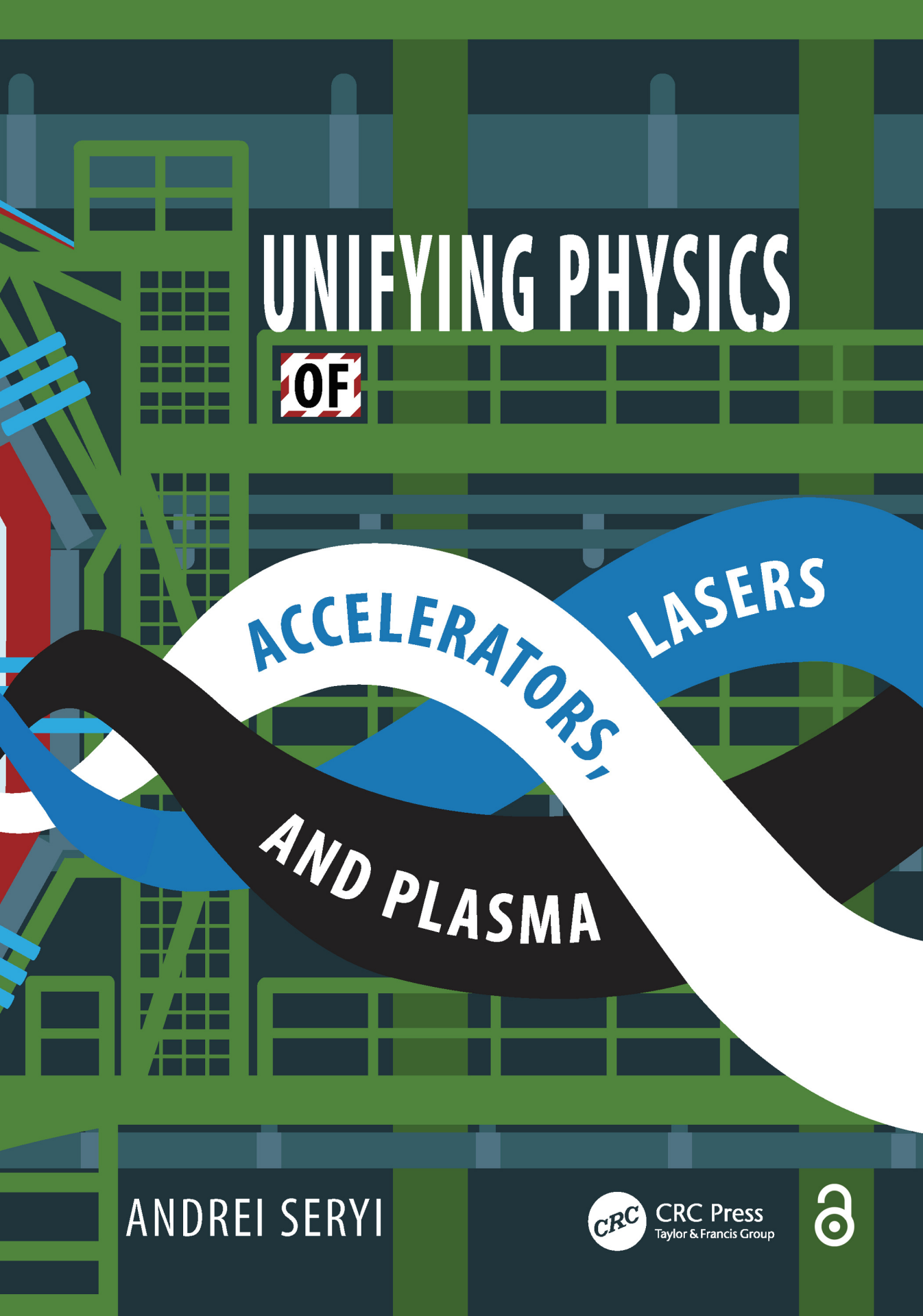
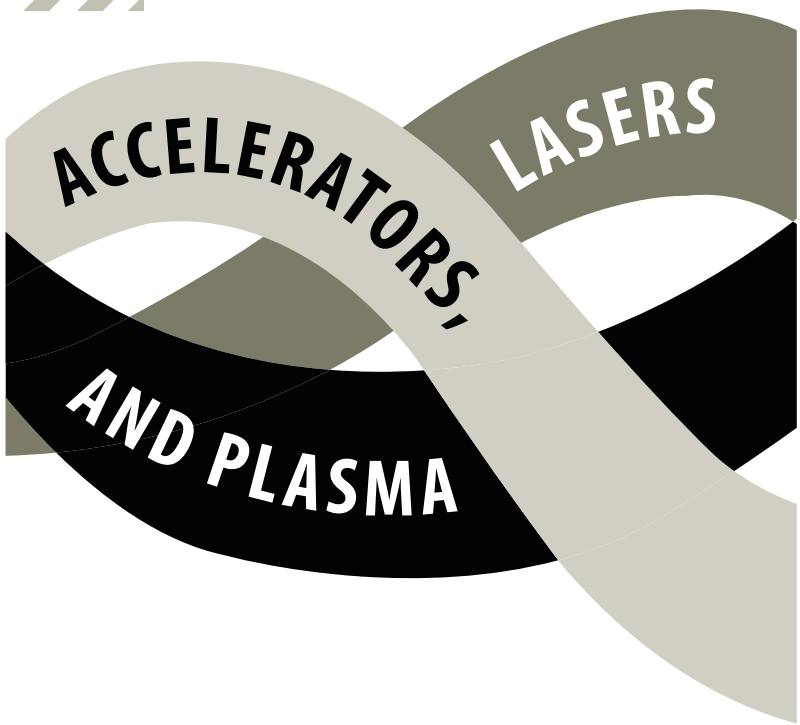



