalline direction and the transverse motion, since the interaction between the projectile and the continuum string is independent of the longitudinal coordinate z.

This continuum approximation, outlined above for the atomic string, can also be used when a particle is incident on a crystal with a small angle to a plane of atoms. In this case the atomic charges are smeared over the plane. The transverse motion is two dimensional for the axial case and one dimensional for the planar case, and it is this transverse motion which describes the motion of channeled particles.

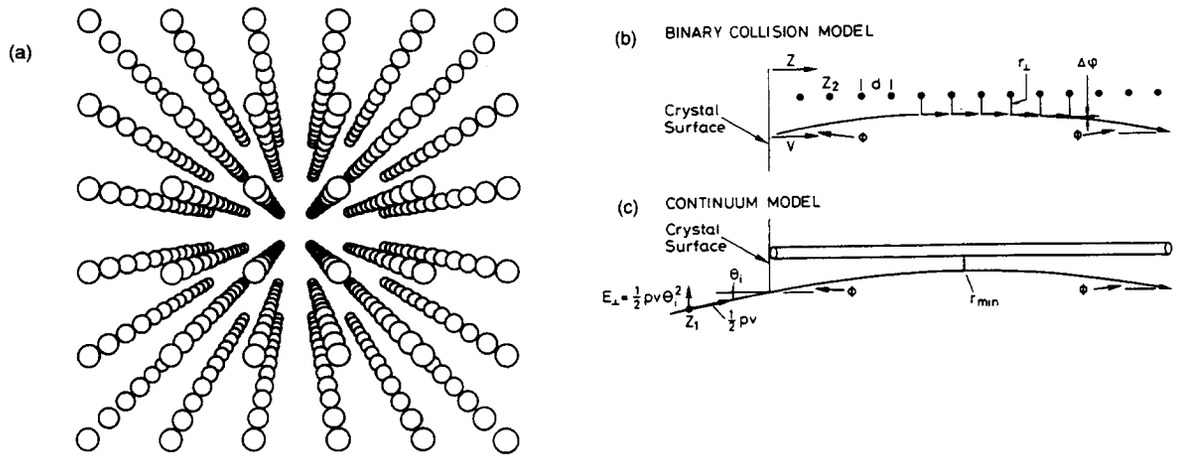


Fig.1 a) Perspective view of simple crystalline structure. b) and c) Schematic drawing of a trajectory in the binary and continuum model

The interaction between the projectile and the string or plane is described by continuum potentials. The string potential is obtained by averaging the atomic potential $V(r)$ along the string, i.e.

$$U(\vec{r}_\perp)=\frac{1}{d}\int_{-\infty}^{\infty} V(\vec{r}_\perp,z)\,dz \tag{1}$$

where $r_\perp$ is the coordinate in the plane perpendicular to the string, i.e. the *transverse plane*. Correspondingly, in the planar case the planar continuum potential takes the form

$$Y(y)=Nd_p\int V(\sqrt{x^2+y^2+z^2})dxdz \tag{2}$$

where N is the atomic density, d_p the planar spacing, and y the transverse coordinate.

In this continuum model the so-called *transverse energy* $E_\perp$ is a conserved quantity, since the potential is independent of the longitudinal coordinates, z and x,z in the axial and planar cases, respectively. This transverse energy is to a very good approximation given by

$$E_\perp=\frac{p_\perp^2}{2\gamma M}+U(\vec{r}_\perp)=\tfrac{1}{2}pv\psi^2+U(\vec{r}_\perp) \tag{3}$$

where the first term is the kinetic energy and the second the potential energy. Here p is the momentum of the particle and ψ the local angle to the string/plane direction. In the continuum approximation, the transverse energy of a given particle is given by it's initial value, i.e. the value it acquired at the surface of the crystal,

$$E_\perp=E_\perp(z=0)=\tfrac{1}{2}pv\psi_0^2+U(r_{\perp 0})\,. \tag{4}$$

This also means that even for a parallel beam perfectly aligned with the crystalline direction, there will be a distribution in transverse energy; an effect called *surface transmission*. In Fig. 2 are shown the single-string potential for the <100> string in Ge at room temperature, and examples of trajectories in the transverse plane of three particles with different transverse energies.

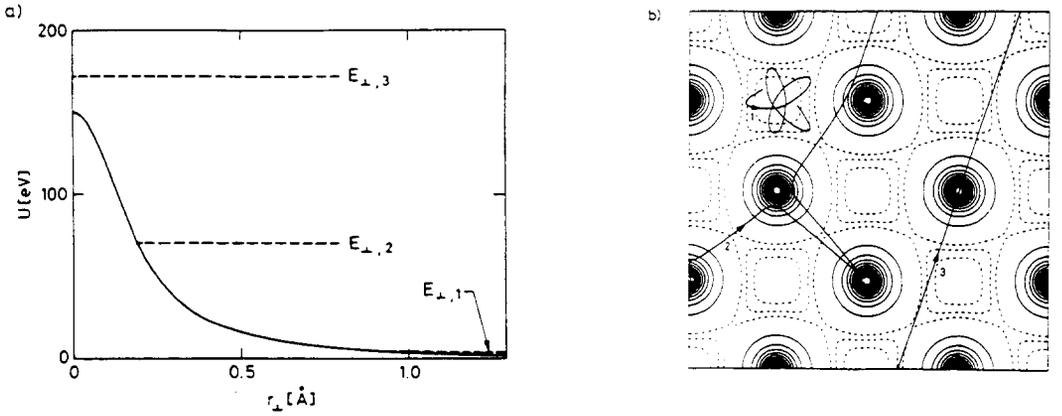


Fig. 2 a) Single-string potential. b) Motion in transverse plane

Previously, we have implicitly assumed the projectile to have a positive charge. Also negative particles may be channeled, and the main change in the above picture is a change of sign in the potentials, which means that contrary to positive particles which are repelled from the strings and planes, negative particles are focused around the strings and planes of atoms.

2.3 Critical Angles

An important quantity, directly observable, is the *critical channeling angle*. This is defined here as the angle of incidence which allows the particle to penetrate into the string or plane. Hence it can be found by equating the transverse kinetic energy to the maximum potential height,

$$E_\perp^{crit} = \tfrac{1}{2} pv\psi_{crit}^2 = U^{max}. \tag{5}$$

The critical angle is approximately given by the Lindhard critical angles ψ_1 and ψ_p for the axial and planar case respectively,

$$\psi_1 = \sqrt{\frac{4Z_1Z_2e^2}{pvd}}, \qquad \psi_p = \sqrt{4Z_1Z_2e^2Nd_pCa/pv}. \tag{6}$$

In the formula for the planar angle, $C \approx \sqrt{3}$ is a constant and $a = 0.8853a_0(Z_1^{2/3}+Z_2^{2/3})^{-1/2}$ a screening length. For the <111> axial and (110) planar direction in silicon, we have

$$\psi_1 = 15\mu\text{rad}/\sqrt{p[\text{TeV/c}]}$$
$$\psi_p = 5\mu\text{rad}/\sqrt{p[\text{TeV/c}]} \quad . \tag{7}$$

Particles with $E_\perp < E_\perp^{crit}$ are said to be *channeled*, and these particles will in first approximation stay channeled. Particles with $E_\perp >> E_\perp^{crit}$ are called *random* particles and are moving through the crystal as if the target was amorphous.

In Fig. 3 are shown calculated surface-transmission probabilities, i.e. the fraction of the beam initially channeled, for 450 GeV protons in silicon at room temperature using the so-called Doyle-Turner potential, see below. The surface transmission is calculated as function of beam divergence for incidence parallel to the (110), (100) and (111) planes.

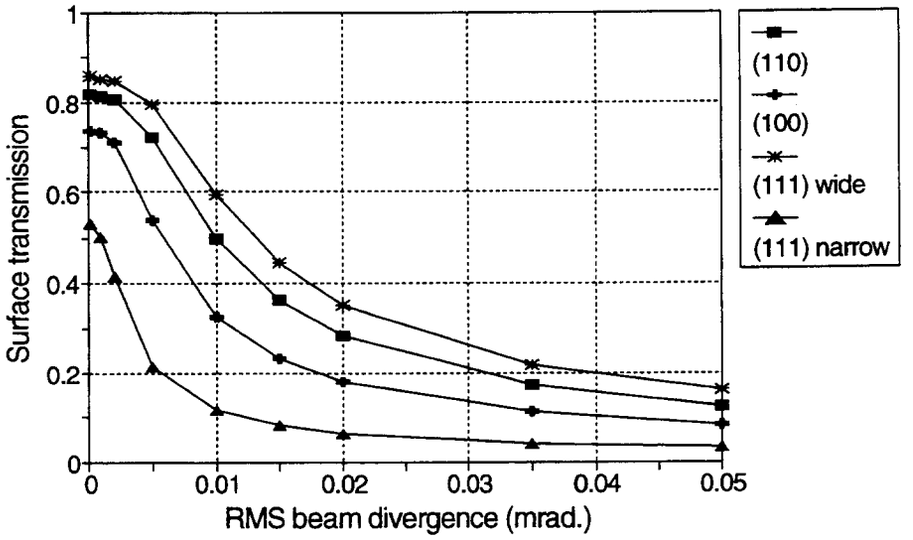


Fig. 3 Calculated surface transmission for a 450 GeV proton beam in a silicon crystal

2.4 Thermal vibrations

Up to this point we have implicitly assumed that the lattice atoms were static. In reality, they are vibrating around their mean position with a temperature-dependent vibration amplitude ρ. In silicon the one-dimensional rms thermal displacement is ρ_1=0.075 Å at 300 K, to be compared to the lattice constant a=5.431 Å. The uncertainty in position owing to these thermal vibrations gives rise to a transverse smearing, and the resulting 'thermally averaged' continuum string potential is given by

$$U_T(\vec{r}_\perp) = \int P(\vec{r}_\perp) U(\vec{r}_\perp - \vec{r}_\perp) d^2\vec{r}_\perp' \quad . \tag{8}$$

The probability distribution P of the target atoms is gaussian, that is

$$P(\vec{r}_\perp) = (\pi\rho^2)^{-1} \exp(-r_\perp^2/\rho^2) \tag{9}$$

where ρ^2 is the two-dimensional mean square thermal displacement from the string. Qualitatively, the main effect of the thermal vibrations occurs at small distances from the

string, where the infinite static potential is modified to have a finite maximum. A similar thermal smearing applies for the planar potential, but the influence of the thermal vibrations is here rather small.

2.5 Potentials

For the atomic screened Coulomb potential, Lindhard used the so-called standard potential, an analytical, easy-to-evaluate potential,

$$V(r)=\frac{Z_1Z_2e^2}{r}\left(1-\frac{r}{(r^2+C^2a^2)^{1/2}}\right) \tag{10}$$

which leads to the critical angles ψ_1 and ψ_p, Eq. (6).

The corresponding single-string axial continuum potential is obtained from Eq. (1)

$$U(r_\perp)=\frac{Z_1Z_2e^2}{d}\ln\left(1+\frac{C^2a^2}{r_\perp^2+\frac{1}{2}\rho^2}\right) \tag{11}$$

where d is the distance between the atoms in the string, and where the thermal averaging, Eq. (8), has been approximated introducing the $\frac{1}{2}\rho^2$ term.

The corresponding planar continuum potential, for a single static plane, is obtained from Eq. (2),

$$U(y)=2\pi Z_1Z_2e^2Nd_p[(y^2+C^2a^2)^{1/2}-y] \tag{12}$$

where y is the distance from the plane and d_p is the planar distance.

A rather accurate potential, for more quantitative evaluations, is the thermally averaged Doyle-Turner single-string potential given by

$$U(r_\perp)=\frac{Z_1e^2}{a_0}\frac{2a_0^2}{d}\sum_{i=1}^{4}\frac{a_i}{B_i+\rho^2}\exp\left(-\frac{r_\perp^2}{B_i+\rho^2}\right) \tag{13}$$

where $B_i=b_i/4\pi^2$, and the constants a_i and b_i are given in Table 1 for silicon and germanium. The corresponding planar thermally averaged Doyle-Turner potential takes the form

$$Y(y)=2\sqrt{\pi}Z_1\frac{e^2}{a_0}a_0^2Nd_p\sum_{i=1}^{4}\frac{a_i}{(B_i+\rho^2)^{1/2}}\exp\left(-\frac{y^2}{B_i+\rho^2}\right) \tag{14}$$

Axially channeled particles are moving in a potential which is the superposition of the single-string potentials pertaining to the individual strings in the transverse plane, see Fig. 2 b). There is, however, only a significant difference between the multi-string and the single-string potential for small transverse energies. Similarly, the planar potential is the sum of the two potentials from neighbouring planes.

Table 1
Parameters for Doyle-Turner potential

Si		Ge	
a_i [Å]	b_i [Å^2]	a_i [Å]	b_i [Å^2]
2.1293	57.7748	2.4467	55.893
2.5333	16.4756	2.7015	14.393
0.8349	2.8796	1.6157	2.4461
0.3216	0.3860	0.6009	0.3415

2.6 Dechanneling

The thermal vibrations of the crystal atoms, the discreteness of the lattice and the presence of the electrons in the target mean that the motion in the continuum potential is perturbed, and changes in the transverse energy take place. On the average the transverse energy will increase with penetration depth. This means in particular that channeled particles may become transferred to the random beam, a process called *dechanneling*. For low transverse energies, it is the scattering on the electrons which contributes, whereas for higher transverse energies it is the nuclear scattering which dominates. In general the dechanneling process can be described by a diffusion equation. For well-channeled particles, in particular planar channeled, the dechanneling process can be approximated by an exponential depletion of the number of channeled particles,

$$n = n_0 \exp(-z/L_0) \tag{15}$$

where the phenomenological dechanneling length L_0 scales with momentum. For the (110) plane in silicon, $L_0 \approx 0.9\text{m}\cdot p[\text{TeV/c}]$. The inverse process of dechanneling is also possible and random particles may become channeled. This process is called 'feed-in' or 'volume capture'.

For not too small transverse energies, the dechanneling is dominated by scattering on the nuclei. This means that there is a temperature dependence of the dechanneling length owing to the thermal vibrations, and channeling effects can be enhanced by cooling the crystal.

In Fig. 4 are shown some measurements of proton dechanneling lengths at high energy for the (110) plane in silicon at room temperature demonstrating the energy scaling.

The dechanneling lengths for negative particles are much shorter than for positive, because negative particles move in areas of high electron and nuclear density.

2.7 Crystals and Radiation Damage

To observe channeling effects and to apply them, it is clearly neccessary to have single crystals, or polycrystalline material with a mosaic spread much less than the relevant critical angles. At high energy, this leaves only silicon, germanium and more recently diamond as target materials. These materials all have diamond structure. At low energy many other crystals are available for channeling.

It is also important that the crystals are almost perfect in other respects, namely the

number of defects should be small. There exists many types of defects in crystals, originating either from displaced crystal atoms, dislocations, or impurities. It is obvious that atoms sitting in the channels between the strings or planes lead to increased dechanneling. Actually, low-energy channeling can be used to identify the lattice positions of impurities. Hence it is important, that the crystal material has a small number of impurities and defects. Radiation damage, that is damage of the crystal owing to the beam exposure, will also lead to an enhanced dechanneling. A close collision between the projectile and a lattice atom may displace the lattice atom from the string position to an interstitial position, and create a vacancy. In general, such point defects are less severe at high energy, where the cross section for scattering to large angles is small and where the number of atoms participating in one collision with a string is enormous ($\sim 1/\psi_1$).

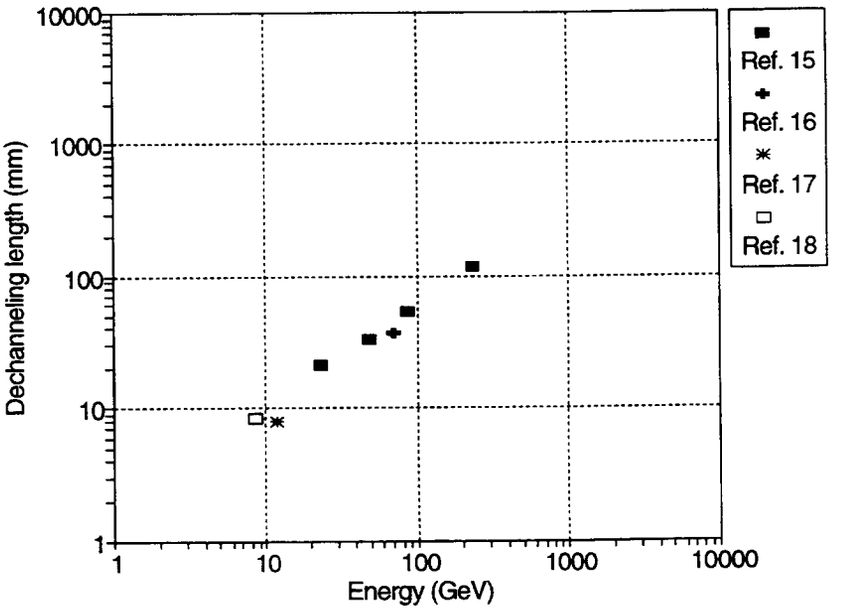


Fig. 4 Measurements of dechanneling lengths in silicon

3. THE STRONG FIELD AND IT'S ORIGIN

It is well-known from Coulomb's law

$$E(r) = \frac{Z_2 e}{r^2} \tag{16}$$

that very high electric fields exist close to point-like particles. For example, the electric field at a distance of 0.1 Å from a silicon nucleus amounts to $2 \cdot 10^{12}$ V/cm. This high field is, however, not of much use as a deflecting field, since it only extends over distances ~0.1 Å.

In the continuum approximation, we have for a <110> string of silicon atoms $E \sim 1.3 \cdot 10^{10}$ V/cm corresponding to a magnetic field (for relativistic particles) of 4000 Tesla. Furthermore, this field extends all along the length of the string, i.e. over the macroscopic distance of the crystal thickness (from mm to m). In the planar case, the corresponding field

is $1.3 \cdot 10^9$ V/cm (~ 400 Tesla), and this field applies all over the extent of the two-dimensional plane. It is these immense fields extending over macroscopic distances, which makes channeling useful for the deflection of charged particles.

For crystals of higher atomic number than silicon, the fields are evidently larger than the values given above.

4. MANIFESTATIONS OF CHANNELING

Channeled positive particles are kept away from the strings and planes of atoms, whereas negative particles are focused around the atomic strings and planes. This will obviously change the interaction between the projectiles and the crystal atoms. For example all close-encounter processes, which require a small-impact parameter collision with a target atom, have a decreased/increased probability for aligned positive/negative particles. Examples of such processes are nuclear reactions, wide-angle scattering and inner-shell excitation. For a beam of positive particles well-aligned with a string direction, such processes are decreased by two orders of magnitude. Also less local processes like multiple scattering are decreased/increased for positive/negative channeled particles.

Processes involving collisions with target electrons are also changed. For example a beam of channeled positive particles experience an energy loss around half of that for a random beam. This is illustrated by Fig. 5, where the energy loss distribution of 15 GeV protons in a 0.74 mm Ge crystal is shown for an axially aligned beam and a random beam [19].

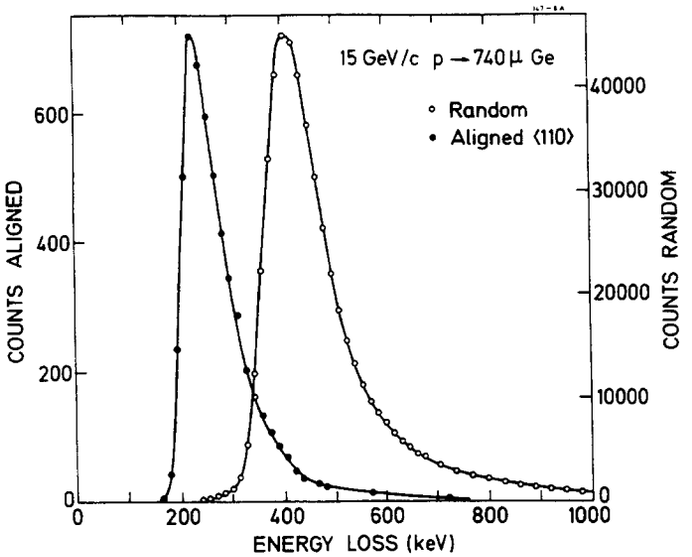


Fig. 5 Random and axially channeled energy loss of 15 GeV protons in 0.74 mm Ge

Channeled electrons/positrons have an increased emission of bremsstrahlung. The regular motion of electrons/positrons through single crystals lead to coherent emission of high-energy photons, in much the same way as undulators are used in electron storage rings. For example axially channeled 100-200 GeV electrons have a radiation energy-loss almost two-orders of magnitude larger than in a corresponding amorphous medium. Also the inverse

process to this *channeling radiation*, namely pair production by photons, is increased in aligned crystals. Crystals may hence be useful as bremsstrahlung and pair-production targets.

The channeling model, and these channeling effects, have been studied for a number of different projectiles: electrons/positrons, muons, pions, protons and heavy ions. The energies used range from keV to TeV corresponding to critical angles from degrees to microradians, and the same, relatively simple, description of particles colliding with strings or planes of atoms holds over this enormous energy range. Note, that at high energy ($v \approx c$), the channeling properties of a particle only depends on it's charge and momentum.

The study of channeling and the applications of the effect have yielded a wealth of results in both applied and basic research. In particular at low energies, standard techniques using channeling are today widely used in e.g. semiconductor and materials science. At high energy, the application of the channeling effect is less mature, but it has started as we are going to see now.

The reader is referred to the bibliography for further details and references to the large subject of channeling effects.

5. CHANNELING IN A BENT CRYSTAL

In the following we shall confine ourselves to planar channeling of positive particles.

It is hardly surprising, that also when a crystal is slightly bent elastically, planar channeled particles will still follow the bent planes. Channeling in a bent crystal with constant curvature is described in the continuum approximation by the introduction of a centrifugal force, giving rise to an effective potential

$$Y_{eff}(x) = Y(x) - \kappa pvx \tag{17}$$

where x is the distance from the centerline between the atomic planes. The curvature is given by $\kappa = 1/R$, where R is the radius of curvature. In Fig. 6 is shown the potential from an unbent crystal (full-drawn curve) and the resulting potential for a bent crystal, in the case of R = 4 m (dotted) and R = 15 m (dash-dotted) at pv = 450 GeV. It is seen, that bending will cause particle trajectories to shift towards the outer atomic plane, and that bending leads to a lowering of the potential barrier. In this way, the critical angle of a bent crystal is reduced.

As was realized by Tsyganov [11], there is a critical radius of curvature R_c for which there is no potential well, i.e. no channeling. This critical radius is reached when Y_{eff} has a minimum at $x = d_p/2$ (or rather $d_p/2$ - ρ) and we get

$$R_c = pv/\pi Z_1 Z_2 e^2 N d_p. \tag{18}$$

Only at high energies can crystals be bent elastically to this limit. For (110) planes in Si we get R_c = 0.4m·p[TeV/c]. This is a very large curvature, corresponding to 4.3 degrees or 75 mrad over 3 cm at 450 GeV!

It is practically not possible to bend a crystal all the way to the edge, and hence there is always a straight entrance, and exit, part of the crystal. This straight part is a useful position for a semiconductor detector which, as we will see later can be used to 'tag' channeled particles.

The influence of the bending is usually described by a bending dechanneling function, the fraction of charged particles lost from the channeling regime (with transverse energies

$E_\perp) Y_{eff}(d_p/2 - \rho))$ [20,21]. This dechanneling fraction F is given in Fig. 7 from ref [15]. It is only accurate for a uniformly populated transverse phase space, i.e. for a beam divergence larger than ψ_p. The abscissa coordinate is a normalized curvature and momentum variable, $\Gamma = \frac{1}{2}\kappa R_c = \frac{1}{2}R_c/R$. Furthermore the bending leads to a modification of the multiple scattering dechanneling, since the average position of the particles shifts towards the atomic planes with higher electron and nuclear densities. It can be approximated by $L_B = L_o(1-F)^2$.

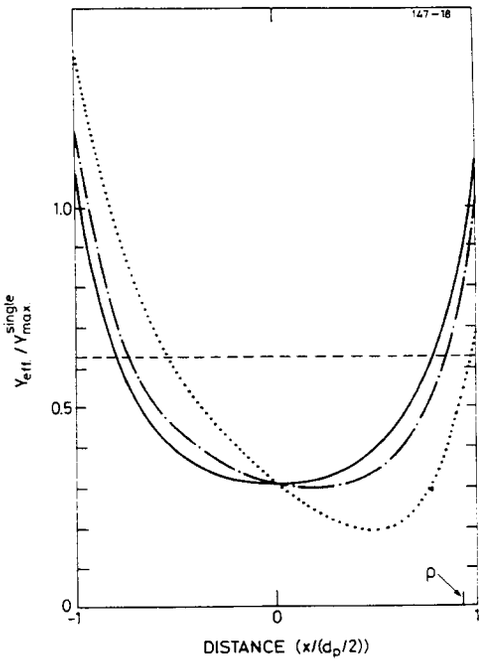


Fig. 6 Planar potential for a straight (full-drawn) and bent crystals

An accurate prediction of the bending efficiency requires a detailed simulation. A rough estimate can be made from the following prescription: 1) The surface transmission, i.e. the number of initially channeled particles, is calculated from e.g. Fig. 3 for a given beam divergence. 2) The multiple scattering dechanneling is taken into account by application of the depletion, Eq. (15). 3) The bending dechanneling is calculated from Fig. 7. The influence of the bending on the normal dechanneling is introduced. It should be noted, that this procedure is only valid for a uniform curvature. For a non-uniform curvature, the most important modification is to determine the bending dechanneling from the maximum curvature in the crystal, which as we will see later can be significantly larger than the average curvature.

6. BENDING BEAMS WITH CRYSTALS

The first observation of beam bending was made at JINR at Dubna at 8 GeV in 1978 [12], two years after the proposal. Since then, bending has been performed up to energies of 800 GeV [22]. Axial bending [17] is much less studied than planar. Planar bending of negative particles has never been investigated experimentally. All the experiments [12,13,15-18,22-30] have been made with protons, apart from [17] and [23] where π^- and C^{6+},

respectively, were used. These previous bending experiments are summarized in Table 2.

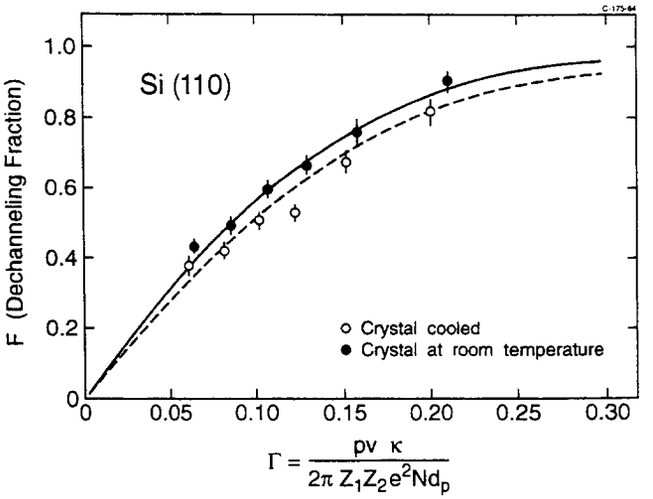


Fig. 7 Dechanneling fraction as function of normalized curvature

We shall in the present paper concentrate on the recent CERN experiments, first of all because the author has been involved in these experiments, but also because only in these recent 450 GeV experiments, large efficiencies up to 50 % have been obtained. Before discussing this experiment in detail, let us look at the, at first sight trivial, problem of bending a crystal.

Table 2
Bending experiments

Year	Energy	Laboratory	Bending details	Ref.
1979	8.4 GeV	JINR, Dubna	2 cm, 1-26 mrad	12
1980	12 GeV	CERN	3 cm, 4-52 mrad (p,π^+,π^-)	13, 17
1984	12-800 GeV	Fermilab	2-5 cm, 3-32 mrad	15, 22, 24
1982	1 GeV	Leningrad	0.1-1 cm, (also focusing)	25, 30
1987	53 GeV/c	JINR, Dubna	2 cm, 65 mrad, (C^{6+})	23
1991	70 GeV	IHEP, Protvino	7-10 cm, 20 mrad, focusing	16, 29
1989	450 GeV	CERN (H8)	5 cm, 2-12 mrad	26, 27
1989	450 GeV	CERN (H4)	5 cm, 10 mrad	28

6.1 Bending devices

The ideal beam bender is a crystal bent at constant curvature, so that the crystalline planes make a section of concentric cylinders. In practice some mechanics is needed to bend the crystal as closely as possible to this ideal curvature.

Historically, the first bending devices used were three-point benders as in Fig. 8a, where three parallel rods are pressed onto the flat sides of the crystal, to bend it. Ideally, such a bender gives a linearly increasing curvature from the first rod to the middle and a correspondingly decreasing curvature from the middle to the last rod. In this way, the maximum curvature is twice the average curvature.

The obvious improvement to the three point bender is a four-point bender, as in Fig. 8b, where ideally the curvature is linearly increasing from the first rod to the second, then constant from the second to the third and then finally linearly decreasing from the third to the fourth rod. A common problem to the three and four point bender is the local distortions at the contact points of the rods, which lead to additional dechanneling, in particular at the points of the crystal side in direct contact with the rod. This local stress is observable experimentally as additional dechanneling at the deflection angles corresponding to these points (see later).

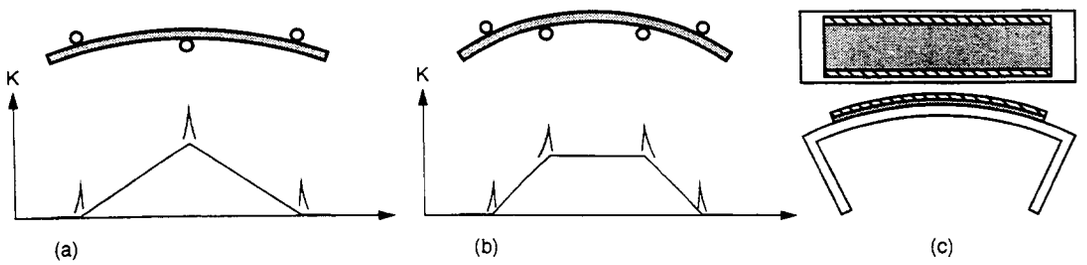


Fig. 8 Crystal benders of 3-point, 4-point and U-type. The graphs show the curvature along the crystals for the 3- and 4-point benders

An elegant way of avoiding mechanical stress and produce a uniform bend is to bend the crystal by coating one side of the crystal with a layer of a different material. In Ref. [15] a 26 mm Si crystal was bent 30 mrad by sputter coating one side of the crystal with a 10 micron layer of ZnO. Obviously, the resulting bending angle is not adjustable, and it is temperature dependent. Furthermore, the coating also produces a transverse bend.

Approximately uniform bends have also been obtained by pressing crystals against precision machined cylindrical surfaces. The crystals have been pressed either by rods at the crystal ends [28] or by a 'fluid' composed of small steel balls [26]. Also in this case, the bending angle is fixed. However, the beam bending angle can be adjusted by using the long sides of the bent crystal as entrance and exit faces [28].

Finally, the bottom of a U-shaped piece is bent uniformly by pressing the top of the two sides together, Fig. 8c. So if a crystal is clamped to such a U-shaped stainless steel piece, the crystal will be approximately bent cylindrically. Even better, the crystal itself can be machined to the shape of a U. In this case, bending in the unwanted directions, in particular variation of the planar direction with position on the entrance side of the crystal, can be avoided. Such crystals have been used in the CERN extraction experiments to be discussed in Section 7.

6.2 The 450-GeV CERN bending experiment

The recent 450-GeV experiments at CERN have been made in two beamlines in the north area of the SPS. At the H4 beam, experiments have been performed as a preparation to the use of a crystal in the new CP-violation experiment NA48 [28,31]. The other experiments

[26,27], which we will discuss in some detail, are made in the H8 microbeam, which is a unique beam for such investigations. This beam has an extremely small emittance, obtained by successive collimation of the primary beam. At the position of the crystal, the beam size is around 2×1 mm^2 (FWHM) to be compared to the crystal size of 0.9×10×50 mm^3. Furthermore, the horizontal divergence is around 3 μrad (rms) to be compared to ψ_p = 7μrad for (110) Si at 450 GeV. The beam intensity is up to 10^6 protons per burst.

Different bending devices have been used in these experiments, but we shall confine ourselves to the discussion of the results obtained with a 3-point bender. The layout of the experiment and the bending jig are shown in Fig. 9 a) and b), respectively. The bent silicon crystal is mounted on a goniometer with 1.7 μrad stepsize. Scintillators (SC) 1, 2 and 3 are used to define a beam trigger, with scintillator 4 as a veto of events creating a nuclear interaction in the crystal. Two drift chambers (DC) are used to measure the position of the beam particles just in front of the crystal and 4 meters behind. Data were accumulated with bending angles from 1 to 11 mrad.

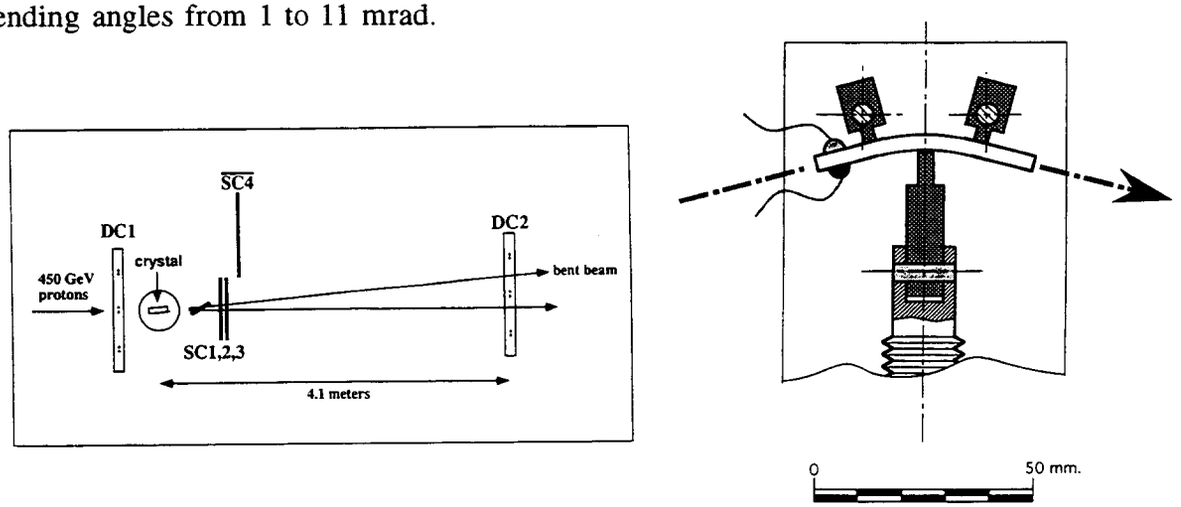


Fig. 9 a) Layout of bending experiment. b) Crystal bender

A surface-barrier semiconductor detector is produced on the straight entrance part of the crystal to tag the channeled particles (Fig. 9 b). In Fig. 10 a) and b) are shown the energy-loss spectra for a non-aligned beam and for a beam aligned with the (111) planes of the silicon crystal. The usual Landau distribution is seen in Fig. 10 a) (with some background) and for an aligned beam in Fig. 10 b) we observe the low-energy loss channeled particles. From Fig. 10 b) we deduce immediately a channeled fraction of the beam upon entrance in excess of 50 %. In. Fig. 10 b) is also plotted the energy loss of those particles deflected 8.9 mrad by the crystal. A narrow energy-loss distribution is observed without a random component, meaning that the bent particles are all channeled from the entrance of the crystal, i.e. no feed-in (Section 2.6).

The crystal acceptance can be measured by monitoring the channeled particles, either by their low-energy loss or by bending, as a function of crystal angle. In Fig. 11 a) is shown the energy-loss window used to tag the channeled particles, and in Fig. 11 b) the measured crystal acceptance. The steepness of the distribution demonstrates that the beam indeed has a very small divergence. The FWHM of the distribution is 17 μrad to be compared to twice the critical angle, in expression 14 μrad.

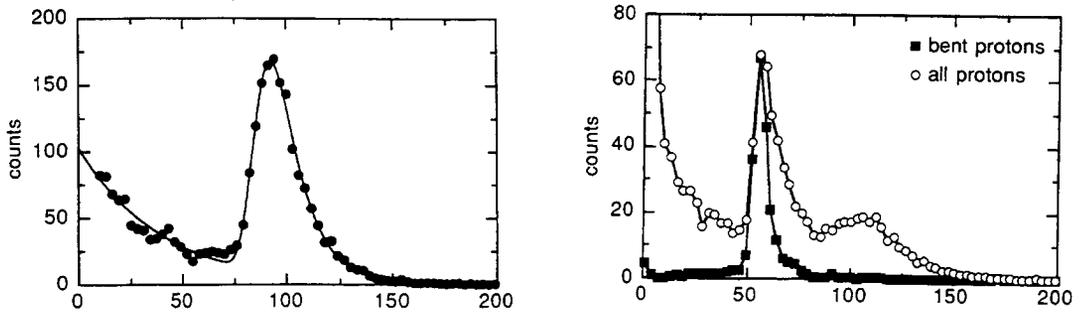


Fig. 10 Energy-loss distributions for a) random and b) aligned crystal

The distributions of the deflected and straight beam, as measured by the second drift chamber are shown in Fig. 12, for crystal bendings of 2.4, 4.6 and 8.1 mrad. The left-most peak is the undeflected beam, and the right-most peak is the bent beam. Between these peaks are the dechanneled particles. The tail from the unbent beam to the middle is the particles dechanneled because of the increasing curvature, whereas from the middle to the bent beam there are very few particles, namely only those dechanneled because of multiple scattering. We also observe, in particular for the 4.6 mrad bending angle, the excess dechanneling caused by the local stress at the middle pin.

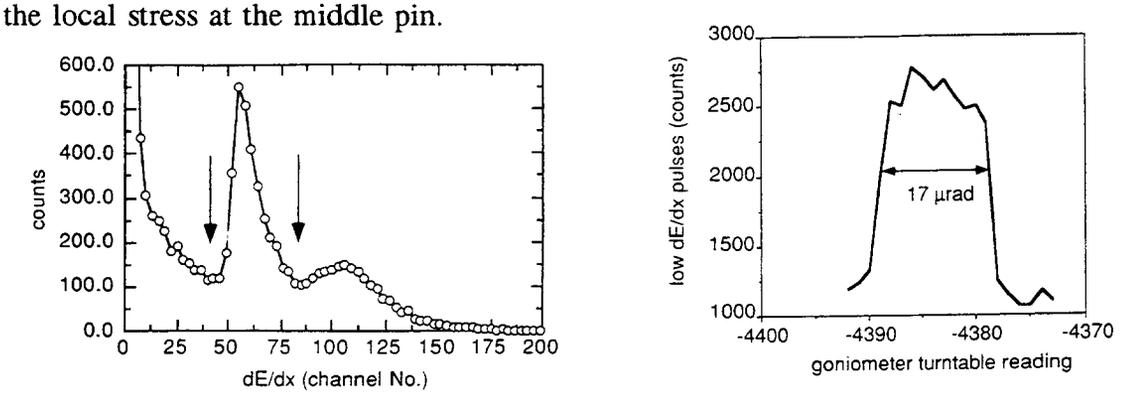


Fig. 11 a) Low-energy loss window. b) Crystal acceptance

The efficiency can be deduced by measuring the number of bent particles relative to the number of incident particles from data as in Fig. 12. This has been done for 9 different bending angles, and the resulting efficiencies as function of bending angle are shown on Fig. 13. The efficiency varies from almost 50 % at 1.4 mrad to 1 % at 11 mrad. The measurements are compared to calculated efficiencies using two extreme models for the curvature. The actual curvature of the crystal bent in the three-point bender is not known, but it is expected to be in between the ideal three-point bender and a uniform-curvature bender. Remember here, that it is the maximum curvature that determines the bending dechanneling. There is a fair agreement between the measurements and the calculations, as the measurements fall in between the two extremes.

6.3 Focusing with crystals

Bent crystals can also be used for focusing of beams. Focusing can be performed in different ways, e.g. by shaping the exit face of the bent crystal [29] or by using a crystal bent

perpendicular to the beam direction [30]. Very short focal lengths (~ meter) and very small beam sizes can be obtained, even at high energies.

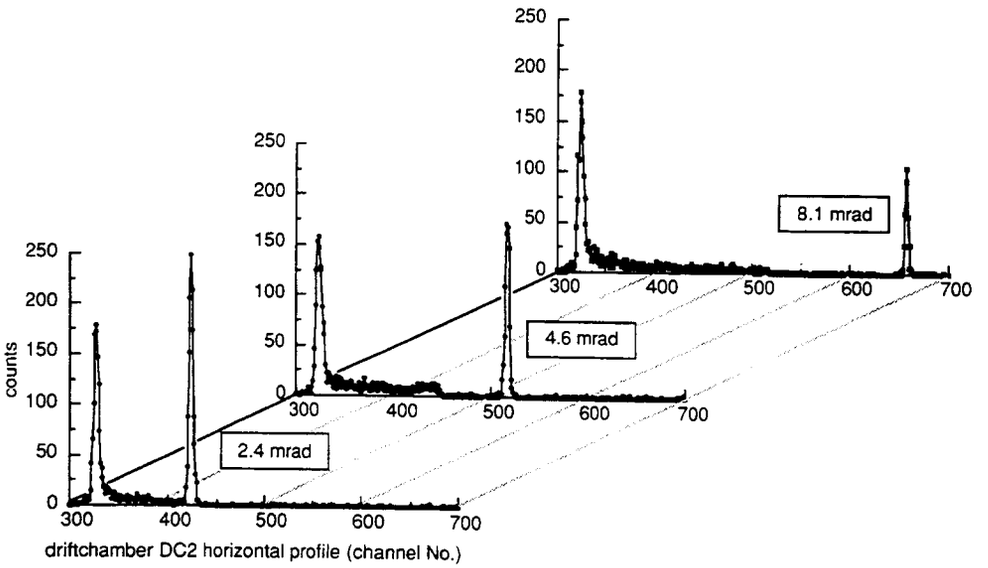


Fig. 12 Profiles of the deflected beams 4 meters downstream of the crystal

7. EXTRACTING BEAMS WITH CRYSTALS

Since a crystal can, to some extent, replace a magnet in a beam line, it can also be used as a septum magnet in a synchrotron/storage ring. The advantage of using a crystal to replace a conventional septum magnet is not only the large deflection power, but maybe more the inherently small septum thickness. The effective septum thickness of a crystal septum is ideally extremely small, namely the planar distance. In practice it will be appreciably larger, determined by surface roughness and miscuts.

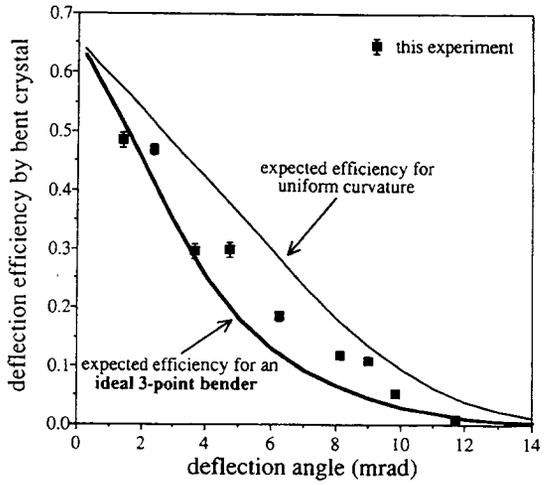


Fig. 13 Measured and calculated deflection efficiencies

Several extraction experiments have already been made, and some details are given in

Table 3. Again, we shall mainly confine ourselves to the CERN experiment for the reasons given above.

Table 3
Extraction experiments

Year	Energy	Laboratory	Crystal length, Bending angle, Efficiency	Ref.
1983	4.2-7.5 GeV	JINR, Dubna	1.1 cm, 35 mrad, ~ 10^{-4}	32
1990-	50-70 GeV	IHEP, Protvino	7 cm, 80 mrad, < 0.7 %	33
1992-	120 GeV	SPS, CERN (LHC)	3 cm, 8.5 mrad, 10 %	34
1994-	900 GeV	Tevatron, Fermilab	1.1 cm, 0.6 mrad	35

The attractive feature of using a crystal to extract a beam from a circular accelerator is besides the inherently small septum thickness also the fact that it works just as well, and in some respects better, the higher the beam energy owing to the decreasing dechanneling length with energy.

Extraction channels in accelerators typically consist of a system of electrostatic and magnetic septum magnets. The first element is usually an electrostatic septum consisting of a long array of thin wires. As an example, we take the CERN SPS [36], where the electrostatic septum consists of three 5 m long arrays of < 0.1 mm thick wires used to produce a field of 120 kV/cm. This septum gives a deflection of 0.4 mrad at 450 GeV. Obviously, for the 7 TeV LHC beam, the length of a conventional wire septum would be enormous, and there would be severe alignment problems. On the other hand, a bending angle of a fraction of a mrad for a silicon crystal for the same 7 TeV beam can easily be envisaged.

A crystal can be used as a septum magnet in at least two different ways. In Fig. 14 are shown schematically two scenarios of 'scraping' the beam with two different edges of the crystal. In both cases, the effective septum thickness will be given by the surface roughness, i.e. it is important that the surfaces are polished (without damage) to optical quality. Furthermore it is essential in (a), that the angle between the extracting surface and the planar direction is very small.

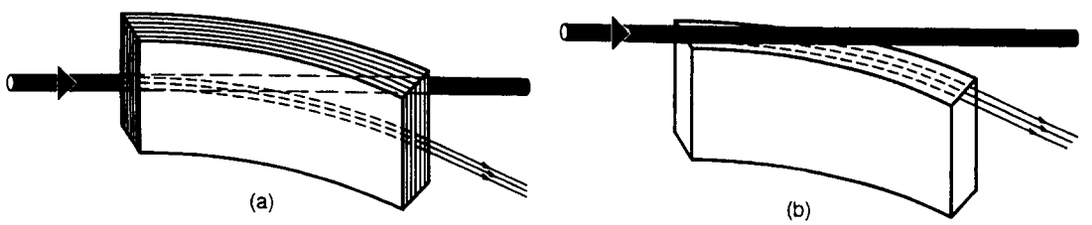


Fig. 14 Scenarios for crystal extraction

The lay-out of crystal extraction is shown in more detail in Fig. 15. Here the importance of surface roughness and miscut angle is evident. It is furthermore seen, that the

impact parameters of the particles incident on the crystal should be large compared to the surface roughness, and that the impact angles should be less than the critical angles.

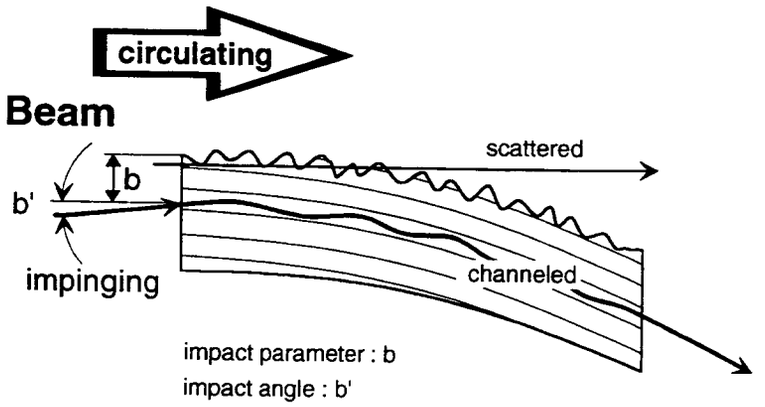


Fig. 15 Impact of beam particles on crystal

The extraction efficiency can be calculated as outlined in section 5. Calculated efficiencies are shown in Fig. 16 as function of pv/R for four bending angles. For a given bending angle, there is a maximum in the efficiency, since the bending dechanneling is increasing for decreasing R, and the multiple-scattering dechanneling increasing for increasing R, i.e. longer crystals. The maximum efficiency for an extraction angle of 8 mrad is 42 % for pv/R=0.4. This gives for the SPS-energy of 120 GeV an optimal crystal length of 2.4 cm. For an extraction angle of 0.2 mrad (relevant for the LHC), the maximum efficiency is 74 % at pv/R=0.07, corresponding to a crystal length of 20 cm at 7 TeV.