ANDREI SERYI

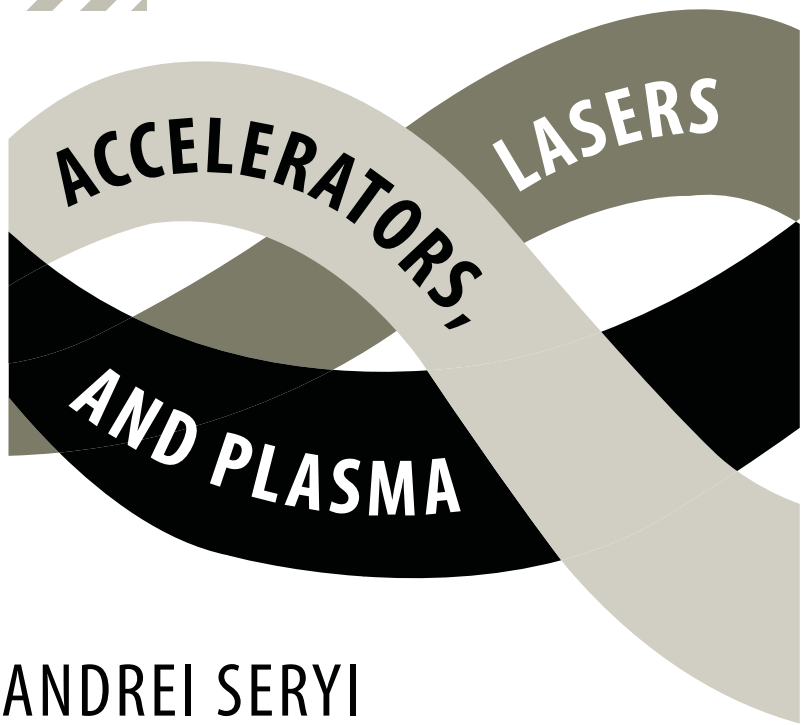
CRC Press
Taylor & Francis Group

# UNIFYING PHYSICS OF



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# UNIFYING PHYSICS OF




ANDREI SERYI

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Sponsoring Consortium for
Open Access Publishing in Particle Physics

*Open Access funded by SCOAP3*



CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

First issued in paperback 2020



ISBN 13: 978-0-367-57562-5 pbk)
ISBN 13: 978-1-4822-4058-0 (hbk)
DOI: 10.1201/b18696



**Library of Congress Cataloging-in-Publication Data**

Seryi, Andrei, author.
Unifying physics of accelerators, lasers and plasma / Andrei Seryi.
pages cm
Includes bibliographical references and index.
ISBN 978-1-4822-4058-0 (alk. paper)
1. Particle accelerators. 2. Lasers. 3. Plasma (Ionized gases) I. Title.

QC787.P3S465 2016
539.7'3--dc23 2015012802

**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**and the CRC Press Web site at**
**http://www.crcpress.com**

To Elena

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Contents in Brief

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Contents

# List of Figures

# List of Tables

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Foreword

Accelerators were invented in the early 1930s and first developed in a significant way by E.O. Lawrence at the University of California, at Berkeley. Initial ideas arose in response to the needs of particle physics expressed by Ernest Rutherford as early as 1924, but right from the outset, Lawrence used his early cyclotron accelerators for both pure research in particle (then called nuclear) physics and more practical applications. Much of the development was funded by their use in the field of medicine for isotope production and therapy. After the World War II years 1939 to 1945, came the invention of the synchrotron — a clever extension of the cyclotron principle to undercut the rising cost of cyclotrons and extend their energy range to allow the production of more massive fundamental particles. Readers who follow the description of the TRIZ inventive process in this book may wonder if this invention was an accidental example of what TRIZ calls the "Russian Doll" technique.

Soon it was realized that electron synchrotrons were a prolific and controllable source of synchrotron radiation, initially in the ultraviolet spectrum and later down to the wavelengths of X-rays. A large number of these SR sources were built and their beams used for scattering experiments, to fathom the structure of new materials for engineering and elucidate the molecular structure of protein and the other complex molecules that have come to dominate our understanding of today's life sciences. To extend the scope of these studies in order to allow single molecules to be reconstructed from their scattering patterns, a completely new concept — the free electron laser — was invented. This has already been used in several countries as a basis for the construction of research facilities that rival the large colliders of particle physics in scale and budget. Finding a new approach to the production of X-rays in this way has been another leap in human imagination, opening the door to new and more powerful applications in a field where those applications bear an almost immediate return on investment. It is the purpose of this book to explore how such leaps in imagination may be stimulated.

To be significant, such inventive processes need to identify a symbiotic relationship between two apparently quite different fields — for the free electron laser these were accelerators and lasers — and then find a common factor that can be used to improve both fields. This technique has been recently developed and refined as a replacement for simple

brainstorming. It is called TRIZ and has already been identified by industry as a means towards more rapid technological progress. We discuss here how it may be used to cut the Gordian Knot[1] of rising costs and complexity, which threaten to impede the development of more powerful instruments for both pure and applied research.

One does not have to look far to find a candidate for a symbiotic technique to pair with today's accelerators. A laser beam penetrating a plasma generates accelerating fields many orders of magnitude greater than today's accelerators. This book for the first time puts these two techniques side by side so that we may identify commonalities of solution and stimulate today's students of accelerator and laser/plasma physics to work together and use TRIZ to solve the problem.

The book's author heads the John Adams Institute, whose participating universities include postgraduate students in both the laser and accelerator fields. We are well placed to kindle the fire of invention amongst them, and we urge those responsible for young and inventive minds elsewhere to join this mission. Our students are already applying our ideas to the Future Circular Collider project of CERN.

Edmund Wilson — Geneva, March 7, 2015

[1] The Gordian Knot is a legend of Phrygian Gordium associated with Alexander the Great. It is often used as a metaphor for disentangling an intractable problem.

# Preface

The aim of this book is to build bridges and connections between three areas of physics that are essential for developing the next generation of accelerators: accelerators, lasers and plasma. These three fields of physics will be introduced in tandem with the industrial methodology of inventiveness, a method which teaches that similar problems and solutions appear again and again in seemingly dissimilar disciplines. This methodology of inventiveness will, ultimately, further enhance connections between the aforementioned fields, and will grant the reader a novel perspective.

The text is suitable for students of various levels between senior undergraduate and graduate physics who are interested in enhancing their ability to work successfully on the development of the next generation of facilities, devices and scientific instruments manufactured from the synergy of accelerators, laser and plasma. I would also recommend this book to anyone interested in scientific innovations.

The idea for the book *Unifying Physics of Accelerators, Lasers and Plasma* came not by accident — it naturally resulted from a search for the best method to suitably train the undergraduate and graduate students at John Adams Institute and its affiliate universities. The aim is to teach several physics disciplines in a coherent way, simultaneously ensuring that this training would develop and stimulate innovativeness.

Materials for this book developed gradually, starting from a single lecture presented at JINR Dubna in March 2014, to several lectures given to Oxford undergraduates in May 2014, and eventually to a full week-long US Particle Accelerator School course given in June 2014. The text, to a large extent, follows the materials developed for this USPAS course. Interaction with the students during this course proved to be a significant inspiration for converting these materials into book form, as well as the lectures, presentations of mini-projects by student teams and tutorial sessions that were focused on analyzing key inventions that shaped the discussed scientific areas.

The style of this book is perhaps different from a typical textbook on accelerator physics, or books on lasers or plasma. Here, we tend to use qualitative discussion and prefer to avoid heavy math, using back-of-the-envelope type derivations and estimations whenever possible (even if they provide approximate formulas). We believe that this is the better

way to convey physics principles to the reader and is certainly preferable when discussing such a broad area of physics.

I would like to express my gratitude to my many colleagues who directly or indirectly helped with this endeavor. Particular thanks go to Professor Riccardo Bartolini, Professor Emmanuel Tsesmelis and Professor Ted Wilson for joint work on the short-option course for Oxford undergraduates, which was the first step toward the development of the US-PAS course that gradually led to development of this book. I am grateful to the students of my USPAS-2014 course for inspiration and enthusiastic discussions and also to Professor Bill Barletta who supported the idea of the USPAS course, discussed the very first time with Bill in early April 2013.

I am grateful to the Rector of Novosibirsk State University, Mikhail Fedoruk, for creating an inspiring opportunity to present a lecture on Science and Inventiveness to a cohort of 1,000 first-year students on September 1, 2014; the preparations for this lecture significantly stimulated the thought process for writing the first chapter of this book.

I am grateful to Professor Zulfikar Najmudin for help with the materials for ion plasma acceleration and thankful to Professor Peter Norreys and Professor Bob Bingham for discussions on the topics connecting chirped pulse amplification and radar inventions.

I am in immense debt to Professor Ted Wilson, who made tremendous efforts to be the first reader of the manuscript and who gave me a lot of valuable comments.

I am grateful to my John Adams Institute colleagues — Professors Riccardo Bartolini, Stewart Boogert, Simon Hooker, Ivan Konoplev, Zulfikar Najmudin and Mike Partridge for reading the manuscript or its individual chapters and giving their valuable comments. I would like to express gratitude to all of my colleagues and friends who helped with the preparation of this book and whom I failed to thank here directly.

I am also delighted to thank Francesca McGowan, CRC Press editor, who on April 30, 2013 appeared out of the blue in my office asking for directions to a meeting — this unexpected conversation helped to crystallize my thoughts toward this book. I appreciate all the aid Francesca and her CRC Press colleagues provided during the writing process.

I would like to express my deep thanks to my family, first of all to my daughter Sasha, who designed the book cover, made several illustrations and helped with grammar.

And most importantly, my eternal gratitude to my wife Elena, who put a massive amount of effort into converting lecture materials to LaTeX, created more than 200 illustrations for this book and gave me support and motivation, without which this book would never have been finished.

# Author

**Andrei Seryi**, currently a professor at Oxford University and Director of the John Adams Institute, graduated from Novosibirsk State University in 1986 and received his Ph.D. from the Budker Institute of Nuclear Physics in 1994. He worked at Stanford Linear Accelerator Center until 2010, where he led the design and first stages of implementation of the Facility for Advanced Accelerator Experimental Tests (FACET) project and coordinated the Beam Delivery efforts for the linear collider. He served as a deputy spokesperson of the ATF International Collaboration for ATF2 Project, is a chairperson or a member of a number of advisory committees and a Fellow of American Physical Society. As well as teaching at numerous accelerator schools, he has contributed to the developments of beam–beam compensation, electron cooling, beam-delivery systems, stability of colliders and novel acceleration methods. His professional interests include: the development of accelerator science applications to discovery science, industry, healthcare and energy; organizing of scientific research; project management; crisis management in scientific and technological areas; inventions and innovations; and developing novel training approaches and methods.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 1 Basics of Accelerators and of the Art of Inventiveness



Can you imagine that electrons
Are planets circling their Suns?
Space exploration, wars, elections
And hundreds of computer tongues

*Author's translation of 1920 poem of Valery Bryusov "The World of Electron"*

In this chapter we will discuss the basic terms related to accelerators. We will also describe the framework and methods of the theory of inventive problem solving.

## 1.1 Accelerators and society

Accelerators are essential for science and society — they are in use in high energy physics, nuclear physics, healthcare and life science. They are important to industry, the development of new materials, energy and security, and can be applied to many other fields.

*About 30% of Nobel prizes in Physics are due to use of accelerators.*

It is enough to state just three reasons why society needs accelerators: tens of millions of patients receive accelerator-based diagnoses and treatments each year in hospitals and clinics around the world; all products that are processed, treated or inspected by particle beams have a collective annual value of more than $500B; and a significant fraction of the Nobel prizes in physics are directly connected to the use of accelerators.

If you are a young person thinking about your future, you might be interested to know that accelerator science can lead to attractive career prospects, because whether you are more inclined towards a theoretical approach, experiments or computer modeling, there will be an array of tasks and opportunities ready for you. Moreover, the knowledge gained and developed in accelerator science has connections and applications to surprisingly remote disciplines, from stock market predictions to planetary motion dynamics.

The recent discovery of the Higgs boson by the Large Hadron Collider — the grandest machine ever built — has once again demonstrated accelerator science's and associated technologies' potential to reveal some of the most fundamental constructs of nature.

DOI: 10.1201/b18696-1

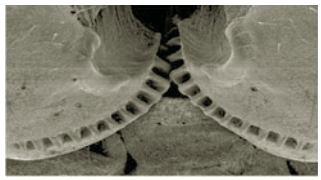


**FIGURE 1.1**
Gear-like structure in jumping insects as an illustration of nature's inventiveness. Burrows and Sutton, 2013. Reproduced with permission.

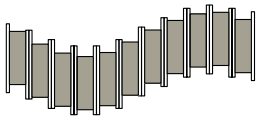


**FIGURE 1.2**
Helical solenoid channel.

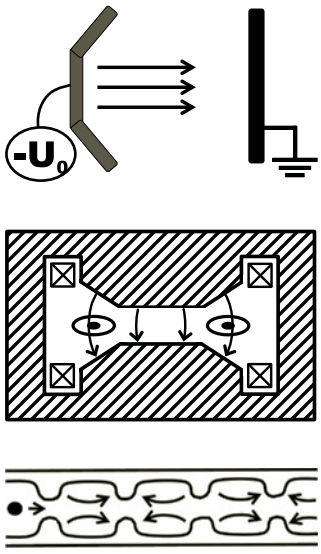



**FIGURE 1.3**
Basic principles of acceleration — electrostatic, betatron, in an EM wave in an accelerating structure.

Accelerator science demonstrates a rich history of inventions, often inspired by nature itself. We are very often motivated by nature and try to compete with it, not always knowing who invented certain things first — nature or humans. For example, it is perhaps a common belief that gears were invented by humans. In fact, insects have been using gears[1] for millions of years! The mechanism allows the insect to jump in a straight line, its left and right leg synchronized by a gear-like connection — see Fig. 1.1 — which works better than a synchronization via nervous signals.