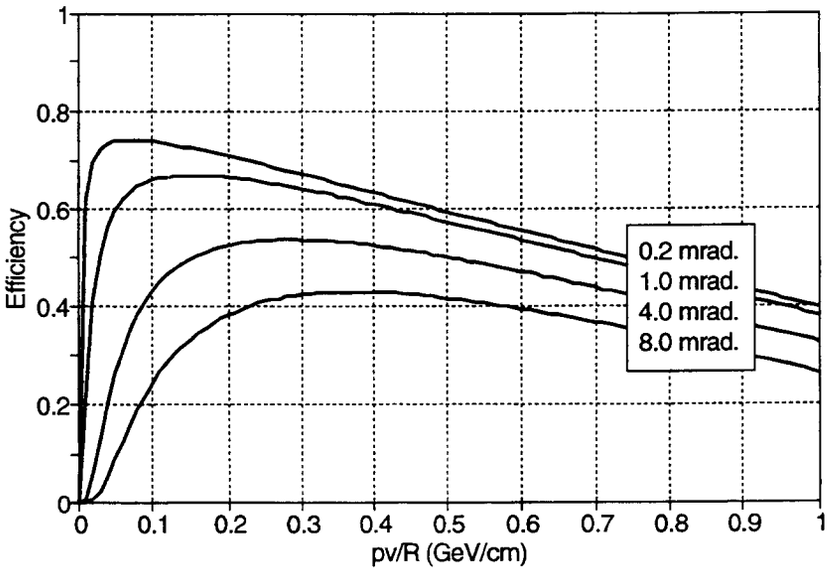


Fig. 16 Calculated bending/extraction efficiencies for Si (110)

7.1 The CERN extraction experiment

The motivation for the extraction experiment at the CERN SPS was the possibility to extract parasitically a proton beam from the LHC for fixed target physics (CP-violation in the b-quark system) [37]. The Large Hadron Collider is a planned superconducting 7+7TeV collider to be built in the LEP tunnel at CERN. An extracted beam of intensity 10^8/sec. would provide 10^{10} BB-bar pairs per year, two orders of magnitude more than from an e^+e^- B-factory with the very large luminosity $\mathfrak{L}=10^{34}cm^{-2}s^{-1}$. At the nominal luminosity in the LHC, around 10^9 protons per second are expected to diffuse from the central core of the LHC beam (10^{14} protons) into the halo, where they will have to be scraped by collimators to avoid hitting the superconducting magnets. Part of this beam could possibly be extracted by a crystal into a fixed-target experiment.

In the extraction experiment, the SPS is operated at 120 GeV where a coasting beam of $5 \cdot 10^{11}$ protons has a lifetime in excess of one hundred hours. The normalised horizontal and vertical emittances are 6-8 mm·mrad for the unperturbed beam. Diffusion is created by exciting the beam horizontally with band-limited white noise. The crystal is placed 10 mm from the closed orbit and the horizontal beam emittance is around 60 mm·mrad when the large-amplitude particles reach the crystal. The typical impact parameters of the particles on the crystal are of the order of one micron. The layout of the SPS for crystal extraction is shown in Fig. 17.

In the SPS experiment [34], 3 cm long (110) Si crystals, mounted in a U-shaped bending device, are used to extract the beam, like in Fig. 14 a). The benders are mounted on a goniometer to align the planar direction with the beam. The benders can furthermore be translated towards the circulating beam. The extraction angle is 8.5 mrad, and the crystals are prealigned and monitored by laserbeams reflected from the polished surface of the crystals. The position of the essential elements in the extraction straight section are shown in Fig. 18, together with the crystal bending device. Several particle detectors allow observation of the extracted beam. We mention scintillators (S), scintillator hodoscopes (H), scintillating screens observed with a camera (TV) and position sensitive microstrip gas chambers (MSGC).

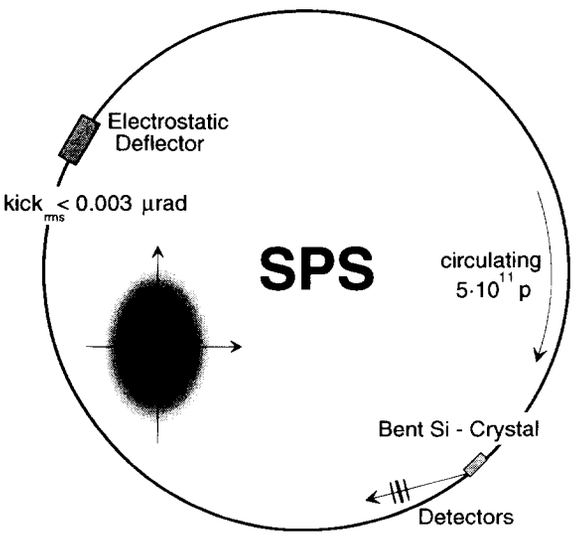


Fig. 17 Layout of SPS for crystal extraction

For prealigned crystals, extraction is easily obtained with efficiencies of up to 10%. This efficiency is determined from intensity monitors in the extracted beam and the beam lifetime in a steady-state diffusion mode. As discussed above, efficiencies up to 40% are theoretically possible.

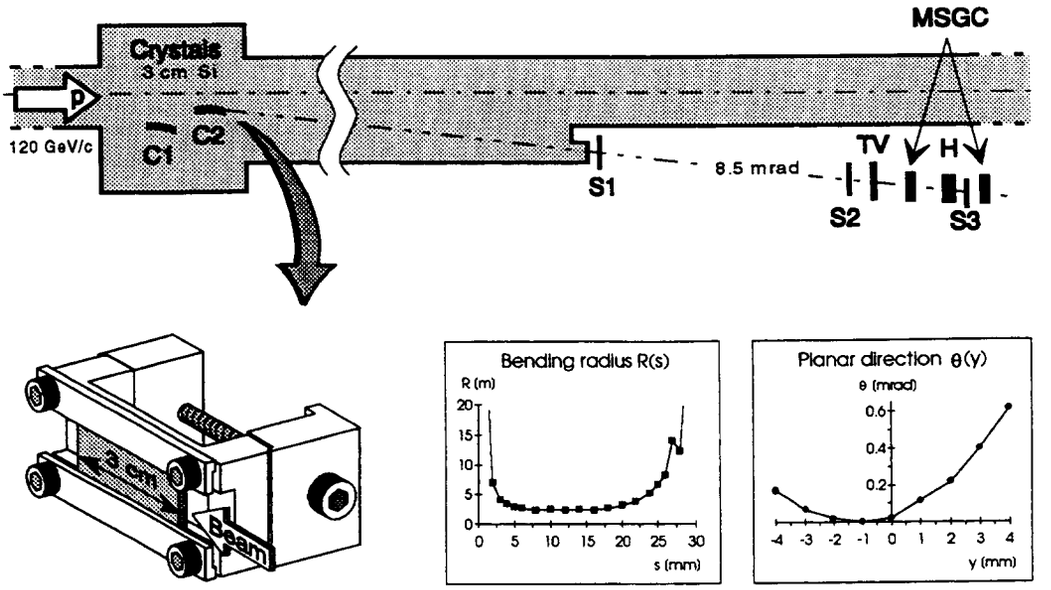


Fig. 18 Layout of straight section used for crystal extraction in SPS

The angular acceptance of the crystal has been measured by monitoring the extracted intensity as a function of planar angle to the beam, Fig. 19. The width of the distribution is much larger than expected, namely around 200 μrad to be compared to $2\psi_p$=28 μrad. Furthermore the observed beam profiles are for some orientations of the crystal wide and show double peaks. These observations can to some extent be explained by the so-called 'anticlastic' bending of the crystal in the bending device used. As seen on the inserts in Fig. 18, the crystal is also bent perpendicular to the direction of the bend. Furthermore, and more seriously, the planar direction varies with the vertical coordinate at the entrance of the crystal. Clearly a beam with a vertical size of more than 1 mm will produce an angular peak larger than $2\psi_p$ with a decreased extraction efficiency. Furthermore, multi-pass extraction may be involved in the explanation. A particle not channeled or dechanneled may stay in the machine even when it is multiple scattered from the 3 cm Si. The particle may then be extracted on subsequent passes of the crystal. Such multi-pass extraction has been discussed in [38,39].

The SPS extraction experiments are continuing in order to understand fully the observations. One improvement is the installation of a bent crystal without anticlastic bending, namely a crystal cut in U-shape. This crystal is shown in Fig. 20.

Furthermore it is very important to understand the influence of the crystal on the machine itself, in particular what the crystal will do to the collimation of the machine. (A silicon crystal 3 cm. long represents 6 % of a nuclear interaction length.) This is important for the background in collider detectors, and also for the quench protection of the superconducting magnets.

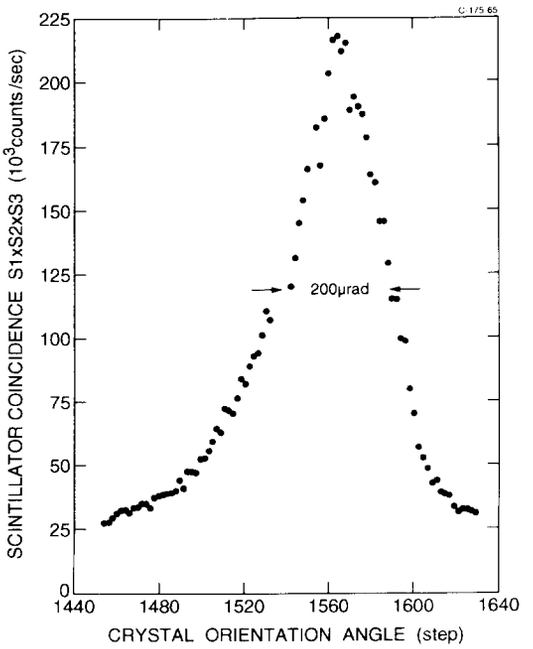


Fig. 19 Measured angular acceptance of extracting crystal

Also the point of radiation damage has to be investigated in detail, in particular because the exposed area of an extracting crystal will be small owing to the small impact parameters. At fluences of 10^{19} protons/cm^2 no noticeable efficiency deterioration at 70 GeV has been seen at Serpukhov [40], whereas a deterioration of low-energy channeling has been observed by Baker et al. [41] for a crystal exposed to $4{\cdot}10^{20}$ protons/cm^2 at 28 GeV. It is, however, not evident how to relate this damage observed at low energy to high-energy dechanneling.

Finally, it is important to study crystal extraction at as high an energy as possible, in order to make a safe extrapolation to LHC energies. At the Tevatron, extraction tests are being prepared for 0.9 TeV [35].

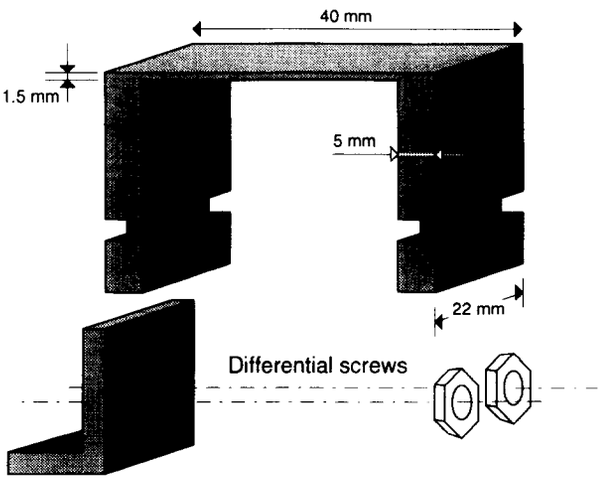


Fig. 20 Size of U-shaped crystal

8. CONCLUSIONS AND OUTLOOK

Channeling is today a well-founded effect occurring when charged particles are incident on single crystals in directions close to crystallographic directions.

The coherent scatterings on the strings or planes of atoms, which a crystal consists of, changes dramatically the transmission of such channeled particles, in comparison to amorphous materials.

In particular, the strong steering effect can be applied in bending and focusing of beams replacing magnets with crystals. The bending powers are very large and the inherent losses acceptable. Efficiencies up to 50 % have been demonstrated.

Crystals are also potentially useful as extraction devices, in particular because of the inherently small septum thickness. Extraction with efficiencies up to 10 % have been demonstrated.

The first real applications of crystals as beamline elements have been made, and the potential for other applications of the immense crystalline fields seems large. Especially applications at very high energies are promising because of the increasing dechanneling length and the slowly decreasing critical angles with energy.

The reader is encouraged to 'invent' his/her own applications of crystalline fields!

ACKNOWLEDGEMENTS

The present lecture was based on the efforts of many people over many years. In particular the author acknowledges the help of A. Baurichter, K. Elsener, M. Gyr, A.H. Sørensen and E. Uggerhøj in the preparation of this Chapter to the CAS proceedings.

REFERENCES

[1] G.R. Piercy, F. Brown, J.A. Davies and M. McCargo, Phys. Rev. Lett. 10 (1963) 399.

[2] R.S. Nelson and M.W. Thomson, Phil. Mag. 8 (1963) 1677.

[3] M.T. Robinson and O.S. Oen, Appl. Phys. Lett. 2 (1963) 3032; Phys. Rev. 132 (1963) 2385.

[4] J. Lindhard, Kgl. Danske Videnskab. Selskab, Mat. Fys. Medd. 34 (1965) No.14; Phys. Lett. 12 (1964) 126.

[5] J. Stark, Physik. Zeitsch. 13 (1912) 973.

[6] O. Fich et al., Phys. Lett. 57B (1975) 90.

[7] A.A. Vorobiev, V.V. Kaplin and S.A. Vorobiev, Nucl. Instrum. Methods 127 (1975) 265.

[8] M.A. Kumakhov, Phys. Lett. 57 (1976) 17.

[9] R.L. Swent et al., Phys. Rev. Lett. 43 (1979) 1723.

[10] J.U. Andersen and E. Lægsgård, Phys. Rev. Lett. 44 (1980) 1079.

[11] E.N. Tsyganov, Fermilab TM-682, TM-684, Batavia, (1976) (unpublished).

[12] A.F. Elishev et al., Phys. Lett. 88B (1979) 387.

[13] J. Bak et al., Phys. Lett. 93B (1980) 505.

[14] N. Bohr, Kgl. Videnskab. Selskab, Mat. Fys. Medd. 18 (1948), nr. 8.

[15] J.S. Forster et al., Nucl. Phys. B318 (1989) 301, and R.A. Carrigan in Relativistic Channeling, R.A. Carrigan and J.A. Ellison (eds.), (NATO ASI series Vol. B 165, 1987).

[16] V.M. Biryukov et al., submitted to Nucl. Instrum. Methods in Phys. Research.

[17] J.F. Bak et al., Nucl. Phys. B242 (1984) 1.

[18] C.R. Sun et al., Nucl. Instr. and Meth. B2 (1984) 60.

[19] H. Esbensen et al., Phys. Rev. B18 (1978) 1039.

[20] H. Kudo, Nucl. Instrum. Meth. 189 (1981) 609.

[21] J.A. Ellison, Nucl. Phys. B206 (1982) 205.

[22] S.I. Baker, Nucl. Instrum. Methods in Phys. Res. A248 (1986) 301.

[23] L.I. Bel'zer et al., JETP Lett. 46 (1987) 381.

[24] S. I. Baker et al., Phys. Lett. 137B (1984) 129.

[25] V. A. Andreev et al., JETP. Lett. 36 (1982) 415.

[26] S. P. Møller et al., Phys. Lett. B256 (1991) 91.

[27] S.P. Møller et al., submitted to Nucl. Instrum. Methods in Phys. Res. B.

[28] M. Clement et al., CERN/SL/92-21, 1992 (unpublished).

[29] A. S. Denisov et al., Nucl. Instrum. Methods in Phys. Res. B69 (1992) 382.

[30] V. A. Andreev et al., JETP. Lett. 41 (1985) 408.

[31] G.D. Barr et al., CERN/SPSC/90-22, SPSC/P253 (experiment NA48), 1990 (unpublished).

[32] V.V. Avdeichikov et al., JINR Communication 1-84, Dubna (1984) [English translation: Fermilab 80/45 (1980)], unpublished.

[33] A.A. Asseev et al., Nucl. Instrum. Methods in Phys. Res., A334 (1993) 283, and references therein.

[34] H. Akbari et al., Phys. Lett. B313 (1993) 491.

[35] R.A. Carrigan et al., Fermilab proposal P853, 1991 (unpublished).

[36] M. Gyr, private communication.

[37] RD 22 Collab., B.N. Jensen et al., CERN/DRDC 91-25, DRDC/P29 (unpublished), and K. Kirsebom et al., CERN/LHCC/93-45, LHCC/I5 (unpublished).

[38] V.M. Biruykov, Nucl. Instrum. Methods B53 (1991) 202.

[39] A.M. Taratin et al., Nucl. Instrum. Methods B58 (1991) 103.

[40] Yu.A. Chesnokov et al., Proc. of the XV Int. Conf. on High Energy Accelerators (Hamburg, 1992).

[41] S. Baker et al., SSCL-Preprint-477, 1993 (unpublished).

BIBLIOGRAPHY

D.S. Gemmell, Rev. Mod. Phys 46 (1974) 129.

D.V. Morgan, Channeling (J. Wiley and Sons, London, 1973).

L.C. Feldman, J.W. Mayer and S.T. Picraux, Materials Analysis by Ion Channeling (Academic Press, London, 1982).

Y. Ohtsuki, Charged Beam Interaction with Solids, (Taylor and Francis, Ltd., 1983).

R.A. Carrigan and J.A. Ellison (eds.), Relativistic Channeling, (NATO ASI series Vol. B 165, (1987).

A.H. Sørensen and E. Uggerhøj, Nuclear Science Applications 3 (1989) 147.

ENERGY MEASUREMENT BY RESONANT DEPOLARIZATION

A. C. Melissinos
Dept. of Physics & Astronomy, University of Rochester, Rochester, NY USA

Abstract

Resonant depolarization can be used to precisely measure the average beam energy of electron-positron storage rings. We discuss how the polarization of the beam is measured by Compton scattering of laser light. Results of transverse polarization measurements at LEP, and of longitudinal polarization at SLAC are presented. Calibration of the LEP energy by this technique has a precision of order 10^{-5}; fluctuations observed at the level of $\pm 5 \times 10^{-5}$ are attributed to the distortion of the LEP lattice by tidal effects. We conclude with a brief mention of the connection between transverse polarization and Hawking-Unruh-Davies radiation.

1 POLARIZATION IN STORAGE RINGS AND SPIN PRECESSION

It is known that in storage rings electrons or positrons tend to become transversely polarized. This effect first proposed by Ternov, Loskutov and Korovina [1] is due to spin flip during the emission of synchrotron radiation favoring the spin orientation of lowest energy in the external field. In a seminal paper Sokolov and Ternov [2] showed that the polarization of the beam is given by

$$P_e = P_o(t - e^{-t/\tau_0}) \tag{1}$$

where the maximum (assymptotic) polarization is

$$P_o = \frac{8}{5\sqrt{3}} = 0.9238 \tag{2}$$

and the polarization time (constant) is

$$\tau_0 = \frac{8}{5\sqrt{3}}\frac{1}{\alpha}\left(\frac{mc^2}{\hbar c}\right)^2 \frac{\rho^3}{c\gamma^5}\left(\frac{R}{\rho}\right) \tag{3}$$

In Eq.(3) ρ is the bending radius and R the circumference radius of the ring, $\gamma = E_e/mc^2$ and $\alpha = 1/137$ is the fine structure constant. The probability for spin-flip is relatively small so that τ_0 is long. For instance for LEP $\tau_0 \sim 340$ minutes.

The observed polarization is in general less than P_o due to depolarizating effects. These arise because of the precession of the spin about non-vertical components of the magnetic field and are discussed in detail in Dr. Koutchouk's lecture [3]. The precession of the spin of a relativistic particle was given by Bargman, Michel and Telegdi [4].

$$\frac{d\vec{s}}{dt} = -\frac{e}{\gamma m}\left[\left(1 + \gamma\frac{(g-2)}{2}\right)\vec{B}_\perp + \frac{g}{2}\vec{B}_\parallel\right] \times \vec{S} \tag{4}$$

Thus in an ideal storage ring ($B_{||} = 0$) the spin precession frequency ω_s is related to the rotation frequency $\omega_c = eB/\gamma m$ by

$$\omega_s = \omega_c \left[1 + \gamma \frac{(g-2)}{2}\right] \tag{5}$$

the anomalous part of the electron's magnetic moment is known to great accuracy [5]

$$(g-2)/2 = 1.1596521884(43) \times 10^{-3} \tag{6}$$

We note that at LEP where $\gamma \sim 10^5$ the spin precesses much faster than the momentum vector, making approximately a hundred turns in one revolution around the storage ring.

Baier and Orlov [6] pointed out that even a small radial magnetic field (see Fig. 1) can lead to depolarization if the ratio of the spin precession frequency to the rotation frequency ω_s/ω_c is an integer. It is customary to define the spin tune

$$\nu_s = \gamma \frac{(g-2)}{2} = \frac{\omega_s}{\omega_c} - 1 \tag{7}$$

For LEP at the Z° mass (E = 45.6 GeV) $\gamma = 8.924 \times 10^4$ and $\nu_s = 103.48$. Evidently if one can measure ν_s one can obtain the energy through Eq.(7).

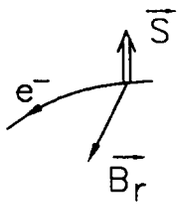


Figure 1. Sketch of radial field component and of a transverse spin orientation.

Depolarization will occur even if ν_s is not an integer, provided we add a weak depolarizing field at frequency ω_D, so as to satisfy the resonance condition

$$n_s\nu_s \pm n_xQ_x \pm n_yQ_y \pm n_zQ_z \pm (\omega_D/\omega_c) = n \tag{8}$$

Here Q_x, Q_y are the betatron tunes and Q_z the synchrotron tune; the n's are arbitrary integers. It is desirable to work in the lowest mode namely $n_s = 1, n_x = n_y = n_z = 0$ so that

$$\nu_s \pm (\omega_D/\omega_c) = n \tag{9}$$

For instance referring to the previous example at LEP where $\nu_s = 103.48$, we have

$$f_D = f_c\delta\nu_s \sim 5 \text{ kHz} \tag{10}$$

As will be seen later, the depolarizing resonance is narrow so that f_D can be measured to good accuracy. In view of the large value of $\nu_s \sim 100$, the accuracy in ν_s and thus in the machine energy is 100 times better than the accuracy in f_D.

2 COMPTON POLARIMETRY

Polarization of electron/positron beams in storage rings was first reported by LeDuff *et al* [7] who measured the spin dependence of Touschek scattering; this is intrabeam electron-electron scattering. Other groups used the asymmetry in $e^+e^- \to \mu^+\mu^-$ [8] to deduce the polarization of SPEAR. Another method for measuring the polarization of high energy electron beams is Möller scattering from thin foils of magnetized iron. However the most generally used technique is Compton scattering of laser photons from the stored electron or positron beam [9,10,11]; first applied at SPEAR in 1979 [12].

Figure 2 shows the kinematics of photon ($\omega_0 \sim 2.3$ eV) scattering from a high energy electron ($E = \gamma/m$). In the laboratory frame the photon is backscattered, within a cone of angle $\theta \sim 1/\gamma$ as a high energy gamma with energy of order $\omega \sim 4\gamma^2\omega_0$. It is convenient to view the scattering in the rest-frame of the electron where the photon energy is $\bar{\omega}_0 = 2\gamma\omega_0$. Thus the energy of the scattered photon is given by the Compton condition

$$\frac{1}{\bar{k}} - \frac{1}{\bar{k}_0} = \frac{1}{m}(1 - \cos\bar{\theta}) \tag{11}$$

where we used the wave-vector $\bar{\mathbf{k}}_0$, $\bar{\mathbf{k}}$ of the photon in the electron rest-frame before and after the scattering; $\bar{\theta}, \phi$ are the polar and azimuthal angles of the scattered photon (with respect to its original direction). In the laboratory frame the energy of the recoil electron is given by

$$\omega = \frac{4\gamma^2\omega_0 \sin^2(\bar{\theta}/2)}{1 + (2\gamma\omega_0/m)(1 - \cos\bar{\theta})} \tag{12}$$

$$= \frac{4\gamma^2\omega_0}{1 + 2\gamma^2(1 - \cos\theta) + (2\gamma\omega_0/m)(1 + \cos\theta)} \tag{13}$$

where Eq.(12) is in terms of the scattering angle $\bar{\theta}$ in the electron rest frame, whereas in Eq. (13), ω is expressed by the laboratory (back) scattering angle θ measured from the electron's direction. Thus $\bar{\theta} = \pi$ corresponds to $\theta = 0$!. Eqs.(12,13) are written in the limit $\beta \to 1$. For LEP or SLAC at the Z° mass where $\gamma = 9 \times 10^4$ using $\omega_0 = 2.3eV$ ($\lambda = 537$ nm) gives $\omega_{\max} = 28.4$ GeV.

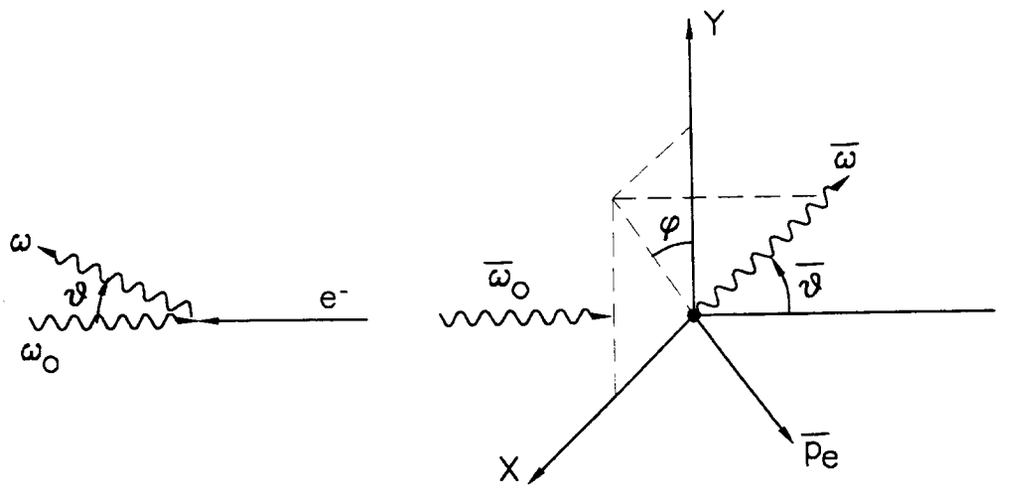


Figure 2. Photon-electron scattering (a) in the laboratory frame, (b) in the electron rest frame.

The spin-dependent Compton cross section for polarized incident light can be found in different forms. We follow Lipps and Tolhoek [13] and give the differential cross-section in the electron rest-frame.

$$\frac{d\sigma}{d\bar{\Omega}} = \frac{R_0^2}{2}\left(\frac{\bar{\omega}}{\bar{\omega}_0}\right)^2\left[(1+\cos^2\bar{\theta}) + (\bar{\omega}_0-\bar{\omega})\frac{1}{m}(1-\cos\bar{\theta})\right.$$
$$\left.+(\xi_1\cos 2\phi + \xi_2 \sin 2\phi)\sin^2\bar{\theta} - \xi_3(1-\cos\bar{\theta})\frac{1}{m}\mathbf{S}\cdot(\bar{\mathbf{k}}_0\cos\bar{\theta}+\bar{\mathbf{k}})\right] \quad (14)$$

Here $r_0 = e^2/mc^2 = 2.82\times 10^{-13}$ cm is the classical electron radius and $\mathbf{S}$ is the electron spin. ξ_1, ξ_2, ξ_3 characterize the polarization of the incident photons and are analogous to the Stokes parameters.

$\xi_1 = \pm 1$	Linear polarization along the Y, X axis
$\xi_2 = \pm 1$	Linear polarization at 45° to the Y, X axis
$\xi_3 = \pm 1$	Circular right, left, polarization

Note for instance that if ξ_1 or ξ_2 are different from zero, an azimouthal asymmetry results even in the absence of electron polarization.

Measurements of the transverse polarization at LEP have been reported [14,15]. In this case $\mathbf{S}\cdot\bar{\mathbf{k}}_0$ is zero and since circularly polarized light is used ($\xi_1 = \xi_2 = 0$) the spin-dependent part of the cross-section is

$$\xi_3 P_e \frac{\bar{k}}{m}(1-\cos\bar{\theta})\sin\bar{\theta}\cos\phi \quad (15)$$

where P_e is the magnitude of the electron polarization along the Y axis (see Fig. (2b)); $P_e = (n_+ - n_-)/(n_+ + n_-)$. When $\cos\phi > 0$ the scattering is "up", when $\cos\phi < 0$ "down" with respect to the plane of the ring as sketched in Fig. 3. Thus it is convenient to use a (strip) detector which measures the scattered photons as a function of the y-coordinate (integrating over x). At LEP the detector is placed at a distance of 275 m from the laser interaction point so that the spin independent cross-section is of the form shown in Fig. 4a. The spin dependent part is shown in (b) of the figure, for 100% electron and photon polarization. As ξ_3 is alternated in sign the centroid of the scattered photons is shifted in y by an amount $\langle y\rangle = \kappa\xi_3 P_e$; for the LEP polarimeter the analyzing power $\kappa \simeq 500\mu$m. For $P_e = 0.1$ a shift of 50 μm must be measured; this requires that a flat orbit be established in the machine.

Results from the LEP measurements [15] are shown in Fig. 5. In (a) of the figure is shown the vertical distribution of the back-scattered photons and in (b) of the figure, the asymmetry as a function of y,

$$A(y) = \frac{n_+(y) - n_-(y)}{n_+(y) + n_-(y)} \quad (16)$$

where $n_\pm(y)$ is the number of counts at a given y for $\xi_3 = \pm 1$. The data correspond to a polarization $P_e \sim 0.1$.

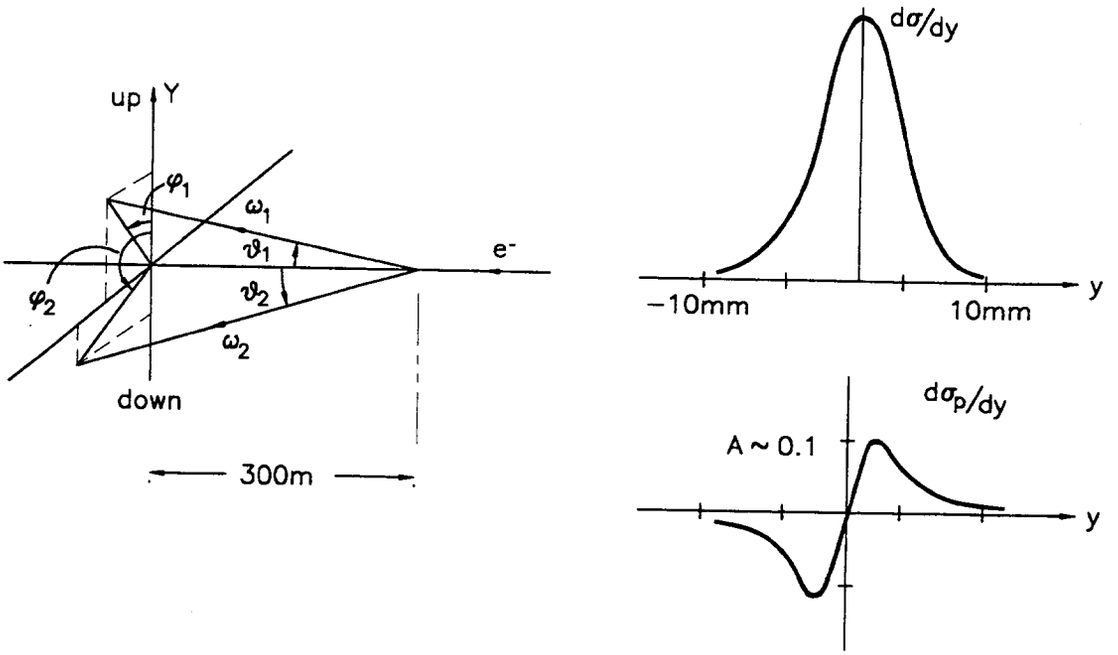


Figure 3. Geometrical arrangement for the detection of backscattered photons. Figure 4. The differential cross-section for the backscattered photons as a function of vertical distance from the median plane. (a) The polarization independent term. (b) The asymmetry for 100% transverse electron polarization and 100% photon circular polarization.

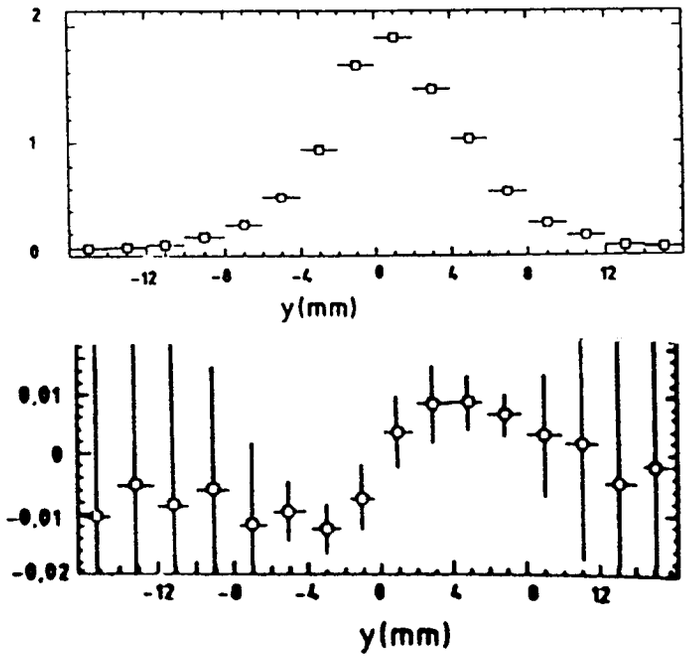


Figure 5. Results from the LEP polarization measurements [Ref. (15)]. (a) Vertical distribution at the detector. (b) Asymmetry as a function of vertical distance.

We can estimate the statistical accuracy of the measurement from the expected counting rate. Typically a NdYAG laser is used which operates in the infrared ($\lambda = 1054$ nm) and is frequency doubled to the the green ($\lambda = 532$). The laser is Q-switched to deliver 50 mJ in a $\Delta t_\omega = 12$ ns long pulse at a repetition rate of 30 Hz; the focal spot has a diameter $d_\omega = 2$ mm and the crossing angle is $\alpha = 3°$. We take the electron beam parameters as $\sigma_h = \sigma_v = 1$ mm with $10''$ electrons/bunch in a $\Delta t_e = 10$ mm long bunch, to obtain

$$N_S = \sigma_C \frac{N_\omega N_e}{\pi (d_\omega/2)^2} \frac{d_\omega}{\sin\alpha} \frac{1}{c\Delta t_\omega} \sim 1.4 \times 10^3/\text{pulse} \tag{17}$$

where we used $\sigma_C = 3 \times 10^{-25}$ cm^2. The error in the assymmetry at each value of y is $\delta A(y) = 1/\sqrt{n_s}$ with n_s the counts in each bin. One can reach $\delta A(y) \sim 0.01$ in 3 seconds. Thus for a polarization $P_e = 0.1$ corresponding to a peak asymmetry $A = 0.01$, a 10% measurement of the polarization can be obtained in 5 minutes.

The case of longitudinal polarization is of interest at the SLC in which case the spin-dependent term (see Eq. 14) is

$$\xi_3 P_e (1 - \cos\bar{\theta}) \cos\bar{\theta} (\bar{k}_0 + \bar{k}) \tag{18}$$

which becomes significant with respect to the spin-independent contribution, especially at $\bar{\theta} = 180°$. Thus it is possible to measure the total yield of γ's as a function of ξ_3 or to measure the differential cross section. The latter method is used at SLC where the asymmetry is recorded as a function of the energy of the recoil electrons. The calculated spectrum for 100% polarization and the results obtained at SLC in 1992 [16] are shown in Fig. 6; the data indicate a longitudinal polarization $P_e = 0.22$.

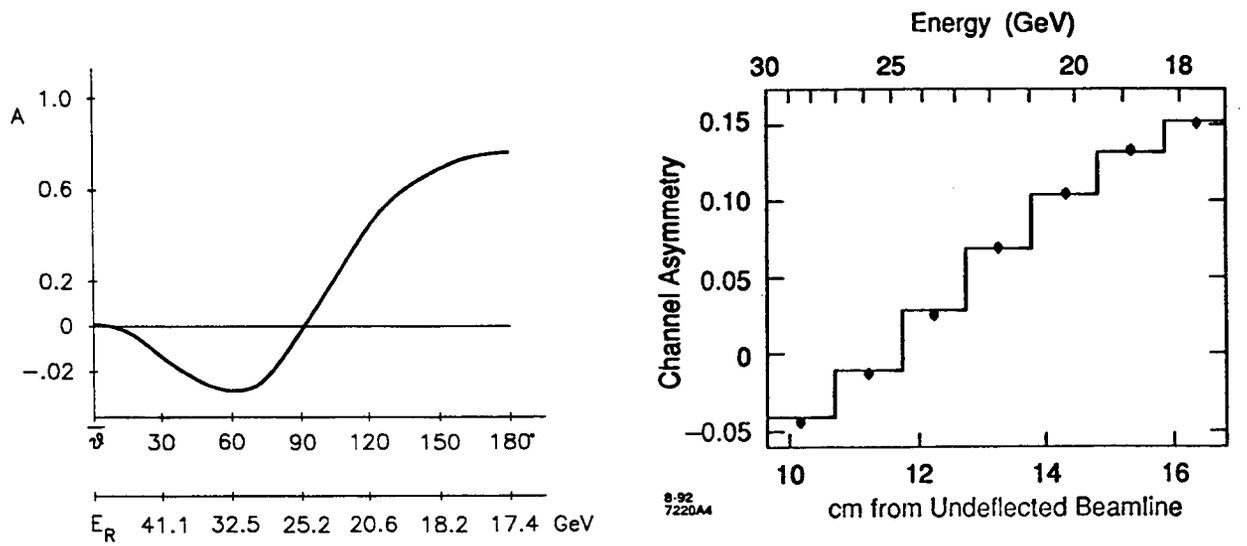


Figure 6. (a) Predicted asymmetry for 100% longitudinal electron polarization and 100% photon circular polarization as a function of recoil electron energy. (b) 1992 data from SLAC [Ref. (16)] corresponding to $P_e \simeq 22\%$

Recently, use of strained GaAs cathodes at SLC has yielded a polarization in excess of 60%. The polarization at the collision point is continuously monitored, as shown

in Fig. 7 [16]. This is because P_e enters directly in the determination of the left-right asymmetry in Z^0 production

$$A_{LR} = \frac{1}{P_e}\frac{N_L - N_R}{N_L + N_R} \tag{19}$$

where $N_L(N_R)$ is the flux normalized count of Z^0's produced with left(right)-handed electrons. Since

$$A_{LR} = \frac{2(1 - 4\sin^2\theta_W)}{1 + [1 - 4\sin^2\theta_W]} \tag{20}$$

the error in $\sin^2\theta_W$ is $\delta(\sin^2\theta_W) = (1/8)\delta A_{LR}$. The asymmetry error δA_{LR} is found from

$$\delta A_{LR} = \left[\frac{1}{N}\frac{1}{P_e^2} + A_{LR}^2\left(\frac{\delta P_e}{P_e}\right)^2\right]^{1/2} \tag{21}$$

With $N = N_L + N_R = 5 \times 10^4$, $P_e = 0.65$, $\delta P_e = 0.02$ we find $\delta A_{LR} = 0.008$ and $\delta(\sin^2\theta_W) \simeq 10^{-3}$. This compares well with the error on $\delta(\sin^2\theta_W)$ obtained at LEP.

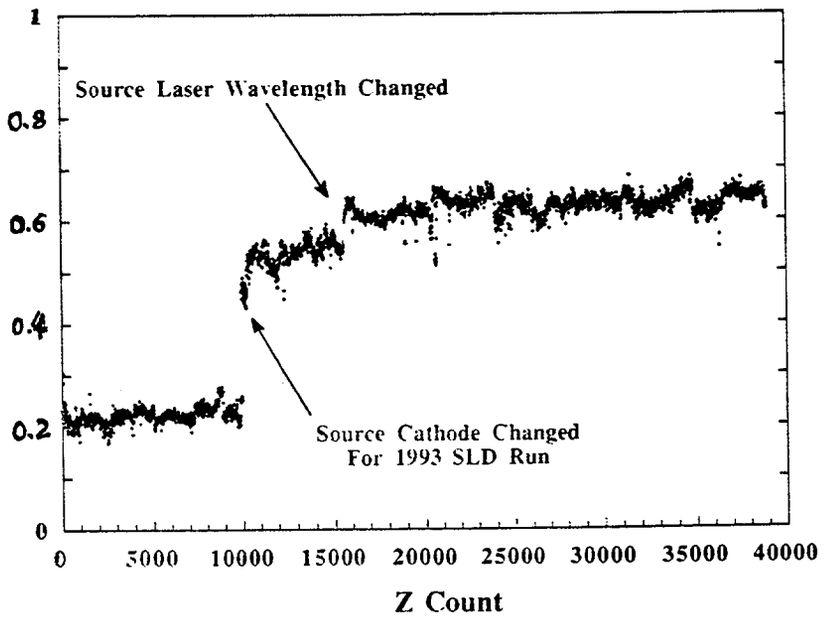


Figure 7. Electron beam polarization for the 1992 and 1993 SLAC runs (Ref. [16]).

3 MEASUREMENT OF THE LEP ENERGY

Energy measurements at LEP made in late 1991 have been reported by Arnandon *et al* [17]. The nominal energy was $E = 46.5$ GeV and thus the spin tune was near $\nu_s = 105.455$. To depolarize the beam a vertical kicker (radial magnetic field) with field integral $\int Bdl = 3.7$ Gauss-m was used. Note that

$$\nu_s(\int Bdl)_{\text{depolarizing}}/(\int Bdl)_{\text{guide}} \sim 10^{-4} \tag{22}$$

so that depolarization will occur only after many orbits. Consequently the depolarizing frequency correctly measures the average beam energy. The probability of spin flip by a

depolarizing field was first calculated by Froissart and Stora [18], who give

$$\frac{P(\text{final})}{P(\text{initial})} = 2e^{-\chi} - 1 \tag{23}$$

with

$$\chi = \frac{[\pi\nu_s(\int Bdl)_D/(\int Bdl)_G]^2}{(d\nu_s/dt)} f_0 \tag{24}$$

Here $f_0(d\nu_s/dt)$ is the rate at which the depolarizing frequency is swept. Clearly for $\chi \ll 1$ no depolarization is achieved, whereas for $\chi \gg 1$ one can expect a reversal of the polarization. Using the value of the depolarizing field given in Eq. (22) we find that the sweep rate must be kept in the range of $d\nu_s/dt \sim 10^{-4}$ Hz/s.

Figure 8 shows that a 9% polarization was established and that it could be destroyed by a static resonance, (introducing a spin bump such that $\nu_s = 106$). When the bump is removed the beam polarizes again and tends to its asymptotic level $P(t \to \infty)$ exponentially. The data are fit by using Eq.(1) but with P, τ given by

$$P(t \to \infty) = \frac{P_0}{1 + \tau_0/\tau_D} \qquad \tau = \frac{\tau_0}{1 + \tau_0/\tau_D} \tag{25}$$

where P_0 and τ_0 are the Sokolov-Ternov values given by Eqs. (2,3).

Next the depolarizing kicker was swept and a series of depolarizing resonances are shown in Fig. 9. Note that spin-flip is observed in certain cases. The details of a particular run are shown in Fig. 10. It is important to note that depolarization is recorded within a time interval of $\sim$ 16 s. Given the rate at which the kicker was swept this corresponds to an uncertainty in the beam energy $\Delta E = 1$ MeV.

The relation between the beam energy and the spin tune was given by Eq. (7); numerically we can write

$$\nu_s = \frac{(g-2)}{2}\frac{E}{mc^2} = \frac{E(\text{GeV})}{0.4406486(1)} \tag{26}$$

The beam energy is continuously monitored from a knowledge of the field of the bending magnets and the r.f. frequency. The best estimate so obtained is referred to as E_{FD} (for field display). The results of the resonant depolarization measurements established that

$$E_0 - E_{FD} = (-37.1 \pm 1.4)\text{MeV} \tag{27}$$

making an important improvement in the absolute value with which the beam energy is known.

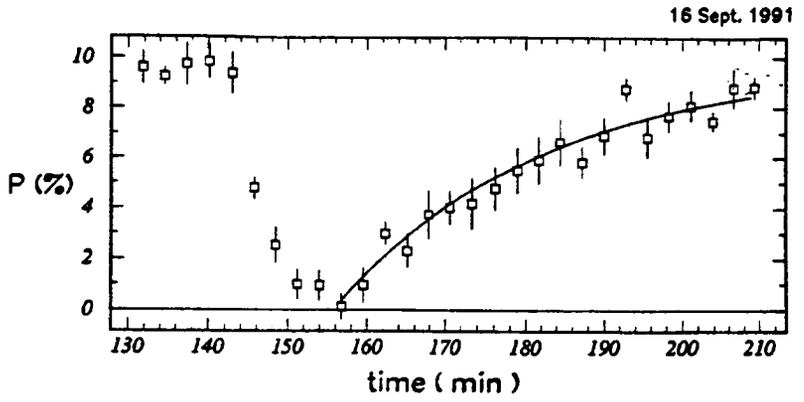


Figure 8. Depolarization of the LEP beam by the introduction of a "spin bump". Note the polarization rise when the bump is removed (Ref. [17]).

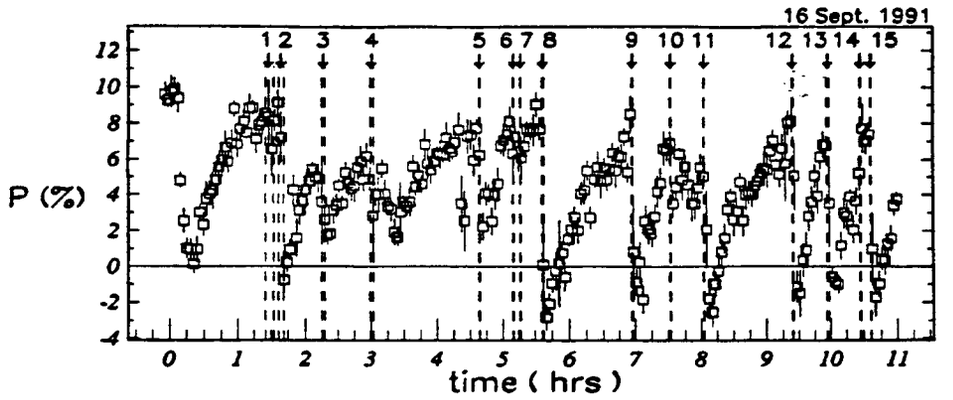


Figure 9. Sequence of resonant depolarizations at LEP. Note that occasionally the polarization can be reversed (Ref. [17]).

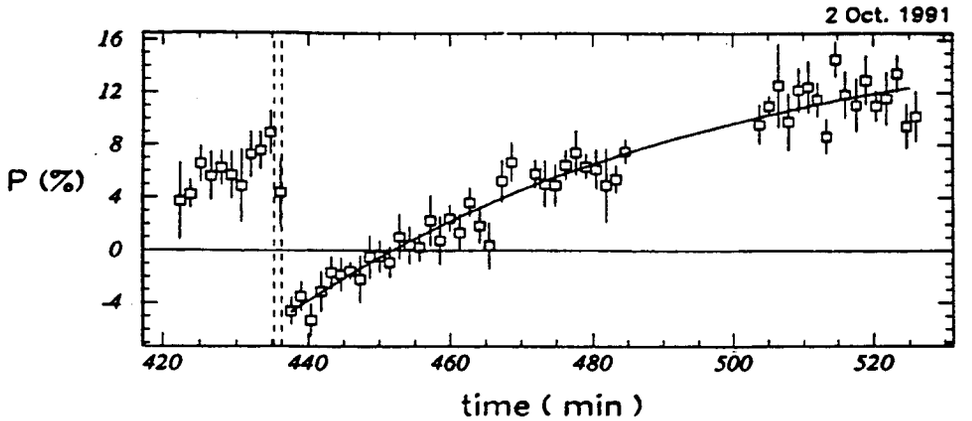


Figure 10. Detailed plot of resonant depolarization at LEP (Ref. [17]).

4 THE EFFECT OF THE TIDES

In reference [17] the authors state that the precision of the energy measurements made at LEP was $\pm 1.5 \times 10^{-5}$ but that the variability of the results was $\pm 6 \times 10^{-5}$ "consistent with the expected stability and reproducibility of the machine". The observed energy spread was almost 10 MeV. Only two effects could cause such an energy change: a change in the magnetic field, which was excluded because of the relative stability of the excitation current, or a change in the orbit radius.

The required change in orbit radius is very small because of the "momentum compaction", $(1/\alpha)$, in strong focussing machines such as LEP

$$\frac{\Delta E}{E} = \left(\frac{1}{\alpha}\right)\frac{\Delta R}{R} \qquad \alpha \sim \frac{1}{Q^2} \tag{28}$$

Here Q is the betatron tune which for LEP is of order $Q = 70$, so that $(1/\alpha) \sim 5 \times 10^3$. Thus if we take $\Delta E/E = 2 \times 10^{-4}$ we find $\Delta R/R = 4 \times 10^{-8}$ or $\Delta R = 0.15$ mm ! However for a fixed lattice, changes in the mean radius are directly reflected in the frequency of the r.f.

$$\Delta f/f = \Delta R/R \tag{29}$$

since $f = 352$ MHz at LEP, $\Delta R/R = 4 \times 10^{-8}$ would imply a frequency shift $\Delta f = 14$ Hz. Such a shift is at least an order of magnitude greater than the observed stability of the r.f. One concludes that the observed energy fluctuations in LEP cannot be explained by magnet current or r.f. instabilities.