The shapes created by nature often inspire our creativity in accelerator physics as well — e.g., the spiral-shaped Muon collider cooling channel (consisting of an integrated helical solenoid and accelerating cavities interleaved with absorbers; see Fig. 1.2) was possibly inspired by the double-helical DNA. We hope that examples like this, together with a rigorous inventiveness methodology described in this book will arm the reader with a new systematic approach that will enable efficient inventiveness.

## 1.2 Acceleration of what and how

Accelerators can be either giant or tiny, but all have similarities and the same subsystems. A giant accelerator, such as SLAC's linear accelerator and a cathode ray tube TV (which is also an accelerator, albeit smaller), both have all of the main components of a modern linear accelerator or collider. This includes a source of charged particles, an acceleration area, a drift region with focusing and steering, and a target or detector (represented in the case of a TV by the phosphorous screen).

When discussing acceleration, we assume that we accelerate a *bunch* of particles — a compact cloud of, for example, electrons or positrons, protons or antiprotons, ions or any other charged particles.

The simplest accelerating mechanism is electrostatic direct acceleration — caused by DC voltage and a corresponding electric field. Another method is betatron acceleration, which is caused by a magnetic field changing in time which, according to Maxwell's equation $\oint \mathbf{E} \cdot d\boldsymbol{\ell} = -d/dt \int \mathbf{B} \cdot d\mathbf{S}$, creates a curl of electric field $\mathbf{E}$ suitable for acceleration. The third method is acceleration in an electro-magnetic wave; however, one should note that an EM wave in free space cannot continuously accelerate particles along the direction of its propagation, as its $\mathbf{E}$ and $\mathbf{B}$ components are transverse to the direction of the EM wave's propagation. Therefore, in order to use an EM wave for acceleration, one needs to change the

[1] M. Burrows and G. Sutton, *Interacting Gears Synchronize Propulsive Leg Movements in a Jumping Insect*, Science, **341**, 13 Sep 2013.

structure of its fields, which can be achieved by propagating the EM wave in an appropriately shaped accelerating structure. These three methods are illustrated in Fig. 1.3.

Assuming we know how to accelerate the beam, we can ask the question of why would we want to do that? That is, how are we planning to use the accelerated beam? One can foresee at least four different uses of the beam, as illustrated in Fig. 1.4. We can direct the accelerated beam onto a target, either for scientific experiments (e.g., in nuclear physics) or for modifying or treating the target itself. We could direct two accelerated beams onto each other, as is typically done for high energy physics experiments. Acceleration could also be used to characterize the beam or perhaps to separate it into different species or isotopes. Finally, we can use the accelerated beam to generate useful radiation.

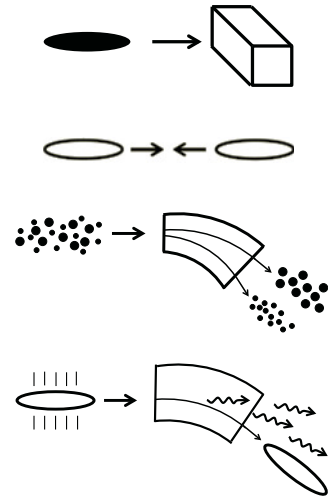


**FIGURE 1.4**
Uses of accelerated beams — sending to target, colliding with another beam, characterization of the beam or separation into species, generation of useful radiation.

### 1.2.1 Uses, actions and the evolution of accelerators

Having discussed in the previous section why we need to accelerate the beam and how an accelerated beam can be used, we will now take this moment to define the basic actions that can be applied to the beam: acceleration, focusing and cooling, and the generation of radiation, as well as the corresponding parameters and characteristics of these actions.

In cases of acceleration we aim to find out the final energy of the particles and usually prefer to achieve as high a rate as possible of the energy change (usually called the accelerating gradient). If the electrostatic accelerating voltage is $U_0$ then the final energy is $E = \gamma mc^2$ equal to $E = eU_0 + mc^2$, where $\gamma$ is relativistic factor, $\gamma = 1 + eU_0/(mc^2)$, $m$ is the rest mass of the particle, $e$ is its charge and $c$ is speed of light.

Whether we plan to send the beam to a target or collide it with another beam, we strive to achieve a certain flux of particles; we therefore may need to focus the beam to a small size on the target or at the interaction point with the oncoming beam. As it is with light, a sequence of focusing and defocusing lenses (in this case electromagnetic lenses) focus the beams, as is illustrated in Fig. 1.5.

Using lenses to focus the beam does not affect its so-called phase-space volume, which is usually called *emittance* $\varepsilon$ and defined as an area of the ellipse occupied by the beam in coordinate-angle phase space (for example, $x$ and $x'$ as illustrated in Fig. 1.5). If the two transverse planes are independent (not coupled), then both $\varepsilon_x$ and $\varepsilon_y$ emittances are conserved. Emittance $\varepsilon$ is usually defined in units of $m \cdot rad$ or $mm \cdot mrad$. If the beam is accelerated, the so-called normalized emittance $\varepsilon_n = \gamma\varepsilon$ is conserved. If the beam emittance is large (which can especially be true for positrons or antiprotons, which are created in "hot" collisions of the initial beam with the target) it can be particularly difficult to focus such

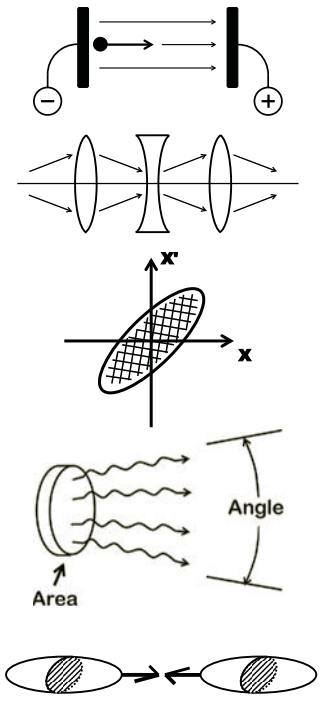



**FIGURE 1.5**
Actions on accelerated beams — acceleration, focusing, generation of radiation, colliding.

a beam to a small size. Therefore, the next important action applied to the beam is *cooling*, which is intended to reduce the emittance of the beam.

In the context of the generation of radiation, we concern ourselves with reducing both the sizes of the beam in the emitting region, and its angular spread (as emitted radiation usually follows the direction of the particles). Therefore, low emittance and beam cooling may again be a necessity. The corresponding characteristic, which describes the radiation generation, is called *brightness* and is defined as the number of photons emitted per unit of time from a certain area into a certain solid angle (see Fig. 1.5). One also typically adds "into a certain bandwidth" (such as, for example, 1% of the central wavelength) as one is usually interested only in a certain spectral range of the emitted radiation. Brightness is therefore defined in the units of $number\,of\,photons/(s \cdot m^2 rad^2(\% \cdot bandwidth))$.

Finally, when two beams are colliding, we aim at focusing the beams into the smallest possible size at the interaction point, in order to maximize the probability of interaction. In this case, we are therefore interested in the characteristic called *luminosity* $\mathcal{L}$, which is defined in such a way that the product of luminosity and the cross section of interaction $\sigma$ (which has the units of $m^2$) gives the number of events per unit of time. The luminosity is then defined in units of $1/(s \cdot m^2)$.

These basic actions or manipulations that can be applied to the beam help to define the evolution of accelerators.

Scientific and technological advances in the area of accelerators have focused on an increase of energy of accelerated beams, mastering acceleration techniques (including the acceleration of different species) and on an increase of accelerating gradients. The need for smaller beam sizes facilitated the development of beam focusing and cooling methods. Demands for higher brightness stimulated mastery of the methods of radiation generation. Desires to increase luminosity of colliding beams led to improvements of a whole class of techniques, from emittance preservation to stabilization of nanometer beams. Lastly, one of the biggest motivations for accelerator progression was their potential to be applied in various scientific and technological areas.

### 1.2.2 Livingston plot and competition of technologies

The history of accelerators and various accelerator technologies can be summarized in a so-called "Livingston plot" where the equivalent energy of an accelerated beam is plotted against time — see Fig. 1.6. One can clearly see that, over the course of many decades, the maximum achieved energy grew exponentially. It was the development of different accelerator technologies that enabled this exponential growth.

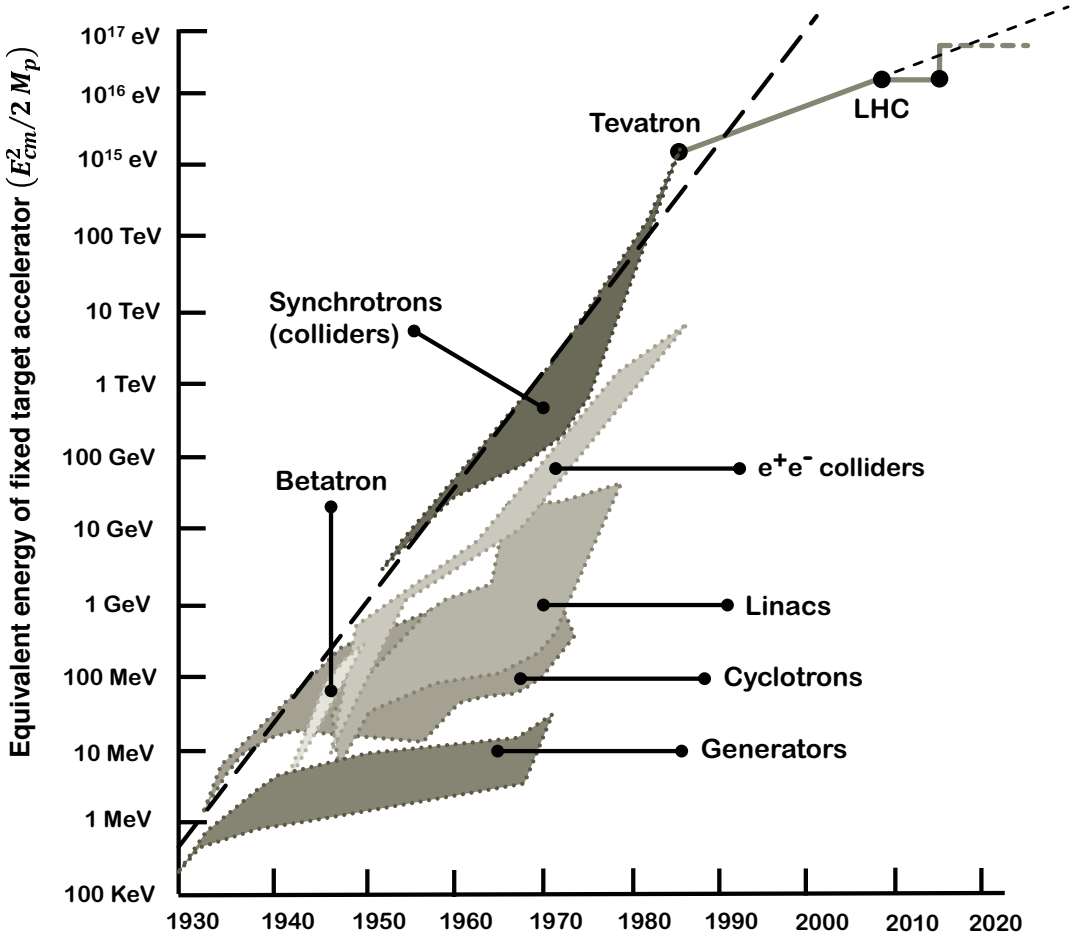



**FIGURE 1.6**
Livingston plot of evolution of accelerators.

The Livingston plot also depicts that new accelerating technologies replaced each other once the previous technology had reached its full potential. This evolution and *saturation* of particular acceleration technologies, and birth of the new technologies, is a common phenomenon in any technological or scientific field — illustrated in Fig. 1.7.

The three most recent decades shown in Fig. 1.6, represented by Tevatron and LHC colliders, exhibit a much slower exponential energy growth over time. This may be an indication that the existing technologies of acceleration came to their maximum potential, and that further progress would demand creation of new accelerating methods — one which was more compact and more economical. There are several emerging acceleration techniques, such as laser-driven and beam-driven plasma acceleration, which can bring the Livingston plot back to the exponential path — which we will further discuss later on.