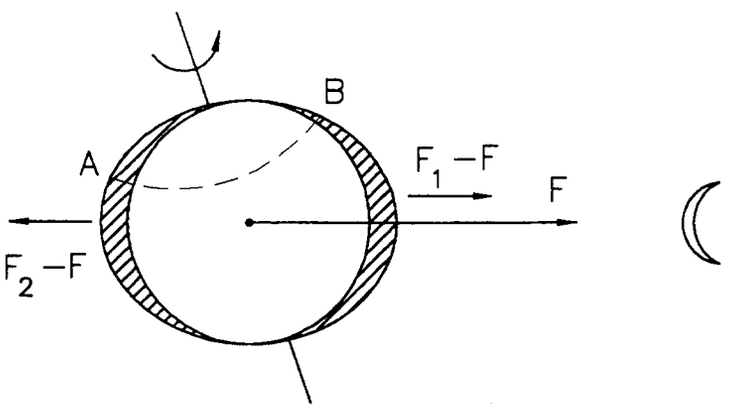


Figure 11. Illustration of the tidal forces acting on the earth due to the moon.

The answer to this puzzle was provided by G. Fisher [19] who suggested that the machine lattice is being deformed by tidal effects. To illustrate that this is possible we calculate the relative change in local gravity, g, due to the attraction of the moon as shown in Fig. 11. The tidal acceleration is given by

$$\frac{\Delta F}{m} = \frac{1}{2m}(F_1 - F_2) = G\frac{M_{\leftmoon}}{R^2_{\oplus\leftmoon}}\frac{R_\oplus}{R_{\oplus\leftmoon}} \tag{30}$$

or

$$\frac{\Delta g}{g} = \frac{M_{\leftmoon}}{M_\oplus}\left(\frac{R_\oplus}{R_{\oplus\leftmoon}}\right)^3 \sim 5.7 \times 10^{-8} \tag{31}$$

It is reasonable to assume that the relative change in the effective radius of the orbit is of order $\Delta g/g$ which reproduces the observed energy variability. Fig. 12 shows the results of a careful measurement of the LEP energy over a 26 hour period [20]. The curve gives the energy change expected from the tidal effects of the sun and moon, based on the above arguments. Further lattice distortions due to ground motion are also observed.

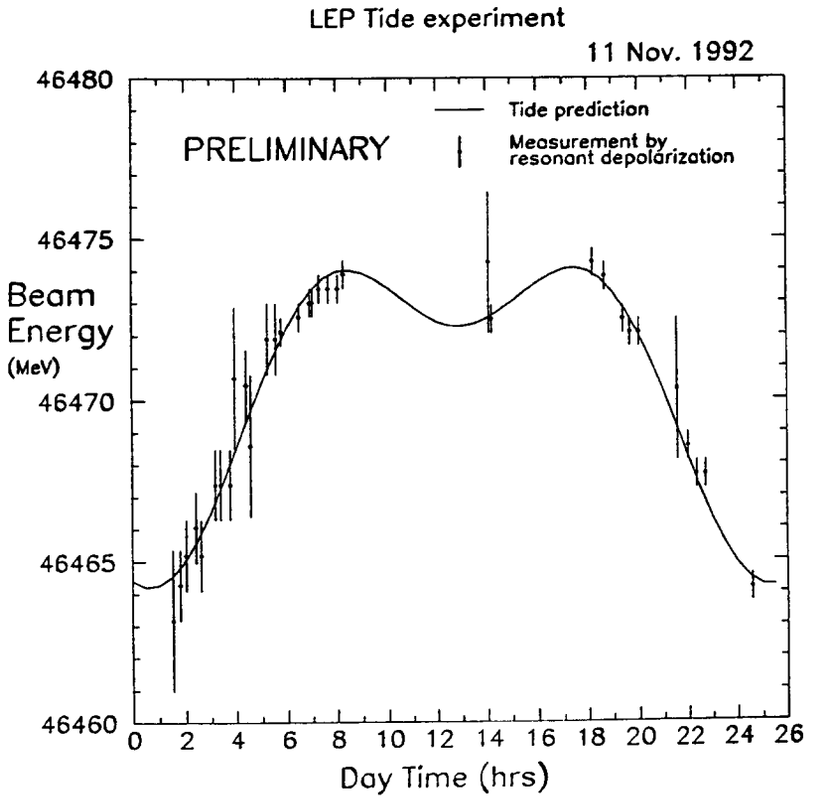


Figure 12. Data from the measurement of the LEP energy over a 26 hour period. The solid curve is the prediction of the tidal distortion of the machine lattice (Ref. [20]).

5 HAWKING-DAVIES-UNRUH RADIATION

In this section we review the possible manifestation of Hawking radiation [21] in the limiting value of the polarization in an electron storage ring. This relationship was first pointed out in a classic paper by Bell and Leinaas [22].

Hawking proved that the fluctuations of the vacuum at the horizon of a black hole result in the emission of electromagnetic radiation with a Planck spectrum at an equivalent temperature

$$kT = \frac{\hbar c}{2\pi} \frac{1}{(4GM/c^2)} \tag{32}$$

where M is the mass of the black hole. We can illustrate this result by considerations based on the uncertainty principle as shown in Fig. 13. We assume that a virtual pair of photons is produced at a distance ϵ from the horizon whose radius is $r_s = 2GM/c^2$. If one of the photons enters the horizon of the black hole it can not any more communicate

with the other photon which must therefore become real. Let $\Delta\tau$ and $\Delta\varepsilon$ be the lifetime and energy of the fluctuation in its own rest frame, so that

$$\Delta\varepsilon\Delta\tau = \hbar \tag{33}$$

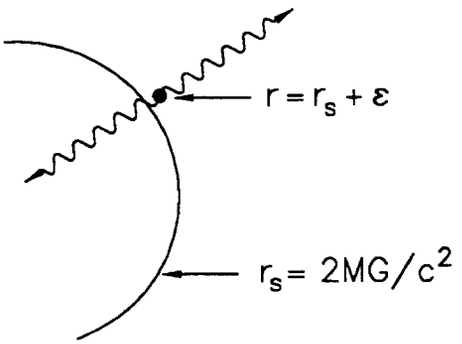


Figure 13. Sketch of virtual photon pair production near the horizon of a black hole.

Because of the strong gravitational field the time $\Delta\tau$ needed for the photon to traverse the distance ϵ to the horizon, is

$$\Delta\tau = (2/c)\sqrt{\epsilon(2GM/c^2)} \tag{34}$$

whereas the energy of the other photon when at a large distance from the horizon, is

$$E = \Delta\varepsilon\sqrt{\epsilon(2GM/c^2)} \tag{35}$$

Combining Eqs. (33-35) we obtain $E = \hbar c/(4GM/c^2)$. If we identify $E/2\pi$ with the mean energy kT of the Planck spectrum, we reproduce Eq. (32).

The acceleration at the horizon of a black hole is $a_s = GM/r_s^2 = c^4/(4GM)$, and this establishes an accelerated frame equivalent to the gravitational field at the horizon. On the basis of this equivalence Davies [23] and Unruh [23] argued that a frame of reference subject to a linear acceleration a, should be considered as a thermal bath at temperature T, with

$$kT = \frac{\hbar a}{2\pi c} \tag{36}$$

These ideas remain conjectural and have not been verified by experiment, because the emitted radiation is very weak. For instance the lifetime due to Hawking radiation, of a black hole of one solar mass is $\tau \sim 10^{54}$ years. Similarly, in the laboratory it is difficult to produce accelerated systems with significant temperature according to Eq. (36). There has however been a considerable theoretical interest in the subject [24,25].

Bell and Leinaas [22] pointed out that because of relativistic effects, the acceleration to which electrons are subject in a storage ring is significant, to the extent that the

influence of the thermal bath of Eq. (36) may be observable [26]. In the rest-frame of the electron

$$a = \frac{\gamma^2 c^2}{R} \qquad \text{or} \qquad T = \frac{\hbar c}{2\pi k}\frac{\gamma^2}{R} \tag{37}$$

where R is the orbit radius. For LEP the equivalent temperature is $T \sim 1200$ K. Considering the electron spins in equilibrium with the thermal bath, the ratio of the population of the two spin states will be given by the Boltzman distribution

$$\rho_+/\rho_- = \exp(-\mu B'/kT) \tag{38}$$

where $B' = \gamma B$ is the magnetic field in the electron rest-frame. For the magnetic moment μ we write $\mu = g\mu = g(e\hbar/2m)$ and note that in Eq. (38) we have assumed a spin-1/2 system. Furthermore $R = \gamma mc/eB$, so that the Boltzman exponent reduces to

$$\frac{\mu B'}{kT} = \left(g\frac{e\hbar}{2m}\gamma B\right)\left(\frac{\gamma mc}{eB}\frac{1}{\gamma^2}\frac{2\pi}{\hbar c}\right) = \pi g \tag{39}$$

The above result is remarkable in that it is independent of the electron energy or of the ring parameters. Thus for all storage rings the population ratio ρ_+/ρ_- should be the same. The beam polarization is a measure of the population ratio,

$$P = \frac{\rho_- - \rho_+}{\rho_- + \rho_+} = \frac{1 - e^{-\pi g}}{1 + e^{-\pi g}} \tag{40}$$

If we use $g = 1$ in Eq. (40) we find $P = 0.9174$. This is to be compared with the maximal polarization $P_0 = 0.924$ given by Eq. (2). Therefore, the assymptotic polarization in electron storage rings derived by considering the quantum-mechanical spin-flip transitions may be the first tangible manifestation of the Hawking-Davies-Unruh effect. The reader will wonder why we used $g = 1$ instead of the known value of $g = 2$ for the electron; this is related to the Thomas precession in the circular orbit. For this point and for estimates of the time constant of the build-up of the polarization one should consult the original paper [22].

6 ACKNOWLEDGEMENTS

It is a pleasure to thank Bernd Dehning of CERN for keeping me posted on the progress of the polarization and energy measurements at LEP, and Mike Woods of SLAC for providing me with unpublished data from the SLC polarization measurements.

7 REFERENCES

[1] I.M. Ternov, Yu. M. Loskutov and K. L. Korovina, Zh. Eksp. Teor. Fiz. 41, 1294 (1961) [Sov. Phys. JETP 14, 921(1962)].

[2] A.A. Sokolov and I. M. Teruov, Dokl. Akad. Nauk SSSR 153, 1052 (1963) [Sov. Phys. Dokl 8, 1203 (1964)]; see also J.D. Jackson, Rev. Mod. Phys. 48, 417 (1976).

[3] J. P. Koutchouk "e^+e^- polarization", these Proceedings.

[4] V. Bargmann, L. Michel and V.L. Telegdi, Phys. Rev. Lett. 2, 435 (1959); see also J.D. Jackson "Classical Electrodynamics", J. Wiley, N.Y. 2nd edition; note that in Eq. (4) we have dropped the contribution of the electric field.

[5] R.S. Van Dyck Jr. *et al.*, Phys. Rev. Lett. 59, 26 (1987).

[6] V.N. Baier and Yu. F. Orlov, Dokl. Akad. Nauk SSSR 165, 783 (1965) [Sov. Phys. Dokl. 10,1145 (1966)].

[7] J. Le Duff *et al.*, "Proceedings of the 1971 Symposium on Electron and Photon Interactions at High Energies", edited by N.B. Mistry, Cornell University, Ithaca, N.Y. 1972.

[8] U. Camerini, D. Cline, J.G. Learned, A.K. Mann and L.K. Resvanis, Phys. Rev. D12, 1855 (1975); J.G. Learned, L.K. Resvanis and C.M. Spencer, Phys. Rev. Lett. 35, 1688 (1975).

[9] L.A. Page, Rev. Mod. Phys. 31, 759 (1959).

[10] R.H. Milburn, Phys. Rev. Lett. 10, 75 (1963).

[11] V.N. Baier and V.A. Khoze, Sov. J. Nucl. Phys. 9, 238 (1969).

[12] D.B. Gustavson *et al.*, Nucl. Instr. Meth. A165, 177 (1979).

[13] F.W. Lipps and H.A. Tolhoek, Physica 20, 85 and 395 (1954).

[14] M. Placidi and R. Rossmanith, Nucl. Instr. Meth. A274, 79 (1989).

[15] L. Knudsen *et al.*, Physics Letters B270, 97 (1991).

[16] M. Woods, private communication.

[17] L. Arnaudon *et al*, Physics Letters B284, 431 (1992).

[18] M. Froissart and R. Stora, Nucl. Instr. Meth. 7, 297 (1960).

[19] G.E. Fisher "Ground Motion - An Introduction for Accelerator Builders", Proceedings of the CERN Acclerator School, Montreux, 1992, S. Turner ed., CERN 92-05 (1992).

[20] M. Placidi *et al*, "Polarization Results and Future Perspectives", Proceedings of the Third Workshop on LEP Performance, CERN SL/993-19DI, p. 281; B. Dehning, private communication.

[21] S. Hawking, Nature 248, 30 (1974); Comm. Math. Phys. 43, 199 (1975).

[22] J.S. Bell and J.M. Leinaas, Nucl. Phys. B212, 131 (1983).

[23] P.C.W. Davies, J. Phys. A8, 609 (1975); W.G. Unruh, Phys. Rev. D14, 870 (1976).

[24] See for instance, T.H. Boyer, Phys. Rev. D21, 2137 (1980).

[26] However Eq. (36) is valid for a linearly accelerated frame and the effects of the circular trajectory in the storage ring are not fully understood as yet.

Plate 1 Nicholas C. Christofilos in 1960. At that time he was group leader in controlled thermonuclear reactions at Lawrence Livermore National Laboratory. Courtesy LLNL.

NICHOLAS C. CHRISTOFILOS: HIS CONTRIBUTIONS TO PHYSICS

A. C. Melissinos
Dept. of Physics & Astronomy, University of Rochester, Rochester, NY USA

Abstract

Nicholas Christofilos is best known for his discovery of the strong focussing principle in 1949. However he continued making important and seminal contributions to physics until his death. These included the invention of the Linear Induction Accelerator, research in thermonuclear fusion and on applications of physics to National defense. A brief discussion of these accomplishments as related by his friends and colleagues is presented in honor of his memory.

1 THE EARLY YEARS

Nick Christofilos was born in Boston in 1916 to Constantine and Eleni Christofilos, who had immigrated from Greece to the US. In 1923 the family returned to Greece and Nick grew up and was educated in Athens. In a twist of fate Christofilos would return to the US and spend the greater part of his scientific career there. In fact he was extremely proud of his origin and had not relinquished his US citizenship. Yet, it was while living and working in Greece that he conceived and detailed the strong focussing principle and his scheme for plasma containment, two of his most creative and original contributions to science.

Nick Christofilos' technical ability was evident at a young age. He was a gifted youngster who as a hobby built radios and repaired electrical equipment. At age 18 he entered the National Polytechnic of Athens which at that time, was the premier higher education institution in Greece. He graduated in 1938 with a degree in Electrical and Mechanical Engineering and started working for an elevator maintenance concern, and later established his own firm.

The work at the firm seems to have allowed enough free time for Christofilos to immerse himself in the reading of physics texts – principally in German – in the area of nuclear physics and accelerators. By his own account he was particularly influenced by S. Fluegge's "Introduction to Nuclear Physics" (1942) and by A. Bouwers' "Electrical High Voltages" (1938). The war years and Greece's occupation (1941-44) caused further isolation but did not reduce Nick's interest in science. After the war he would use the USIS (United States Information Service) library in Athens to regularly read the Physical Review. Already in 1946 he had developed an accelerator scheme in principle similar to the synchrotron. He was not inclined to publish but instead submitted a patent application both in the US and Greece with copies to Berkeley. That proposal had shortcomings and in the meantime "phase stability" had been discovered by E. McMillan and simultaneously by V. Veksler.

Continuing his involvement with accelerators, in 1949 Christofilos discovered the strong focussing principle in a bold departure from the then accepted technology. Again

he spurned publication, in spite of the urging of his friend the physicist Th. Kouyoumzelis, and instead submitted patent applications in the US and Greece [1]. He also sent copies to Berkeley where the paper was passed from scientist to scientist without being understood or appreciated. In the next section we discuss in some detail Christofilos' proposal as it appears in the patent application, a document of extreme clarity and covering a wide range of applications. Nick always considered himself a "practical physicist" and in all his papers the calculation of physical magnitudes and the technical feasibility of his ideas are prominently presented. An example was his early proposal of flywheel excitation for energy recovery in synchrotrons.

2 STRONG FOCUSSING

The application for a patent entitled "Focussing System for Ions and Electrons" was made on March 10, 1950, and the patent was granted in 1956. In Figs. 1, 2 we reproduce the first page of the patent and a figure showing the design of a strong focussing magnet. The author considers a particle moving along the x-axis with restoring forces along the y- and z-axes. He argues that in the case of ideal focussing it should hold

$$F_x = 0 \qquad F_y = -\epsilon_y y \qquad F_z = -\epsilon_z z \tag{1}$$

The restoring forces will be electromagnetic and derivable from a potential, $\vec{F} = \vec{\nabla}\phi$. Since in the region of the beam the potential must obey Laplace's equation, $\nabla^2\phi = 0$, the focussing force has zero divergence

$$\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} = 0 \tag{2}$$

It therefore follows that in Eq. (1)

$$\epsilon_y = -\epsilon_z \tag{3}$$

Namely an ideal focussing system does not exist. Instead, one could alternate the focussing between the y and z axes. By definition in a defocussing region the displacement is smaller than in a focussing region, and since the force is proportional to the displacement, the net effect is focussing.

In the patent a continuously varying gradient was considered,

$$F_y = -\epsilon y \sin(\frac{2\pi x}{\lambda}) \quad F_z = \epsilon z \sin(\frac{2\pi x}{\lambda}) \tag{4}$$

so that a particle injected at $z = z_0$, with $dz/dx = 0$ would follow a trajectory

$$z = z_0[1 - \mu \sin(\frac{2\pi x}{\lambda})] \tag{5}$$

where

$$\mu = \frac{\epsilon\lambda^2}{4\pi^2\beta^2 m\gamma c^2} \tag{6}$$

and $\gamma = E/m$, $\beta = \sqrt{\gamma^2 - 1}/\gamma$ have their usual meaning. Clearly, the stability condition is $0 < \mu < 1$. We recognize the usual formalism for alternating gradient machines simplified by the sinusoidal variation of the gradient.

United States Patent Office

2,736,799
Patented Feb. 28, 1956

2,736,799

FOCUSSING SYSTEM FOR IONS AND ELECTRONS

Nicholas Christofilos (or Philos), Athens, Greece

Application March 10, 1950, Serial No. 148,920

8 Claims. (Cl. 250—27)

The present invention relates to a new focussing system for ions and electrons and application thereof in particle accelerators.

A major problem in the design of particle accelerators is the provision of suitable means for focussing the accelerated particles towards a predetermined orbit and compensating the mutual electrostatic repulsive forces.

An ideal focussing system must accelerate the moving particles towards a predetermined orbit from all directions and the focussing forces must increase as the distance from said orbit increases.

If we consider an orthogonal coordinate system, x, y, z, and suppose that the particle's orbit coincides with the x-axis and considering as P_x, P_y, P_z the x, y, z, components of the focussing forces, then, in an ideal focussing system the equations of the P_x, P_y, P_z would be

$$P_x=0 \tag{1}$$

$$P_y=-\epsilon_y \cdot y \tag{1a}$$

$$P_z=-\epsilon_z \cdot z \tag{1b}$$

But simultaneously the Laplace equation $\Delta\varphi=0$ must be satisfied so that it must be

$$\frac{\delta P_x}{\delta x}+\frac{\delta P_y}{\delta y}+\frac{\delta P_z}{\delta z}=0 \tag{1c}$$

or

$$\epsilon_y=-\epsilon_z \tag{1d}$$

From the above equations is shown that a focussing field capable to accelerate ions or electrons towards a predetermined orbit from all directions simultaneously is impossible. Therefore the focussing system proposed herein is based in a new principle, namely:

If, along a predetermined orbit of ions or electrons an electrostatic or electromagnetic field is produced by means of suitably arranged conductors (connected to a high voltage source or energized by high intensity current) exerting on the moving, along said orbit, particles (ions or electrons) forces directed normally to said orbit and varying periodically, in direction and magnitude along said orbit, and increasing in magnitude as the distance from said orbit increases, then the mean value of the focussing forces is negative (directed towards the orbit) and the particles are focussed towards the orbit from all directions.

The focusing forces acting on the particles resulting from the field which is produced electrostatically or electromagnetically, increase as the distance from the orbit increases. The particles move at some finite distance from the orbit and in a direction substantially parallel to the orbit by virtue of the periodically varying exciting focusing forces due to the field. The particles undergo forced oscillations and are subject to the alternately converging and diverging forces from the field. The electrically produced force field, electromagnetic or electrostatic, imposed upon the orbit and the path of the particles exerts forces on the particles within a plane whose normal is substantially parallel to the velocity vector of each of the particles. The path of the particles becomes concave towards the orbit in a converging section and convex towards the orbit in the diverging section. Since the forces are greater as the distance from the orbit becomes greater, the mean value of the converging and diverging forces along the converging section is greater than the mean value of the forces along the diverging section. The resultant force and net effect of the mean value of these alternating forces causes the particles in the path to be forced towards the orbit from all directions and focusing is thereby obtained.

In a focussing system based upon this principle the x, y, z, components of the focussing forces are

$$P_x=0 \tag{2}$$

$$P_y=-\epsilon \cdot y \cdot \sin\frac{2\pi z}{\lambda} \tag{2a}$$

$$P_z=\epsilon \cdot z \cdot \sin\frac{2\pi z}{\lambda} \tag{2b}$$

As it is obvious

$$\Delta\varphi=\frac{\delta P_x}{\delta x}+\frac{\delta P_y}{\delta y}+\frac{\delta P_z}{\delta z}=0$$

$$\epsilon_y=-\epsilon_z \tag{1d}$$

so that the Laplace equation is satisfied and therefore the production of such a field is possible. If we consider a particle moving parallel to the x-axis and at a distance $z=z_0$, $y=0$ the force exerted on said particle is

$$P_z=\epsilon \cdot e \cdot z_0 \sin\frac{2\pi z}{\lambda} \tag{3}$$

As the force P_z varies periodically as the particle moves along the x-axis, said particle undergoes forced oscillations of frequency

$$f=\frac{\beta c}{\lambda} \tag{4}$$

where βc the velocity of the particle.

The result of these oscillations is that the distance from the orbit oscillates around the mean value z_0 according to the equation

$$z=z_0\left(1-\mu \sin\frac{2\pi z}{\lambda}\right) \tag{5}$$

where

$$0<\mu<1 \tag{6}$$

In the region where

$$\sin\frac{2\pi z}{\lambda}$$

is negative the mean value of the distance from the orbit is greater than z_0 while in the region where

$$\sin\frac{2\pi z}{\lambda}$$

is positive the mean value of the distance is less than z_0, so that the mean value of the force in the first region is greater than the mean value in the second region, with the result that the mean value of the force in a length λ is negative, focussing the particle towards the x-axis, from all directions.

If the maximum value of the force is

$$P_{max}=e \cdot \epsilon \cdot z_0 \tag{7}$$

then the mean value P_m is

$$P_m=e \cdot \epsilon_m \cdot z_0 \tag{8}$$

where

$$\epsilon_m=\epsilon \cdot \frac{\mu}{2} \tag{9}$$

and

$$\mu=\frac{\epsilon\lambda^2}{4\pi^2\beta^2 1^2} \tag{10}$$

Figure 1: The first page of the strong focussing patent issued to Nicholas Christofilos.

Feb. 28, 1956 2,736,799

NICHOLAS CHRISTOFILOS (OR PHILOS)

FOCUSSING SYSTEM FOR IONS AND ELECTRONS

Filed March 10, 1950 4 Sheets-Sheet 2

Fig. 5

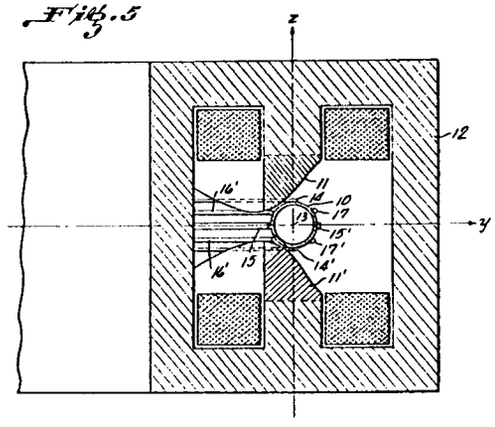
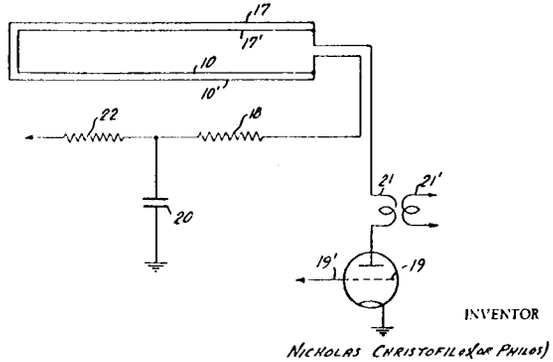


Fig. 6



INVENTOR

NICHOLAS CHRISTOFILOS (OR PHILOS)

BY Woodcock, Knight & Pennock

ATTORNEYS

Figure 2: A figure from the patent showing the design of a strong focussing magnet.