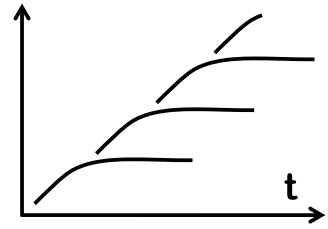



**FIGURE 1.7**
Evolution of technologies — saturation and replacement by newer technologies.

## 1.3 Accelerators and inventions

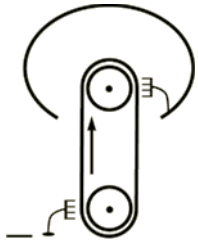


**FIGURE 1.8**
Van der Graaf accelerator.

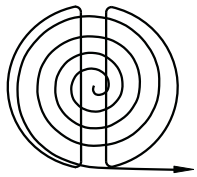


**FIGURE 1.9**
Cyclotron accelerator.

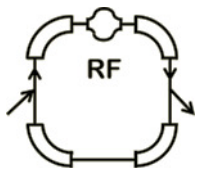



**FIGURE 1.10**
Synchrotron accelerator.

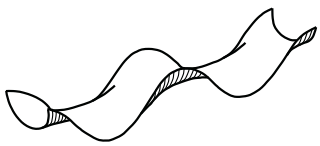


**FIGURE 1.11**
Strong focusing concept.

Accelerator science exhibits a rich history of inventions. Let us briefly skim through some of the most influential inventions. Those interested in further reading are recommended to explore the fascinating story of accelerators described in *Engines of Discovery*.[2]

Between 1900 and 1925, radioactive source experiments initiated by Rutherford created a demand for higher energy beams.

From 1928 to 1932, Cockcroft and Walton developed electrostatic acceleration using voltage multiplication created out of diodes and oscillating voltage — ultimately reaching around 700 kV of voltage. At about the same time, Van der Graaf created a method of voltage charging wherein a rubber mechanical belt would carry charges deposited onto the belt by sharp needles via the ionization of gas molecules — which achieved voltages at around 1.2 MV (see Fig. 1.8).

Resonant acceleration development commenced in 1928 with Ising establishing the concept and Wideroe building the first linac.

In 1929, Livingston built his small prototype of the cyclotron as his PhD thesis, inspired by Lawrence, who studied Wideroe's thesis (see Fig. 1.9). This concept was then realized later in large scale by Lawrence.

In 1942, the principle of magnetic induction helped Kerst build the first betatron.

In 1944, the synchrotron (see Fig. 1.10) was invented by Oliphant, while MacMillan and Veksel independently invented the principle of RF phase stability, which made longitudinal focusing of beams possible.

In 1946, Alvarez built the proton linac by using an RF structure with drift tubes (in progressive wave in $2\pi$ mode).

In 1950, Christofilos developed strong focusing, which he later patented as the alternate gradient concept (transverse strong focusing). Fig. 1.11 illustrates the strong focusing as a gutter with its edges bent up and down in a sine-like manner. The story of the invention of strong focusing has an interesting aspect — it is usually attributed to Courant and Snyder, since the Christophilos patent was pointed out only after the Cosmotron team had announced the idea. This example is certainly relevant for anyone considering whether to publish or to patent their ideas.

In 1951, a tandem Van der Graaf accelerator was developed by Alvarez, thusly upgrading the electrostatic acceleration concept. A charge-exchange stripping foil was placed at the high voltage point and the source of negative ions was placed at ground potential (which was also much more practical for its servicing). This resulted in a voltage twice as large.

[2]Engines of Discovery, A Century of Particle Accelerators, A. Sessler and E. Wilson, World Scientific, 2014.

In 1953, Courant, Snyder, and Livingston built the weak focusing 3.3 GeV Cosmotron in Brookhaven, and in 1957, Veksler built a 10 GeV (which was the world record at that time) synchrophasotron in Dubna, whose magnet weighed 36,000 tons and was registered in the Guinness Book of World Records. This record in energy was overtaken by the CERN Proton Synchrotron in 1959, constructed under the leadership of Sir John Adams. The CERN PS was the first strong focusing accelerator, closely followed by the AGS at Brookhaven.

In 1956, Veksler suggested the principle of collective acceleration (see Fig. 1.12), which became the predecessor to a variety of collective methods based on plasma acceleration.

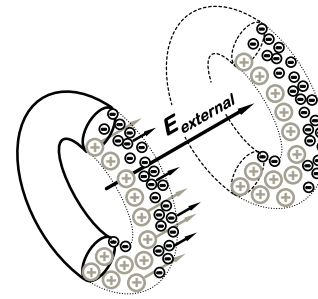



**FIGURE 1.12**
Collective acceleration.

In 1956, Kerst discussed the concept of colliding beams and in 1961, the e+e- collider — the concept for a particle-antiparticle collider — was invented by Touschek.

The collider concept created the need to develop the methods used to decrease beam emittances, in particular for antiparticles, and in 1967, Budker proposed electron cooling as a way to increase the proton or antiproton beam density (see Fig. 1.13). Shortly thereafter, Van der Meer proposed stochastic cooling (see Fig. 1.14) in 1968 as a way to compress the beam's phase space.

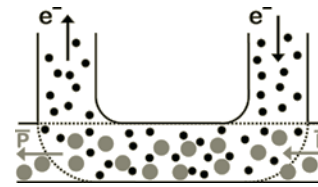



**FIGURE 1.13**
Electron cooling concept.

In 1970, RFQ was invented when Kapchinski and Telyakov build the first radiofrequency quadrupole that allowed a simultaneous focusing and acceleration of the beam. This technology allowed a much more efficient and compact acceleration of ions and protons from a very low energy to an energy as high as a few MeV.

In 1971, Madey developed the principle of a Free Electron Laser (FEL), a method that allowed the production of coherent, hard X-rays of unprecedented brightness.

In 1979, Tajima and Dawson proposed acceleration of the beam in plasma waves excited by a laser (see Fig. 1.15). It was only much later that suitable lasers became available and this method started to become competitive.

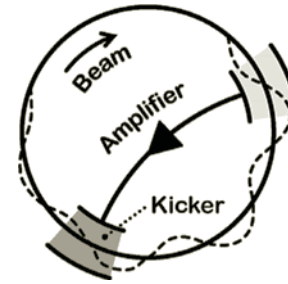



**FIGURE 1.14**
Stochastic cooling concept.

From around 1980, superconducting magnets were developed in various labs around the world, which allowed for a drastic increase of beam energy in circular acceleration. Development of this technology still continues today (8 Tesla magnets are routinely used in LHC, and magnets exceeding 10 Tesla are being developed for LHC upgrade).

Likewise, the superconducting RF (radio-frequency) technology (created around 1980) is currently being used and is still being developed today in many labs and industrial companies, allowing an increase in the RF gradient and efficiency, and thus the eventual energy and power of the beams.

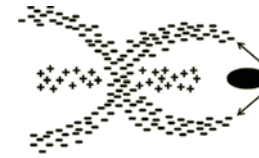


**FIGURE 1.15**
Plasma acceleration concept.

The most recent decades have been rich with inventions as well. The years between 1990 and the present day have seen a photon collider concept, an (experimentally verified) crab

waist collision and integrable optics for storage rings, just to name a few.

The variety of the inventions (many of which will be discussed later on in further detail) demonstrates the *past successes* of our field, but we are concerned with a different question: what can we look forward to inventing in the future, and how are we to invent it more efficiently?