To achieve the configuration of Eq.(4) a radially varying magnetic field with index K is proposed

$$B_z = B_0[1 - (1 - R/R_0)K\sin(n\phi)] \tag{7}$$

where R_0 is the equilibrium radius. In the usual way this leads to restoring forces

$$\begin{aligned} F_z &= -z(\beta^2 m\gamma c^2/R_0^2)K\sin(n\phi) \\ F_y &= y(\beta^2 m\gamma c^2/R_0^2)K\sin(n\phi) \end{aligned} \tag{8}$$

as required by Eq.(4). The index K can exceed unity and is related to the periodicity n, through

$$n = \sqrt{K/\mu} \tag{9}$$

The betatron frequencies are given by

$$\omega_z = \omega_0\sqrt{K_d} \qquad \omega_y = \omega_0\sqrt{K_d + 1} \tag{10}$$

where

$$K_d = K^2/2n = K\mu/2 \tag{11}$$

and $\omega_0 = eB_0/\gamma m$ is the orbital frequency.

As an example, Christofilos calculates the parameters for a 6 GeV machine with field index $K = 250$. He shows the important reduction in the beam pipe diameter and the corresponding economy in magnet size. He is concerned about coupling of vertical and radial oscillations and introduces the concept of the quadrupole lens.

Regretfully, Christofilos' discovery went unnoticed for almost three years. Since it did not appear in a physics journal, it did not attract attention and those who received private copies either did not read them, or did not appreciate the importance of the new principle. In 1952 strong focussing was rediscovered by E. Courant and H. Snyder, following a suggestion by M. S. Livingston to examine a lattice with alternately placed C-magnets [2]. This was an independent discovery promptly published and immediately applied to the design of the new accelerators at Brookhaven, Cornell and CERN. In contrast, Christofilos' discovery was made in isolation from scientific colleagues and with only very limited access to the literature.

In 1953 Nick Christofilos visited the U.S. and while reading the Physical Review at the Brooklyn Public Library came upon the Courant, Livingston and Snyder article. He thought that his idea had been used without acknowledgement and rushed to Brookhaven where he was shown around by John Blewett. After some excited discussion it became clear that Christofilos deserved the credit for the earliest enunciation of the principle of strong focussing but also that the Brookhaven discovery was completely independent. Christofilos was immediately offered a position at Brookhaven where he joined the team designing the 28 GeV Alternating Gradient Synchrotron (AGS) under G.K. Green. Eventually a settlement was reached with the Atomic Energy Commission for the use of his patented idea in accelerators then under construction.

3 "ASTRON" AND THE LINEAR INDUCTION ACCELERATOR

Nick Christofilos was an intense and colorful personality. He bristled with ideas and loved to explain them and argue about them. At Brookhaven he contributed to the de-

Plate 2 Nicholas Christofilos, Paul Weiss (left) and Eugene Lauer in front of the 90 foot chamber of Astron in 1964. Courtesy LLNL.

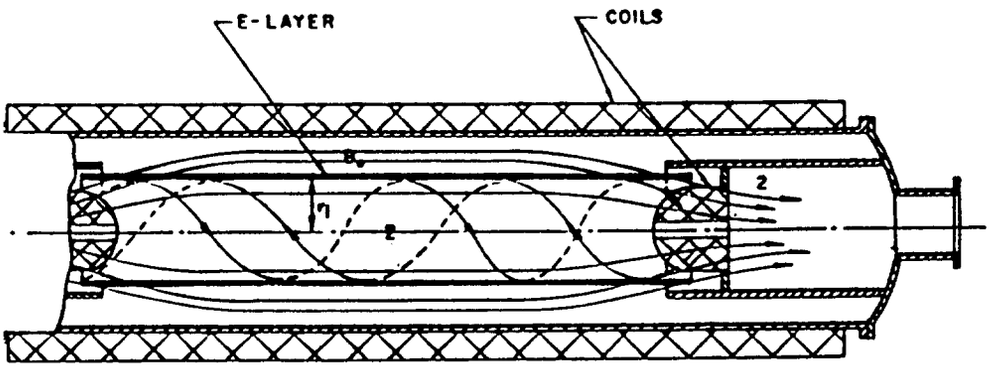


Figure 3: Conceptual Schematic of the "Astron" containment vessel (from Ref. [4]).

sign of the drift tubes for the 50 MeV proton linac [3] but his principal interest was the Astron proposal. Astron is the Greek word for "star" which Nick, appropriately coined for his thermonuclear reactor project. Since all fusion work at that time was classified, it was decided that the work on Astron would be carried out at Livermore (now Lawrence Livermore National Laboratory, LLNL). This led to many amusing incidents until Nick's clearance came through. In one instance he was to present his proposal at a meeting of the "Sherwood project" held in a rented movie theater in Berkeley. While the audience could freely ask questions of the speaker Nick was frustrated since his queries could not be answered. In 1956 Christofilos joined Livermore where he spent the rest of his career.

The fundamental idea of Astron [4,5] was to create closed magnetic lines by establishing a cylindrical layer of electrons at high energy and density. Since electrons and ions would spiral around the magnetic field lines [6] they would remain confined. Furthermore the ions would be heated by colliding with the energetic electrons in the so-called "E-layer". The magnetic lines would by necessity be closed if the field inside the E-layer was reversed with respect to the external field that established the E-layer in the first place. This is shown schematically in Fig. 3. Consider an external axial field B_0 as shown, so that electrons injected perpendicular to the field will follow a circular orbit of radius $R = m\gamma\beta/eB_0$. By appropriate field shaping the electrons can be made to drift axially so as to form a cylindrical layer characterized by a linear density of n_0 electrons/cm. The field generated by this electron layer is directed opposite to B_0 (Lenz's law) and is given by

$$B_{\text{in}} = 2\pi(I/\ell) = 2\pi e n_0 \omega_0 = n_o(e^2 B_0/\gamma m) \tag{12}$$

Therefore "inversion" will be achieved when $\vec{B}_{\text{in}} = -\vec{B}_0$ or

$$n_o = \gamma/(e^2/m) \tag{13}$$

where $e^2/m \equiv r_e = 2.82 \times 10^{-13}$ cm is the classical electron radius. Thus for 50 MeV electrons ($\gamma \sim 100$) the required linear electron density is of order $n_0 \sim 3 \times 10^{14}$ cm^{-1}, a rather demanding condition.

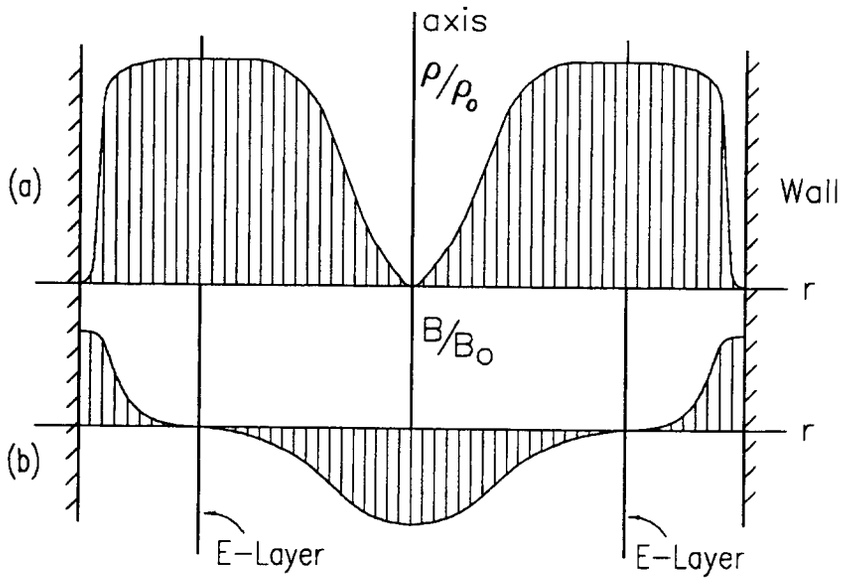


Figure 4: (a) Ion density and (b) magnetic field radial distributions calculated for "Astron" (from Ref. [5]).

A set of design parameters for Astron is given in Table 1. Note that the required external field is an order of magnitude larger than the field seen by the E-layer; the calculated field variation and the ion density are shown in Fig. 4. However the electron density needed for inversion was never achieved because of instabilities in the E-layer, in particular when stacking multiple electron pulses. This in spite of a large effort by many talented physicists who worked on the project. The project was terminated in 1973.

Table 1. Parameters for the Astron Thermonuclear Reactor

E-layer electron energy	50 MeV
E-layer radius	50 cm
B_0	3 T
Length of E-layer	30 m
Plasma temperature	25 keV
(D-T ignition)	8 keV
Ion density	$6 \times 10^{14}/\text{cm}^3$
Net electric power output	0.5 GW

The lasting contribution that resulted from the Astron program was the invention and development of the Linear Induction Accelerator. A high current electron source was needed to establish and maintain the E-layer. Christofilos was aware of the inadequacy of travelling wave linacs for this purpose and set out to invent the "induction linac". Again the principle is simple: the electrons pass through a region of changing magnetic flux and gain energy by Faraday's law [7]. A simplified schematic is shown in Fig. 5 where a time dependent magnetic field is excited in a torroidal loop of area A; electrons cross

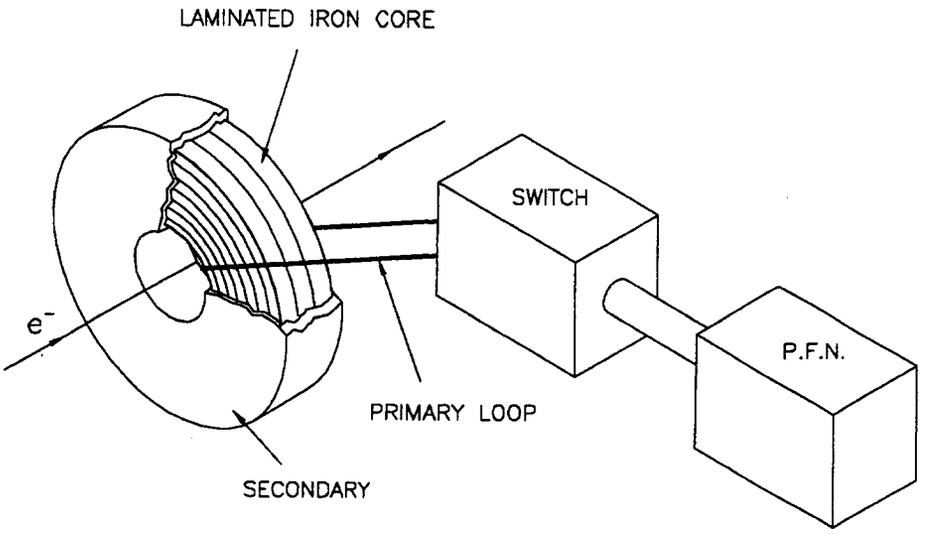


Figure 5: Principle of the linear Induction Accelerator (from Ref. [8]).

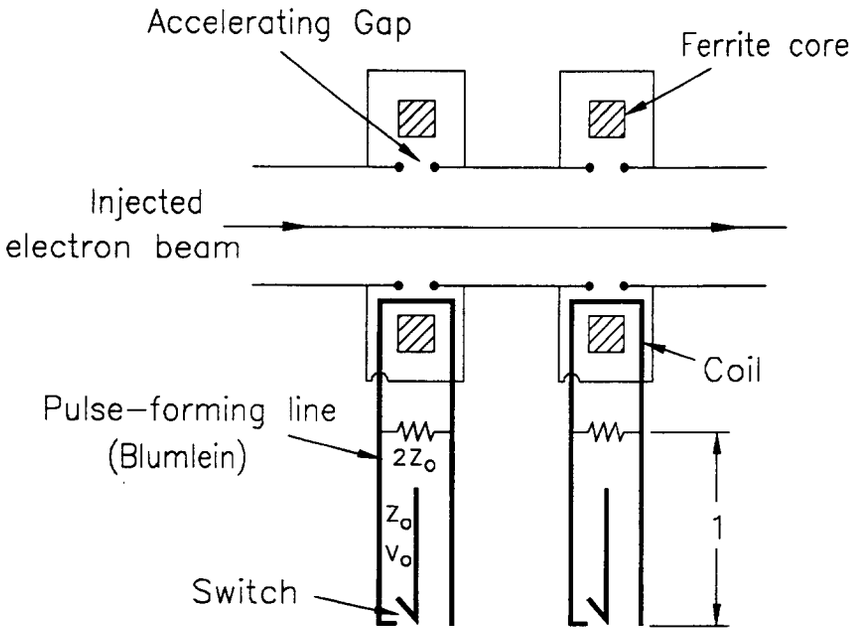


Figure 6: Schematic of a 2-Cavity Linear Induction Accelerator (from Ref. [7]).

Figure 7: The Advanced Test Accelerator. Photograph courtesy Lawrence Livermore National Laboratory.

the center of the torroid so that the energy gained is given by the rate of magnetic field change integrated over the torroid area

$$\Delta E = e \oint \vec{\mathcal{E}} \cdot d\vec{\ell} = -e \int \frac{\partial \vec{B}}{\partial \vec{t}} \cdot d\vec{A} \tag{14}$$

For practical values of $\partial B/\partial t$ and magnetic core area one can easily reach $\Delta E \sim 100$ keV. Several such cores can then be stacked in series to achieve the desired energy.

Induction accelerators are often compared to transformers. In the case of the betatron, the beam is considered as a multiturn secondary which is raised to high voltage by the changing magnetic flux threading the orbit. In the induction linac the beam is viewed as a single turn secondary which can carry high current. In practice, when several induction stages are used to reach the desired voltage, the primary function of the magnetic cores is to isolate one accelerating cavity from the next. This is shown in Fig. 6 which is a schematic of a two-cell linear induction accelerator such as the Advanced Test Accelerator (ATA). A 250 kV pulse is applied directly across the acceleration gap but also excites the ferrite core. This places a high impedance between the pulse feed and its return so that the current loop is closed through the beam [8].

The first induction linac was that used as an injector for Astron. It delivered a 4 MeV pulse at 150 A and was completed in 1964. In 1968 the accelerator was upgraded to deliver 600 A of 6 MeV electrons. A prototype for a second generation accelerator the ETA (Experimental Test Accelerator) began operation in 1979 and was capable of 10 kA of 5 MeV electrons. This was followed by the ATA (Advanced Test Accelerator) which achieved 10 kA operation in 1984. The parameters of the ATA are given in Table 2 and a view from the electron gun end is shown in Fig. 7. Both the ETA and ATA were capable of continuous operation at 1 Hz and of 5-10 pulse bursts at 10 kHz.

Table 2. The Advanced Test Accelerator

Primary Voltage	250 kV
Pulse duration	35 ns
Repetition rate	1-5 Hz
Number of cavities	200
Length	85 m
Electron Energy	50 MeV
Average current in pulse	10 kA
Energy in pulse	18 kJ

4 NATIONAL DEFENSE

A significant part of Christofilos' activities at Livermore was devoted to National defense. He understood the military needs of the country at the height of the cold war and was proud to contribute his vast scientific knowledge and mental agility to such problems. Much of that work especially on directed energy weapons still remains classified. There are today in excess of 200 declassified technical reports which he authored or co-authored while at Livermore. Among the better known of these projects are the "Argus" experiment and the proposal for low frequency communications code named "Sanguine".

In 1957 Christofilos proposed that one could create an artificial belt of electrons around the earth by injecting at sufficient altitude energetic electrons into the earth's magnetosphere [9]. The electrons would spiral along the magnetic field lines but would be reflected at the reentry points due to the "magnetic mirror" effect. This is shown in Fig. 8 and follows from the (adiabatic) invariance of Ba^2 where a is the radius of the spiral and B the local value of the magnetic field. It follows that

$$\frac{p_\perp^2}{B} = \frac{mv_\perp^2}{B} \text{ is conserved and } = m\frac{v_{\perp 0}^2}{B_0^2} \tag{15}$$

and therefore

$$v_{||}^2 = v_0^2 - v_\perp^2 = v_0^2 - v_{\perp 0}^2 \frac{B(z)}{B_0} \tag{16}$$

Thus the axial velocity $v_{||}$ will go through zero and reverse direction for a suitable increase of the axial field.

Plate 3 Nicholas Christofilos in 1958 discussing the "Argus" experiment. Courtesy LLNL.

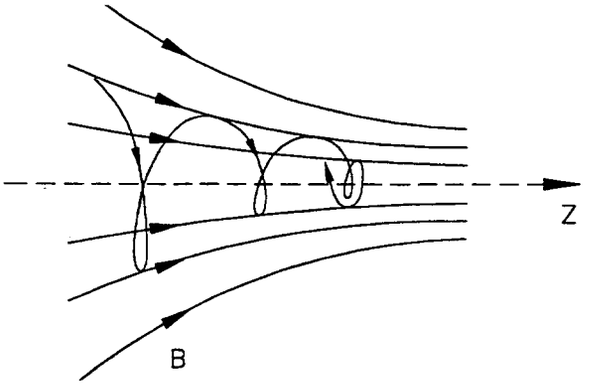


Figure 8: Mirror effect for a trapped particle in an increasing magnetic field.

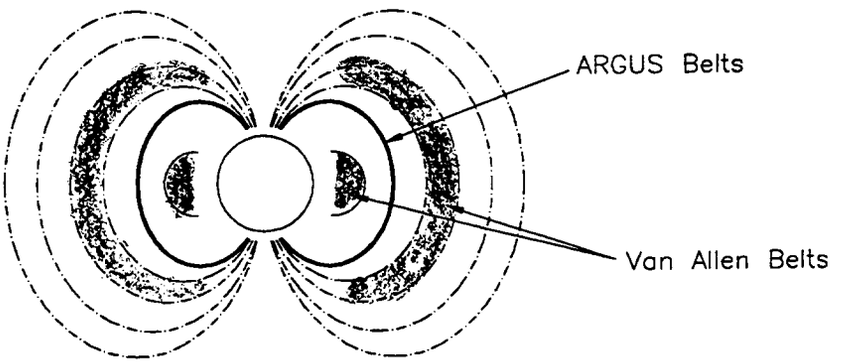


Figure 9: Approximate location of the electron belts produced by the "Argus" experiment.

It is important to note that the Argus proposal was made before the discovery of the Van Allen belts. The source of electrons would be a nuclear detonation and would result in a very dense electron layer which could, for instance, be useful for defense against incoming missiles. The experiment was carried out from August 27 to September 6, 1958 by detonating three small (1-2 kiloton) nuclear devices over the South Atlantic Ocean, at a nominal altitude of 180 km. The radiation was probed by the Explorer 4 satellite which was placed in a near earth orbit.

Minutes after the detonation the radiation could be detected in the Northern hemisphere. Figure 9 shows the position of the Argus belts in relation to the natural Van Allen belts. In total 180 crossings of the sattelite through the artificial electron belts were observed, the typical energy of the detected electrons being 3 MeV; a crossing lasted approximately 30 seconds. The lifetime of the belts, in one case, was as long as 15 days, in excess of the calculated effect. Argus was an experiment on a global scale, and pointed out how vulnerable the earth can be to modern technology.

"Sanguine", was the code name of the U.S. Navy for its project of extremely low frequency (elf) communication with submerged submarines. The idea was to use the Schumann resonances for the propagation of these low frequency waves around the globe. These resonances arise because the ionosphere and the surface of the earth form a waveguide in which TM waves can propagate as shown in Fig. 10. Neglecting the distance to the ionosphere the frequency of the lowest modes is given by

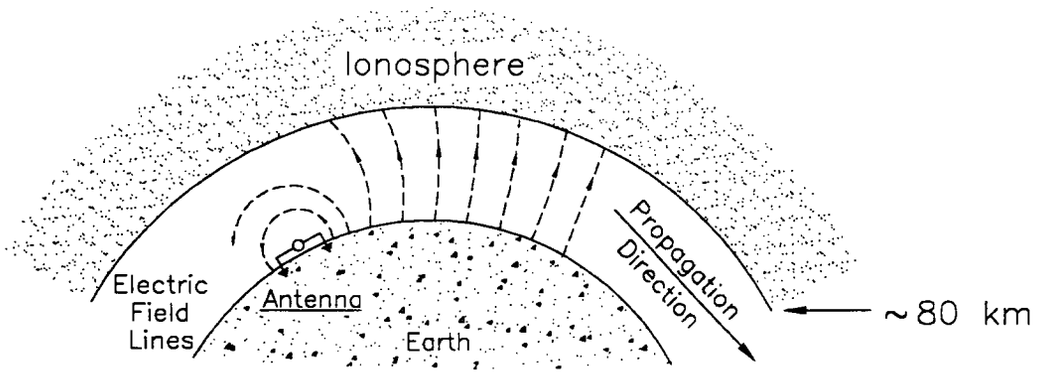


Figure 10: Excitation of a Schumann ELF mode in the earth-ionosphere system.

$$\nu \simeq \frac{c}{2\pi a}\sqrt{\ell(\ell+1)} \tag{17}$$

with a the radius of the earth [10]. The observed frequencies are closer to 8, 14, ... Hz.

The problem solved by Nick Christofilos was how to excite the resonances [11]. He proposed a linear antenna shorted into the earth at both ends. Because of the reflection at the conducting surfaces of the ionosphere and of the earth the antenna is equivalent to an infinite sequence of dipoles and couples effectively into the TM mode. A site in Wisconsin was under consideration for a long time envisaging a 150 km long antenna operating at 40 Hz. It would radiate 4 kW for a power input of 38 MW.

These two examples are typical of Christofilos' practical solutions to problems on any scale. He was a scientific genius ahead of his time and would always seek a novel approach to a problem. Yet he was widely read and had extensive knowledge of the work done by others. The traits of his Greek upbringing stayed with him to the end, evident in his supreme self-confidence, his occasional impatience and in his ultimate generosity to all.

5 EPILOGUE

Nicholas Christofilos died of a heart attack on September 24, 1972 in Hayward CA. Typically, the previous night he had worked late at Livermore. In his life he had been honored in 1963 by the Elliot Cresson medal of the Franklin Institute and in 1959 by the National Science Achievement Award of the American Hellenic Progressive Association (AHEPA). Posthumously the ATA was dedicated in 1980 to his memory. Christofilos was married twice: in 1954 to Elly Christofilos and in 1960 to Joan Jaffray. He left two sons, Nicholas Christofilos, Jr. from his first marriage and Jason Christofilos from the second marriage.

Nick worked extremely hard at his "physics" and was at the Laboratory every day from ten in the morning to midnight. He asked a lot from the physicists, engineers and technicians who worked with him. Through the force of his personality and the

example of his own dedication he got extraordinary performance from his team. It is at Livermore that he formed his most lasting scientific friendships and to this day his former collaborators speak of him with great affection. I had met Nick Christofilos only briefly, however it seems to me that his personality is best summed up by the closing paragraphs of his obituary written by his friends, and which appeared in Physics Today [12]:

> "... Many of his ideas were controversial and defending them would sometimes seem to demand more ingenuity and unremitting effort than even his great spirit and energy could sustain.
>
> He was devoted to music, especially Beethoven, whose birthday he shared. He had studied concert piano as a child and when he found time he still played, as he did physics: fortissimo. Intensely preoccupied with work, he made few casual friends. Instead he leaves behind many who admired him greatly and a select group of loyal, close associates who will pursue his dreams. For them, there can never be another Nick."

6 ACKNOWLEDGEMENTS

I sincerely thank the organizers of the school for this opportunity to review for a younger generation of accelerator physicists the work of Nicholas Christofilos. I am heavily indebted to the many persons who shared with me their recollections of Nick and of the early days of strong focussing; in particular to J. P. Blewett, T. Fessenden, K. T. Fowler, S. Humphries Jr., D. L. Judd, and A. B. Langdon. The kind cooperation of Mr. Stephen Wofford of Lawrence Livermore Laboratory in providing the archival material and the photos of Nick Christofilos, is greatly appreciated.

7 REFERENCES

[1] Nicholas Christofilos, U.S Patent 2,736,799 "Focussing system for ions and electrons".

[2] E. Courant, M.S. Livingston and H. Snyder, Phys. Rev 88, 1190 (1952).

[3] N.C. Christofilos, " Method of computation of drift tube shapes", CERN Symposium on High Energy Accelerators and Pion Physics p.176 (1956).

[4] N.C. Christofilos, Proc. Second U.N. International Conference on Peaceful Uses of Atomic Energy, Geneva 32, 279 (1958).