In the next section we will introduce the methodical approaches to inventiveness used in industry, and will then explain how these methods can be applied to, in this case, accelerator science.

## 1.4 How to invent

Methods described in this section are rarely known in science, but they are widely used in industry.

*Formal inventive approaches rarely known in science but widely used in industry*

In his March 7, 2013 contribution to Forbes,[3] "What Makes Samsung Such An Innovative Company?" the author, Haydn Shaughnessy, wrote: "But it was ... that became the bedrock of innovation at Samsung. In 2003 ... led to 50 new patents for Samsung and in 2004 one project alone, a DVD pick-up innovation, saved Samsung over $100 million. ... is now an obligatory skill set if you want to advance within Samsung." The reader may speculate on what the dots signify — but read on.

What is this magic method, and what is the word intentionally omitted in the above quote? The answer will be given in just a few pages.

### 1.4.1 How to invent — evolution of the methods

Let's start by recalling the techniques of problem solving, starting from a well-known brute force or exhaustive search method. In this case, any potential solutions and ideas are considered and evaluated one by one. It is easy to imagine how inefficient this method may be.

An improved variation of the exhaustive search method is called *brainstorming*. Its author, Alex Osborn, introduced this method in the 1950s. Brainstorming is a psychological approach that helps to solve problems and create inventions by separating the process of idea generation from the process of critical analysis.

The method of brainstorming has its limitations — the absence of critical feedback, which is the main feature of the method, is simultaneously its handicap, as feedback is required to develop and improve an idea.

The next method that emerged on the scene, the so-called

[3] http://www.forbes.com/sites/haydnshaughnessy/2013/03/07/why-is-samsung-such-an-innovative-company/

*synectics*, was introduced by George Prince and William Gordon in an attempt to improve brainstorming. One of the main feature of synectics is that it is assumed to be practiced by a permanent group of experts, whose members, with time, become less sensitive to critics from among their peers and thus become more efficient in inventing. This method emphasized the importance of seeing familiarity in an unknown and vice versa, which helps to solve new and unfamiliar problems using known methods. The authors of synectics stressed the importance of approaching a problem with a fresh gaze, and also stressed the use of analogies to generate this attitude.

The analogies employed by synectics can be direct (any analogy, e.g., from nature), emphatic (attempting to look at the problem by identifying yourself with the object), symbolic (finding a short symbolic description of the problem and the object), or even metaphorical (describing the problem in terms of fairy-tales and legends).

Synectics is still a variation of the exhaustive search method and is in fact the limit of what can be achieved, maintaining the brute force method of an exhaustive search.

There must, then, be a better inventive approach free from the irrationality of synectics. Indeed, why would one employ analogies and metaphors and other irrational factors in order to come to a formula — *the action has to happen itself* — a natural and universal inventive formula of an ideal invention? Such a formula (which means that the action needed to solve the problem should happen by itself, i.e., without introducing any additional systems) should indeed be programmed into the process of any inventive solution, aimed at an appropriately selected part of the object, and with precise identification of the physical contradiction and intended physical action. These features are the characteristics of *TRIZ*, the powerful inventive method described in the next section.

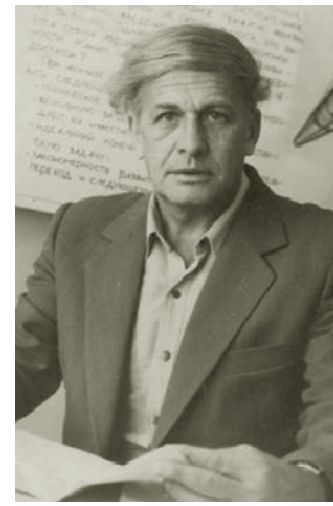
## 1.5 TRIZ method

The abbreviation TRIZ (pronounced *[treez]*) can be translated as the Theory of Inventive Problem Solving. TRIZ was developed by Genrikh Altshuller in the Soviet Union in the mid-20th century. The author, while working in the patent office since 1946, analyzed many thousands of patents, trying to discover patterns to identify what makes a patent successful.

Between 1956 and 1985, following his work in the patent office, Altshuller formulated TRIZ and, together with his team of supporters, developed it further.

Altshuller and his team devised four key constructs of the TRIZ methodology:

- The same problems and solutions appear again and again but in different industries.



The creator of TRIZ G.S. Altshuller (October 15, 1926 – September 24, 1998). Published with permission of the Official G.S. Altshuller Foundation.

- There is a recognizable technological evolution path for all industries.
- Innovative patents (about a quarter of the total) used science and engineering theories outside their own area or industry.
- An innovative patent uncovers and solves contradictions.

Together with these main clauses, the authors created a detailed methodology, which we will review in the next section before making the connection to physics.

And as the reader has likely already guessed, it was the word "TRIZ" that was skipped in the quote at the beginning of the previous section. It was TRIZ that became the bedrock of innovation at Samsung, and which can do so much more if applied to other existing fields such as accelerator science.

### 1.5.1 TRIZ in action — examples

Let us consider the TRIZ approach with an example that is often cited in TRIZ textbooks. Imagine that we need to polish an optical lens with an abrasive stick (see Fig. 1.16) and we need to do it quickly. However, there is a problem: polishing the lens generates heat, which degrades the optical properties of the produced lens. Existing cooling methods are ineffective, as one cannot supply adequate and uniform cooling to each abrasive particle.

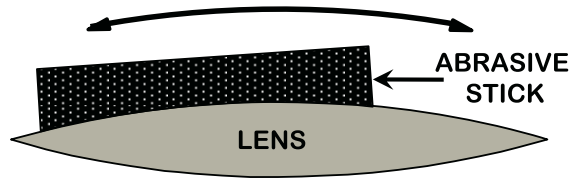



**FIGURE 1.16**
Illustration of TRIZ in action — initial specific problem.

The first step of the TRIZ approach consists of identifying the pair of contradicting parameters: the one that needs to be improved and the one that simultaneously gets worse. In the considered case of the lens polishing, these parameters are the *speed* (the one to be improved) and the *temperature* (the one that gets worse).

The second step of the TRIZ algorithm starts by identifying if anyone else has solved such a contradiction in the past. And it is here that the power of TRIZ lies — hundreds of thousands of analyzed patents allowed the TRIZ team to put together a table, *the contradiction matrix*, which helps to find relevant *inventive principles* corresponding to a particular pair of contradicting parameters.

The beauty of the TRIZ method is also illustrated by the fact that the entire field of engineering can be described by the TRIZ contradiction matrix using only 39 parameters, and that the number of TRIZ inventive principles is also remarkably small — only 40.