[5] N. C. Christofilos, Nuclear Fusion 1962 Supplement Part I p.159.

[6] See for instance J.D. Jackson "Classical Electrodynamics", John Wiley and Sons, New York (1975) p.588.

[7] N.C. Christofilos et al., Rev. Scient. Instr. 35, 886 (1964).

[8] See for instance S. Humphries, Jr., "Principles of Charged Particle Accelerators", John Wiley and Sons, New York (1986). See also C. Kapetanakos and P. Sprangle, Physics Today, February 1985 p.58.

[9] N.C. Christofilos, Proceedings of the National Academy of Sciences 45, 1144 (1959) and related articles.

[10] See for instance J.D. Jackson, "Classical Electrodynamics", John Wiley and Sons, New York (1975) p. 360.

[11] N.C. Christofilos, "Ocean 1972, IEEE International Conference on Engineering in the Ocean Environment" p. 145. This is a report dated August 29, 1960 and declassified on June 29, 1972. See also related articles.

[12] Physics Today, January 1973 p. 109. Obituary of Nicholas Christofilos by J.S. Foster, T K. Fowler and F.E. Mills.

MEDICAL APPLICATION OF ACCELERATORS IN TUMOR THERAPY

G. Kraft
Biophysik, GSI Darmstadt, Germany

Abstract
Heavy charged particles are the most advanced tool of an external subcutaneous radiotherapy of deep seated tumors. Small angular- and lateral-scattering and the increase of the energy deposition with penetration depth are the physical basis for a more efficient tumor targeting. High biological efficiency in the tumor is the prerequisite for a successful treatment of tumors resistant to sparsely ionizing radiation. The possibility to perform target-conform irradiation and to control the achieved actual distribution using PET techniques guarantees that highly biological-efficient particle stopping can be restricted to the tumor volume only. Although the physical and radiobiological properties of ion beams are very favorable for therapy, the necessity to produce these particles in an accelerator has restricted a general application of heavy ions up to now. Particle therapy could be applied at a dedicated medical machine in the cost range of conventional cancer therapy in those cases where surgery and conventional therapy cannot be applied. Presently many particle therapy units are in preparation or starting operation. However, most of these activities are focused on proton therapy. The use of heavy ions such as carbon or oxygen is still restricted to a few locations, for example Chiba (Japan) or Darmstadt (Germany) where heavy-ion therapy units are in preparation.

1. INTRODUCTION

Every year one million humans in the European Union (EU) are hit by cancer. Approximately 52% of these suffer from non-localized tumors, the remainder having a local tumor. Because of improved diagnosis and irradiation schedules, nearly one half of all tumor patients have a chance to be cured or at least to survive the next five years. To the survival rate of 45%, surgery contributes with 22%, radiation treatment with 12%, and the combination of both with another 6%. But this cure rate is concentrated on the localized tumors and only 5% can be cured by chemotherapy acting predominantly on the non-localized disease. Improvements in cancer treatment are urgently needed because of the large number of patients that cannot be helped so far. The development of gene therapy is a great hope for the large group of patients suffering from non-localized disease but also to some extent for patients having a solid tumor. Gene therapy could also promote a better understanding of cancer in general and its cure. However, our present knowledge is far from that point and all therapy methods have to be examined to see whether they can contribute to a better cure rate in the near future.

For radiation therapy, the better localization of the radiation dose applied to the target volume yielded the major progress over the last decades. This was possible because of a better definition of the target volume using improved diagnostics such as Computer Tomography CT, Magnetic Resonance Imaging MRI and Positron Emission Tomography PET. In the treatment itself, the transition from low-energy X-ray to Co-gamma rays and finally to Röntgen-Bremsstrahlung of highly-energetic electrons allowed a more target-conform treatment. Computerized treatment planning, and in consequence, more sophisticated treatment schedules using nearly 4π geometry increased the concentration of the radiation dose to the target nearly to the maximum value theoretically possible for electromagnetic radiation. However, the limits of

this selection process are clearly given by the physical properties of the electromagnetic radiation where the energy deposition decreases exponentially with the penetration depth, at least for penetration depths of more than a few centimeters. Therefore, for a deep-seated tumor the dose to the healthy tissue in front of and behind the tumor is normally the limiting factor for the tumor dose. Higher tumor doses that would lead to a better tumor control involve higher complication rates that are not tolerable.

Further progress in external radiation therapy can be obtained by using different types of radiation. In the last years two concepts have been pursued: First the improvement of the physical dose distribution, or physical selectivity using particle radiation like pions, protons or heavier ions [1], secondly the improvement of the biological efficiency using neutrons [2]. Pion and neutron trials have been started with great enthusiasm. In the case of pions an increase in the physical selectivity, and also an increase in the biological efficiency, by a factor of three or more was expected in the stopping region of the pions where the nuclear "star" reaction takes place. In the clinical trials it turned out that only a 20% increase of biological efficiency was found. The better physical selectivity yielded better clinical results but there were great problems to produce high pion fluences. Even using high-intensity proton beams and a very sophisticated pion delivery system like the superconducting Piotron at Villigen, the overall treatment time is too long. For one patient a total beam time of nearly one day is required. This time is split into 20 or more fractions but even then only a few patients can be treated per month. Therefore pions are too expensive for a routine therapy, and at most sites pion therapy has been terminated. A similar situation is found for the application of neutrons. The first trials at the Hammersmith Hospital, London, yielded very good initial results with excellent tumor control, that encouraged others to make further neutron trials. But the early success of neutron therapy was spoiled by late effects that were so severe that most of the neutron trials have now been terminated. Presently, it is believed that only a few radioresistant tumors can successfully be treated with neutrons.

In contrast, heavy charged particles have proven to be very successful. Up to now approximately 15000 patients have been treated mainly with protons, and 500 patients with heavier ions, mainly neon. In both cases the clinical results are superior to the conventional therapy. This is especially true for the treatment of melanomas in the eye but also for the treatment of many other inoperable tumors in the brain or head and neck region. The great success especially of the Harvard proton therapy has stimulated the first privately operated proton therapy at the Loma Linda University Medical Center. Other proton therapy projects in Europe are in preparation. In Japan the Heavy Ion Medical Accelerator at Chiba (HIMAC) will start in 1994 with patient treatment using heavier ions like carbon and neon.

For physical as well as for biological reasons, the heavier ions should yield better clinical results than protons [3]. These differences have so far not been confirmed by clinical experience. Theoretically an enhanced cell killing should support the therapeutic effect in the target region because of the increased ionization density of heavy ions at the end of the particle range. However, it has not yet been possible to restrict the region of the stopping ions exactly to the target volume only. Because of the limitation of passive beam shaping methods that have been applied in all particle therapy units to date, a strict tumor conform treatment could not be realized and highly efficient stopping ions are distributed to the tumor surroundings. Therefore, particle therapy using heavy ions like carbon or oxygen is still at the stage of radiobiological research and technical development, while proton therapy, not confronted with changes in biological efficiency, is more a subject of clinical research.

In the following section the physical and radiobiological advantages of particle beams for therapy will be presented and the role of active beam delivery systems using fast magnetic scanning will be explained.

2. PHYSICAL AND RADIOBIOLOGICAL ADVANTAGES OF PARTICLE BEAMS

Heavy charged particles exhibit a different interaction mechanism with matter compared to electromagnetic radiation like X- or gamma-rays or bremsstrahlung. For electromagnetic radiation the primary intensity of the beam decreases exponentially with the penetration depth and the beam has no finite range. In addition, the lateral scattering of the primary beam due to Compton scattering and the scattering of the secondary electrons is large.

Swift heavy charged particles when passing through matter dissipate their energy mainly by direct interactions with the electrons of the target material. Because of the large difference between the electron mass and the ion mass, the deflection of the projectile ions is very small and only multiple collision processes can cause a net deflection of the particle beam. In Fig. 1 the lateral beam deflection by multiple scattering in water is compared for different ions having the same range. From protons to helium ions the angular scattering of the primary beam is reduced by a factor of two and decreases again from helium to carbon. A further reduction of the angular deflection of heavier ions is beyond the precision of clinical diagnosis and represents no further clinical improvement.

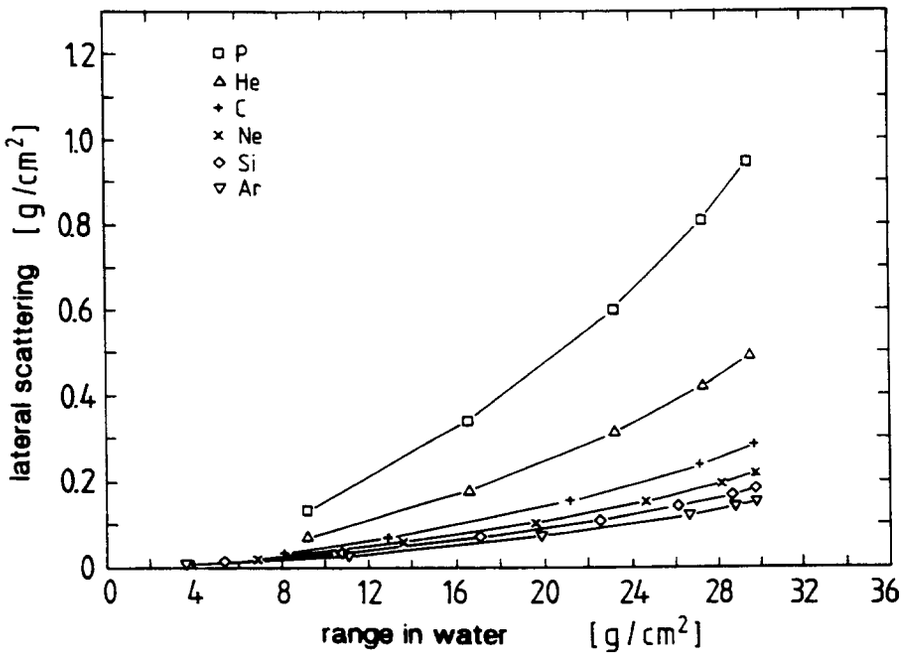


Fig. 1. Lateral beam deflection as a function of penetration depth in H_20 for different ions. For carbon and heavier ions a precision of 1 mm can be reached for tumor depths of 15 cm.

Because the energy dissipating electronic collisions are very frequent, the energy loss of the projectile is nearly continuous. The energy deposition increases with decreasing velocities. For larger penetration depth it culminates in a sharp maximum, the so-called Bragg peak, just before the end of the particle range [3]. Beyond the Bragg peak the energy deposition decreases rapidly giving a finite range with a sharp cut off at the end of the particle range. In consequence the dose profile of ions is inverted compared to electromagnetic radiation. Both properties, the small scattering and the inverted dose profile are major advantages for the application of heavy charged particles in radiotherapy. In Fig. 2 depth/dose profiles normalized to the entrance dose are compared for beams with electromagnetic radiation and neutrons [4].

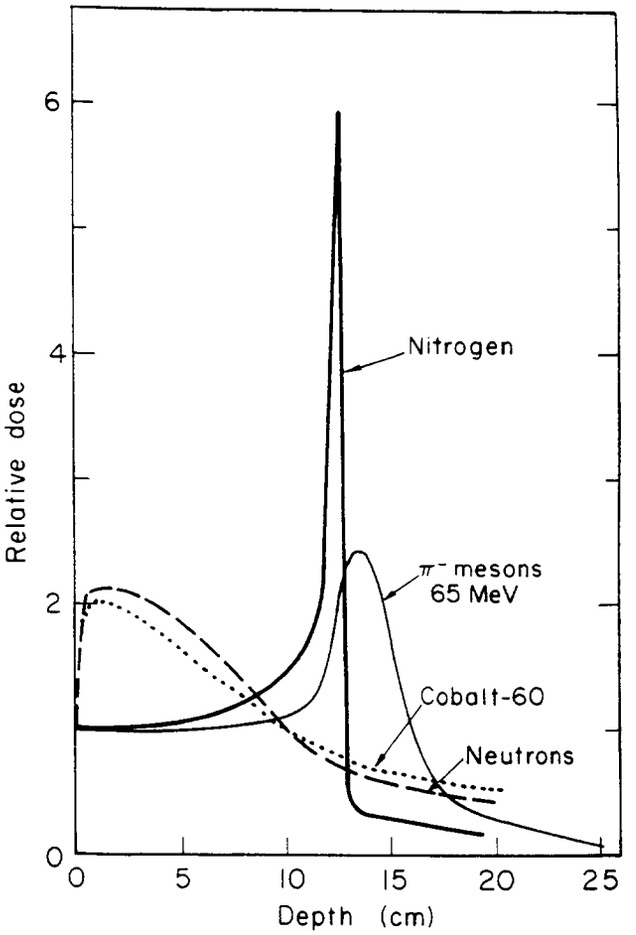


Fig. 2 Comparison of depth/dose distributions of Co-gamma with neutrons, π-mesons and nitrogen ions. The charged particles exhibit an inverted dose profile compared to X-rays and neutrons. This allows a higher tumor dose.

The increase of dose with increasing penetration depth is characteristic for particles like π-mesons and nitrogen. π-meson therapy has not been very successful because of the low pion intensities and, in consequence, the limited number of patients that can be treated. Protons or heavier ions can be produced with high intensity directly by accelerators.

For practical applications, the sharp Bragg peak has to be adapted to the tumor size, varying the penetration depth by energy variation. In Fig. 3 a superposition of various Bragg peaks corresponding to different particle energies is shown to produce a flat dose profile over the depth of the target volume. This figure also illustrates that in general the dose at a given point of the target volume is produced by ions of different energies and, due to nuclear fragmentation also by ions of different atomic numbers [5]. For the use of protons this is of minor importance because fragmentation of the primary beam does not exist and the regime of elevated biological efficiency is small. However, for the heavier ions the variation of the biological efficiency depends strongly on atomic number and energy. Both parameters have to be known precisely at each target point in order to determine the biological effect at any location of the superimposed Bragg peaks.

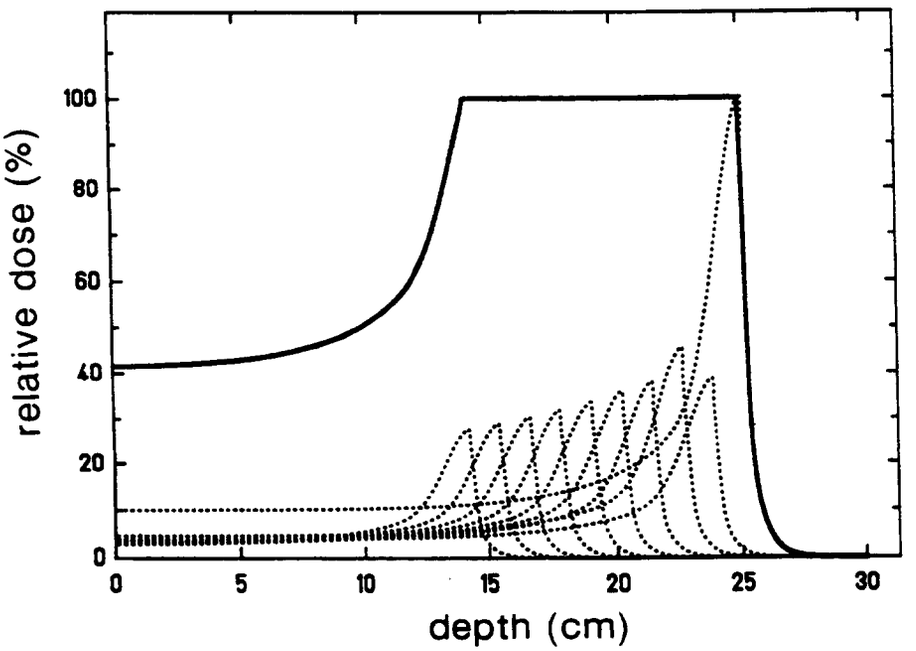


Fig. 3 Superposition of several Bragg maxima produces a flat top over the target area

3. HIGH-LET EFFECTS

An increased biological efficiency is observed in the region of elevated energy loss or linear energy transfer (LET) [6]. The different action of high-LET radiation, like heavy ions, compared to low-LET radiation, like X-rays, is characterized by the relative biological efficiency (RBE) which is defined as the ratio of X-ray to particle dose necessary to produce the same biological effect. LET values of 100 keV/μm or greater are obtained for the lighter ions like oxygen in the Bragg peak only. For protons the region of high-LET in the Bragg peak is too small to be of clinical relevance. For ions heavier than neon significantly elevated RBE values already occur in the entrance channel and saturate in the Bragg maximum.

The reason for an increased biological efficiency can be qualitatively explained by comparing the local ionization density produced by the electrons around the particle track. In Fig. 4 the calculated track structure of protons and carbon ions at 1 MeV/u are compared with the size of a DNA molecule as the critical target inside the cell nucleus [7]. Obviously a larger amount of DNA double-strand breaks can be produced by the traversal of a single carbon ion than by protons or other low-LET radiation. For low-LET radiation, the electrons act mostly independently producing isolated DNA lesions like single-strand breaks. Only in a few cases do these single-strand breaks interact, forming a biologically more relevant double-strand break [8]. Using heavy ions, the a priori probability for the production of double-strand breaks is increased by the high local density of the electrons in a heavy particle track. This yields an increase of the relative biological efficiency up to RBE values of 3–5 in the Bragg maximum of the lighter ions. However, in the case of heavier ions a further increase in local damage cannot be expected to produce a greater biological effect because of physical, chemical and biological saturation effects ("overkill").

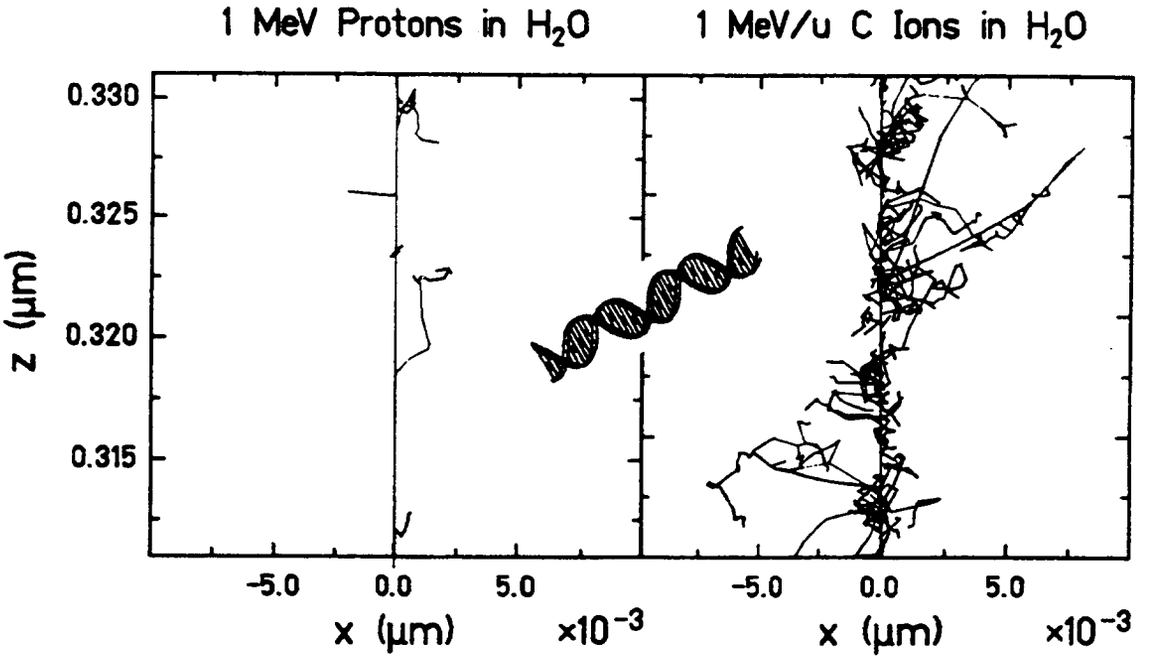


Fig. 4 Comparison of proton and carbon track structure, both at 1 MeV/u, with the dimension of a schematically shown DNA molecule [7]. Carbon ions have a much higher probability to simultaneously produce breaks at two opposite strands.

For radiotherapy the optimal ions are those for which biological high-LET action can be restricted to the tumor volume while in the healthy tissue in front of the tumor mainly repairable damage is produced. In Fig. 5 the relative biological efficiency of various ions is plotted as a function of the penetration depth. For the heavier ions such as neon, silicon or argon the RBE is already greater than 1 in the entrance channel. Therefore, these ions are candidates for the treatment of superficial tumors only. For deep seated tumors, beams of carbon or oxygen are more suited because they produce less damage to the healthy tissue in the entrance channel.

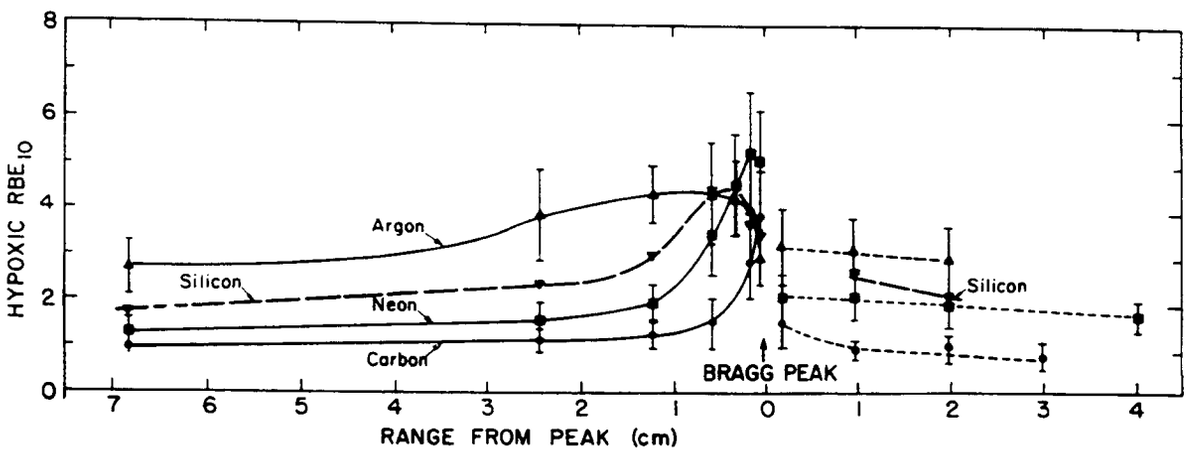
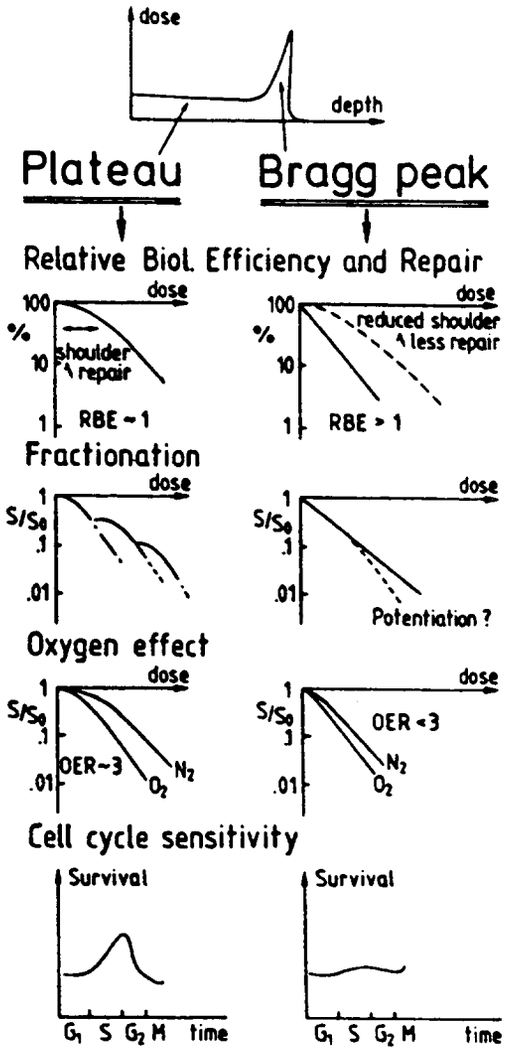


Fig. 5 Relative biological efficiency of various heavy-ion beams as a function of the penetration depth in water [6]. Light ions such as carbon have RBE levels close to 1 in the entrance channel and more elevated only in the tumor volume.

Up to now, the increased biological efficiency was treated simply as a factor independent from the cell type and biological reaction under consideration. However, this is a very raw approximation. The increase in RBE is caused by a reduced repair capacity and affects many cellular reactions which depend on repair, for instance the response to fractionated exposure and at different phases of the cell cycle, and probably most important the response of slowly proliferating cells having an immense repair capacity against sparsely-ionizing radiation.

Some of the critical issues for a comparison of plateau and Bragg maximum are summarized in Table 1. It is expected that the properties of heavy charged particles can be used to restrict the damage to radioresistant (against sparsely-ionizing radiation) and deep seated tumors. This concept infers that the high-LET part of radiation, i.e. the stopping ions can be restricted to the target volume only.

Table 1
Comparison of the various biological reactions in the entrance channel (plateau)and the Bragg maximum



Repair and fractionation

In the low LET region, cell survival is characterized by high repair capacity corresponding to an extensive shoulder in the survival curve. The high LET regime exhibits reduced repair and an exponential or nearly exponential survival curve.

In the high LET regime, a drastically diminished fractionation effect has been observed. In a few cases a potentiation of the radiation damage has been found for fractionated treatment. At the present time, it is not clear whether the potentiation effect is useful for therapy. However, diminshed repair in the tumor volume combined with full repair in the entrance channel will be of clinical relevance.

Oxygen effect

At low LET values, the radiation sensitivity of hypoxic cells can be reduced by a factor of three compared to oxic cells. For high LET radiation, these differences are decreased and for LET values around 100 keV/μm both types of cells, oxic and hypoxic, exhibit the same radiation sensitivity.

Cell cycle sensitivity

In vitro cell cultures show differences in radiosensitivity of more than one order of magnitude when exposed to sparsely ionizing radiation in different phases of the cell cycle. This is true especially for cells in late S phase which are radioresistant in comparison to G_1 phase cells.

In the range of high energy transfer, the differences in radiosensitivity between the different phases of the cell cycle are reduced and cells with a higher DNA content like G_2 cells, are more radiosensitive.

These changes in the radiobiological action of heavy Ions all act in the same direction: In the tumor volume cell killing is enhanced while in the healthy tissue around the tumor and in the tumor bed the cells are spared to a greater extend.

4. ON-LINE CONTROL OF THE BEAM BY PET TECHNIQUES

A precise beam delivery (see next paragraph) needs precise control. Heavy ions can be monitored using the production of radioactive isotopes that decay by positron, β^+ emission. The β^+ annihilation can be localized by the coincident detection of the two 511-keV gamma quanta in two adjacent gamma cameras. Because the positrons are emitted isotropically with respect to the beam direction, the center of the observed distribution of positron emission corresponds to the stopping point of the radioactive isotope.

Originally it was proposed for PET techniques to produce a radioactive β^+ beam instead of a stable beam and to measure its distribution in order to verify the treatment planning prior to the therapy. This has been done successfully at Berkeley in cases where the treatment volume was close to a critical organ like the spinal cord [9].

Meanwhile it has been demonstrated in phantom measurements that β^+ radioactivity produced by a stable beam is sufficient to monitor the stopping points of the projectile ions [10]. Beams of stable ions undergo nuclear reaction i.e. projectile and target fragmentation. Most likely are stripping reactions leading to lighter ions, a fraction of which are β^+ isotopes. In the case of target fragmentation these β^+ emitters are smeared out over the particle path. For the projectile fragmentation the β^+ isotopes have nearly the same range due to their similar mass-to-charge ratio and the strongly forward-peaked reaction kinematics. In Fig. 6 the induced β^+ activity of a ^{20}Ne beam directed into a plastic block is compared with the calculated range distribution and with the range of the primary ^{20}Ne. From the β^+ distribution the range of the ^{20}Ne can be calculated with high accuracy. It should be mentioned that measurements of the β^+ activity are not possible during the beam spill due to a large background. However, the off-spill time intervals are sufficient for PET measurements.

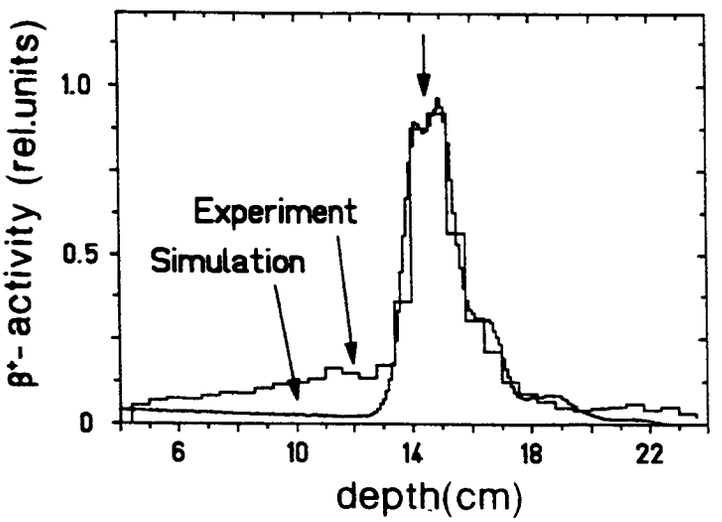


Fig. 6 Measured and calculated profiles of β^+ emission are compared with the range to the primary ^{20}Ne beam indicated by an arrow [10]

Figure 6 also illustrates another problem of dose localization for the heavier ions: with increasing atomic number more and more lighter fragments are produced and which have a longer range than the primary beam so causing a long dose tail behind the Bragg maximum. Therefore, the excellent dose distribution of the heavy charged particles deteriorates for ions heavier than neon. According to these findings atomic numbers around that of oxygen seem to be optimal for therapeutical application of ion beams for both physical and biological reasons.

5. IMPROVED DOSE DELIVERY BY MAGNETIC SCANNING AND FAST ENERGY VARIATION

In all particle therapies presently in operation, the beam is distributed over the target volume using passive beam-forming elements like degraders, range shifters, boli and fixed or movable collimators. These tools are expensive and inadequate compared to the properties of a beam of charged particles which in principle can be controlled by magnetic fields. They are more or less a straightforward translation of the techniques conventionally used in X-ray radiotherapy.

In order to exploit the properties of a particle beam, magnetic scanning [11] is now introduced into particle therapy (Fig. 7). In this technique the target volume is divided into several slices of equal particle range and each slice is treated independently by scanning the beam over it. If one slice is treated, the beam energy and therefore the particle range is reduced and the next slice is treated in the same manner.

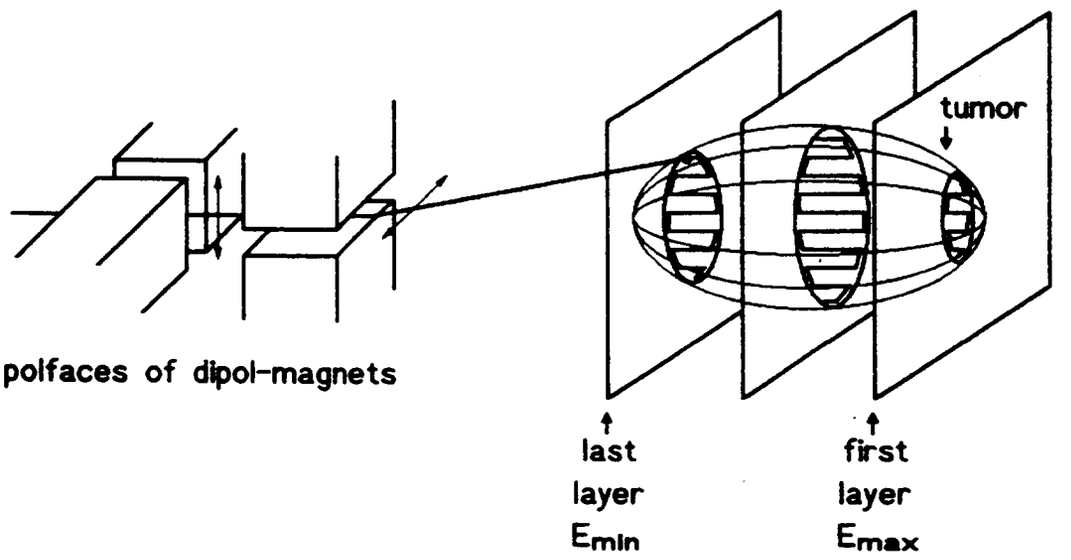


Fig. 7 Schematic view of a magnetic scanning system

Depending on the geometry of the target volume, the proximal parts of the tumor will have been irradiated when the distal parts were treated . Therefore, the additional particle covering has to be distributed inhomogeneously over the slice in order to achieve a homogeneous biological effect over the whole treatment volume. Two strategies have been proposed to produce inhomogenous particle covering: pixel scanning and raster scanning.

In pixel scanning the area of a slice is divided into single picture elements, i.e. pixels. The coordinates for each pixel are chosen such that the beam spots of neighboring pixels overlap. For the irradiation procedure the scanning magnets are set to the coordinates of the first pixel and the beam is turned on. The beam is switched off when the required number of particles is reached. Then the setting of the magnets is moved to the next pixel, the beam is turned on and the preselected number of particles is delivered, and so on. For raster scanning, the beam is moved continuously in a raster-like pattern over each slice and the writing velocity is controlled by the beam intensity and the required particle covering. Extensive theoretical computer simulations have demonstrated that with both methods an optimal particle distribution

can be produced [12]. However, the differences between raster and pixel scanning seem to be academic when it comes to the construction of a magnetic scanning system.

In order to use a real particle beam from an expensive accelerator, and to reduce the time the patient is immobilized in a very inconvenient manner, exposure time has to be short and both scanning methods have to be optimized. For the pixel scanning, this means that a minimal path from pixel to pixel has to be selected which is obviously a path from neighboring pixels to each other. For technical reasons with raster scanning the path of the continuously moved beam has to be divided into a large number of individual coordinates which can be stored in a computer so that the beam can be moved in small steps from each coordinate to the next. The basic difference between raster and pixel scanning finally reduces to whether it is necessary to turn off the beam while moving from one point to the next.

In reality the beam has a finite diameter of several millimeters. If the distance between two pixels is small compared to this diameter, the beam need not be interrupted while moving from pixel to pixel. Because pixels are dense, slight over- or under-exposure of one pixel, and the exposure in the distance between two pixels, can be corrected to a large extent during the irradiation of the next pixel. A digitized version of the raster scanning system, i.e. a hybrid of raster and pixel scanning has been constructed at the biophysics cave at GSI [11]. In this set up the beam is guided in the x and y direction by two magnets while the beam intensity is measured by means of a transmission counter. The beam path of one slice is divided into 16000 picture points and the writing velocity (maximum 1 cm/msec) of the beam is controlled by a secondary electron monitor employed as a transmission counter. The use of a transmission counter to control the writing velocity has been demonstrated to be a very elegant and efficient method to cope also with the intensity fluctuations of the beam as extracted from the accelerator. First experiments demonstrated that particle coverings with a homogeneity of 5% or better can be produced. But in the same way any inhomogenous distribution can be produced with the same accuracy if required. In order to fill a 3-dimensional volume the beam energy has to be varied. In a modern synchrotron this can be done from pulse to pulse by the machine. It is also possible to tune the beam line synchronously to achieve a stable beam spot at the target position. In Fig. 8 experimental results using fast energy variation by the accelerator and fast tuning of the beam line are shown. These experiments demonstrate that 3-dimensional beam delivery is possible using fast energy variation by the accelerator and fast beam line tuning by computer control. Meanwhile 3D volumes of simple shapes have been exposed to an isodose (Fig. 9).

6. TUMOR-CONFORM TREATMENT PLANNING USING HEAVY-ION BEAM SCANNING

The novel technique of fast magnetic scanning combined with an active energy variation by the accelerator opens a new area of target-conform tumor treatment. However, the progress in tumor targeting necessitates a quantum leap in the complexity of the beam control and treatment planning.

In conventional and neutron therapy the beam has the same biological efficiency regardless of its penetration depth. Using suitable range-dose relationships, the tumor dose has to be maximized and the integral dose to the healthy tissue minimized either by using multiport irradiation or cross-fire techniques. For proton treatment planning, the depth/dose relationship has to be changed according to the increase of the energy deposition with increasing penetration depth. The small region of elevated biological efficiency in the Bragg peak is taken into account by changing the overall RBE from 1 to 1.1.

Heavy ions like carbon or oxygen have an extended region of significantly increased biological efficiency in the Bragg region. Using magnetic beam scanning this region of high-LET action will be restricted to the target volume only. But as shown in Fig. 3 the dose in the various points of the target volume is a mixture of doses originating from plateau and from

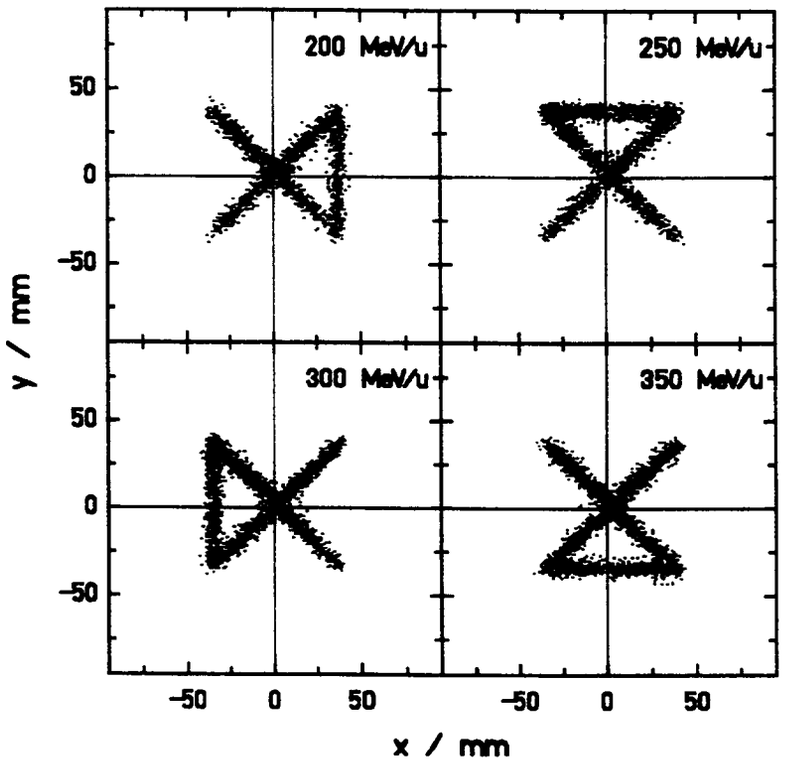


Fig. 8 Combining beam scanning and energy variation one can produce simple figures [11]. The energy in the accelerator and the settings of the beam line were switched within 3 seconds

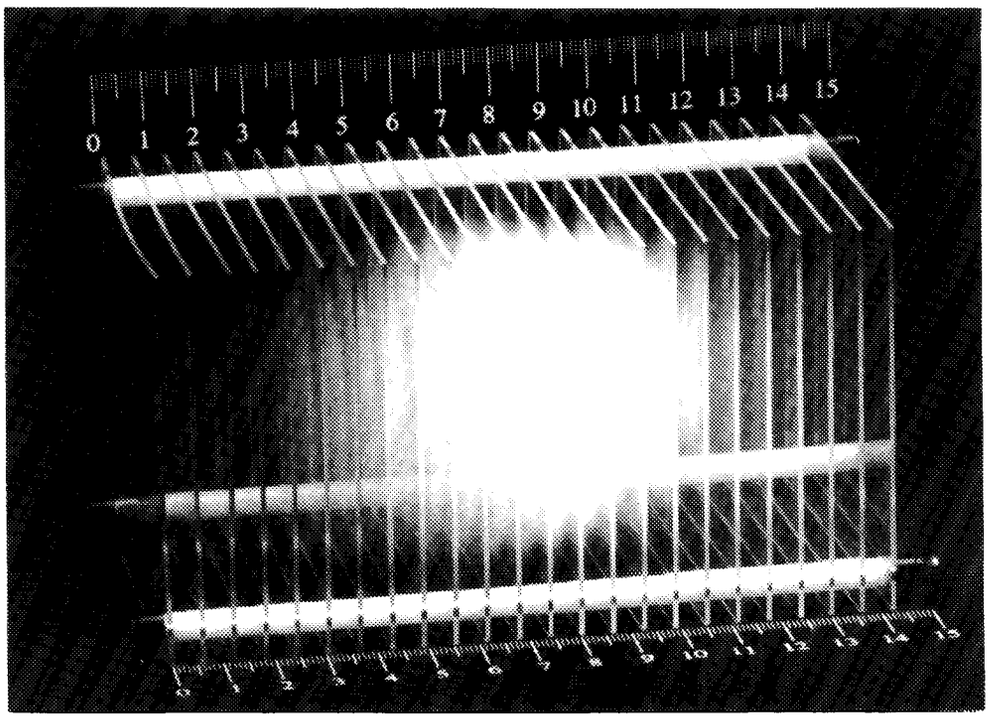


Fig. 9 Recording of a 6 cm diameter spherical isodose volume produced by raster scanning in a depth between 9 and 15 cm in a water phantom. The 270 MeV/u carbon beam entered the water tank from the left. The volume was dissected into 30 layers of equal particle range that were irradiated separately [16].

Bragg peak ions. Because all these ions have completely different RBE values, the biological effect depends critically on the composition of the beam in atomic number and energy [5]. Therefore, the physical properties of the beam such as energy distribution and lateral scattering or nuclear fragmentation have to be known precisely in order to predict the composition of the beam at each target point. In addition, the dependence of the biological effect has to be known for the different energies and particles [8]. Because it is not possible to measure all the necessary data, a model of the biological action of the particles has been developed, which makes it possible to calculate these data [14]. The deeper physical and biological understanding of the properties of heavy charged particles is the prerequisite for using focused particle beams in a similar way to a surgeon using his knife.

7. FUTURE ASPECTS OF PARTICLE THERAPY IN EUROPE

The use of heavy particles is the last possible step in improvement in physical selectivity and biological efficiency that can be reached for external radiotherapy. Up to now, accelerators originally designed for nuclear research have been used in the explorative phase of particle therapy. In consequence the advantages of particle therapy could be exploited only to a limited extent. Tumor-conform treatment could not be achieved at these units. Furthermore, the treatment at these machines is not cost effective. For the consideration of particle therapy as a routine treatment in Europe, the costs at a medically-dedicated machine have to be compared with the costs and the success of conventional radiotherapy. However, it must be stressed that, it is not the purpose of particle therapy to replace the existing efficient treatments like surgery or conventional radiation therapy. Particle therapy should be used in those cases where conventional treatment is not successful.

Conservative estimations performed within the Eulima project revealed that in Europe some 100000 patients a year could benefit from particle treatment. In a big radiation therapy clinic 2000 patients can be treated every year so that 50 therapy centers spread around Europe seems to be a realistic estimate for the future development. But it is not clear how this number should be distributed between proton and heavy-ion centers. For cost efficiency, the Eulima project showed that particle treatment can be performed in the cost range of conventional therapy (see Table 2). However a patient that is not cured from cancer is much more costly and from an ethical point of view it is not tolerable not to cure a patient who could be helped.

Table 2
Comparison of the average costs of the various cancer treatments

	Cost (kECU)
Conventional radiotherapy	4
Surgery	7.5
Chemotherapy	30
Heavy ion treatment	7.5
Not cured patient	~35

The economic efficiency of heavy-ion therapy depends on the extent to which the theoretically possible advantages of heavy particles can be verified in the clinical routine. The presently existing particle therapy units as well as the proposed machines are given in Tables 3 and 4, respectively. In addition to the optimized treatment, the absolute investment costs play an important role for the development of particle therapy. Presently, 10-12 MECU are realistic estimates for a proton machine including a simple beam delivery system. If this price could be lowered further, the addition of proton machines to existing therapy centers would be more likely than separate installation of dedicated particle-therapy centers. In consequence, it is very important to explore the potential of particle therapy at the existing accelerator centers and to prepare for a widespread therapeutical use of particle accelerators in Europe.

Table 3
World-wide totals (Jan 1994) of patients treated with charged particles according to Ref. [15]

Who	Where	What	Date first RX	Date last RX	Recent Patient Total	Date of Total
Berkeley 184	CA. USA	p	1954	1957	30	
Berkeley	CA. USA	He	1957	1992	2054	Jun-91
Uppsala	Sweden	p	1957	1976	73	
Harvard	MA. USA	p	1961		6010	Dec-93
Dubna	Russia	p	1964	1974	84	
Moscow	Russia	p	1969		2550	Oct-92
St.Petersberg	Russia	p	1975		719	Jun-91
Berkeley	CA. USA	heavy ion	1975	1992	433	Jun-91
Chiba	Japan	p	1979		86	Jun-93
PMRC,Tsukuba	Japan	p	1983		354	Sep-93
PSI (SIN)	Switzerland	p	1984		1363	May-93
Dubna	Russia	p	1987		24	Aug-92
Uppsala	Sweden	p	1989		34	May-93
Clatterbridge	England	p	1989		463	Jan-94
Loma Linda	CA. USA	p	1990		682	Dec-93
Louvain-la-Neuve	Belgium	p	1991		21	Nov-93
Nice	France	p	1991		216	Apr-93
Orsay	France	p	1991		235	May-93
N.A.C	South Africa	p	1993		6	Nov-93
Indiana Cycl.	IN. USA	p	1993		1	Dec-93
					2487	ion beams
					12951	proton beams

Table 4
Proposed new facilities for proton and ion-beam therapy, from Ref. [15].

Institution	Place	Type	Date 1st RX
PSI	Switzerland	p	1994
HIMAC,Chiba	Japan	ion	1994
TRIUMF	Canada	p	1994
Berlin	Germany	p	1995
Munich	Germany	p	1995?
Novosibirsk	Russia	p	1995?
Proton Development N.A.	IL. USA	p	1996
GSI Darmstadt	Germany	ion	1996
ITEP Moscow	Russia	p	1996
Jülich (KFA)	Germany	p	1997
KVI Groningen	Netherlands	p	1997?
NPTC (Harvard)	MA. USA	p	1998
TERA	Italy	ion	?
Clatterbridge	England	p	?
Tsukuba	Japan	p	?
Krakow	Poland	p	?

REFERENCES

[1] G. Kraft, Radiobiological and physical basis for radiotherapy with protons and heavier ions, Strahlenther, Onkol, 166 (1990) 10-13.

[2] A. Wambersie, G.W. Barendsen, N. Breteau, Overview and prospects of the application of fast neutrons in cancer therapy, European Journal of Radiation Oncology 5 (1984) 248-264.

[3] G. Kraft, M. Kramer, M. Scholz, LET, track structure and models, Radiat. Environ. Biophys., 31 (1992) 161-180.

[4] H. Jung, Biologische Wirkung dicht ionisierender Teilchenstrahlen. In: L. Diethelm, F. Heuck, O. Olsson, F. Strand, H. Vieten, A. Zuppinger (Hrsg.), Handbuch der Medizinischen Radiologie Bd XX, Springer-Verlag Berlin, Heidelberg, New York, Tokyo (1985).

[5] J. Llacer, J.B. Schmidt, C.A. Tobias, Characterization of fragmented heavy-ion beams using a three-stage telescope detector: measurements of 670 MeV/amu ^{20}Ne beams, Med. Phys. 17 (1990) 151.

[6] E.A. Blakely, C.A. Tobias, F.Q.H. Ngo, S.B. Curtis, Physical and cellular radiobiological properties of heavy ions in relation to cancer therapy applications, LBL Report 11220 (1980).

[7] M. Kramer, G. Kraft, Heavy ion track structure calculations, in: Biophysical modelling of radial. effects, K.H. Chadwick, G. Moschini, M.N. Varma (eds.), Bristol, Philadelphia, New York (1992) 361-368.

[8] G. Kraft, Radiobiological effects of very heavy ions: Inactivation, induction of chromosome aberrations and strand breaks, Nucl. Sci. App. 3 (1987) 1-28.

[9] J. Llacer, Positron emission medical measurements with accelerated radioactive ion beams, Nucl. Sci. Appl. 3 (1988) 111.

[10] W. Enghardt, W.D. Fromm, H. Geissel, H. Keller, G. Kraft, A. Magel, P. Manfraβ, G. Münzenberg, F. Nickel, J. Pawelke, D. Schardt, C. Scheidenberger, M. Sobiella, The spatial distribution of positron-emitting nuclei generated by relativistic light ions, Phys. Med. Biol. 37 (1992) 2127.

[11] Th. Haberer, W. Becher, D. Schardt, G. Kraft, Magnetic scanning system for heavy-ion therapy, NIMRD A330 (1993) 291-305.

[12] Biophysics Group GSI, Design, construction and first experiments of a magnetic scanning system for therapy, GSI-Report 91-18 (1991).

[13] G. Kraft, G. Gademann, Einrichtung einer experimentellen Strahlentherapie bei der GSI, GSI-Report 93-23.

[14] M. Scholz, G. Kraft, A parameter-free track structure model for heavy-ion action cross sections. In: Biophysical Modelling of Radiation Effects. Eds. K. Chadwick, G. Moschini, M.V.Varma Adam Hilger 1992 185-192.

[15] J. Sisterson, Particles 13 (1994)

[16] U. Weber et al., GSI Nachrichten 11 (1993) 6-8.