The second step of the TRIZ algorithm involves checking[4] the TRIZ matrix for this pair of contradicting parameters and then identifying the relevant, in this case, inventive principles.

Having identified the necessary inventive principles, we are ready for the third and final step of the TRIZ algorithm — translating a generic inventive principle into a specific solution. Graphically, the flow of the TRIZ algorithm is shown in Fig. 1.17.

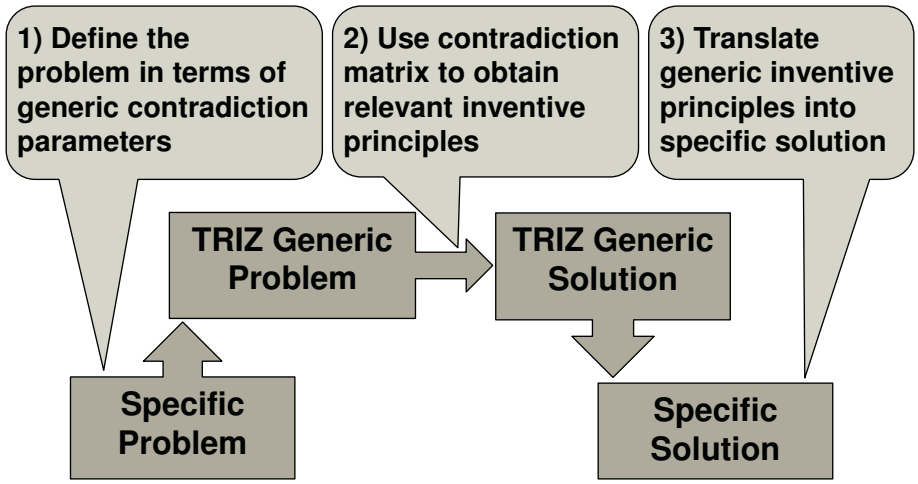



**FIGURE 1.17**
Illustration of the flow of the TRIZ algorithm.

In our particular example of the lens that needs to be polished, the contradiction matrix zoomed into the crossing of the *speed* parameter, which needs to improve without damaging the *temperature* parameter, is pictured in Table 1.1.

The inventive principles listed (according to their numerical index) in the speed–temperature cell of the contradiction matrix are: 2-Taking out; 28-Mechanics substitution; 30-Flexible shells and thin films; 36-Phase transitions.

The next step is to select the inventive principle that is most suitable which in our example can be judged to be the latter one — 36-Phase transition (use of phenomena occurring during phase transitions, such as volume changes, loss or absorption of heat, etc., according to the description of this TRIZ principle).

The corresponding specific solution to the lens polish-

[4]See, e.g., http://www.triz40.com/

**TABLE 1.1**
TRIZ contradiction matrix for speed–temperature parameters and indexes of the corresponding inventive principles

| Improving parameter | Parameter that deteriorates | | | | |
|---|---|---|---|---|---|
| | ... | 9.Speed | ... | 17.Temperature | ... |
| ... | ... | ... | ... | ... | ... |
| 9.Speed | ... | ... | ... | 2,28,30,36 | ... |
| ... | ... | ... | ... | ... | ... |
| 17.Temperature | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |

ing dilemma (which can be suggested according to the selected inventive principle of phase transition) is to use ice together with abrasive particles, which provides efficient uniform cooling, as illustrated in Fig. 1.18.

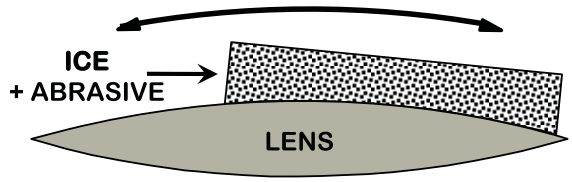



**FIGURE 1.18**
Illustration of TRIZ in action — specific solution.

Before we discuss the connection between TRIZ and science, allow us to present a complete list of the standard TRIZ contradicting parameters (shown in Table 1.2) as well as the list of inventive principles (Table 1.3). A detailed description of the TRIZ principles is also available at the web page mentioned in footnote 4 of this chapter.

**TABLE 1.2**
Elements of TRIZ contradiction matrix

| No.Parameter | No.Parameter | No.Parameter |
|---|---|---|
| 1. Weight of moving obj. | 14. Strength | 27. Reliability |
| 2. Weight of stat. obj. | 15. Durability of mov. obj. | 28. Measurement accuracy |
| 3. Length of moving obj. | 16. Durability of stat. obj. | 29. Manufacturing precision |
| 4. Length of stat. obj. | 17. Temperature | 30. Object-affected harmful |
| 5. Area of moving object | 18. Illumination intensity | 31. Object-generated harmful |
| 6. Area of stationary obj. | 19. Energy use by mov. obj. | 32. Ease of manufacture |
| 7. Volume of moving obj. | 20. Energy use by stat. obj. | 33. Ease of operation |
| 8. Volume of stat. obj. | 21. Power | 34. Ease of repair |
| 9. Speed | 22. Loss of energy | 35. Adaptability or versatility |
| 10. Force (Intensity) | 23. Loss of substance | 36. Device complexity |
| 11. Stress or pressure | 24. Loss of information | 37. Difficulty of detecting |
| 12. Shape | 25. Loss of time | 38. Extent of automation |
| 13. Stability of the object | 26. Quantity of substance | 39. Productivity |

**TABLE 1.3** TRIZ inventive principles

| No. Principle | No. Principle | No. Principle |
|---|---|---|
| 1. Segmentation | 15. Dynamics | 29. Pneumatics, hydraulics |
| 2. Taking out | 16. Partial or excessive actions | 30. Flexible shells, thin films |
| 3. Local quality | 17. Another dimension | 31. Porous materials |
| 4. Asymmetry | 18. Mechanical vibration | 32. Color changes |
| 5. Merging | 19. Periodic action | 33. Homogeneity |
| 6. Universality | 20. Continuity of useful action | 34. Discarding, recovering |
| 7. Russian dolls | 21. Skipping | 35. Parameter changes |
| 8. Anti-weight | 22. Blessing in disguise | 36. Phase transitions |
| 9. Preliminary anti-action | 23. Feedback | 37. Thermal expansion |
| 10. Preliminary action | 24. Intermediary | 38. Strong oxidants |
| 11. Beforehand cushioning | 25. Self-service | 39. Inert atmosphere |
| 12. Equipotentiality | 26. Copying | 40. Composite materials |
| 13. "The other way round" | 27. Cheap short-lived objects | |
| 14. Spheroidality — Curvature | 28. Mechanics substitution | |

## 1.6 TRIZ method for science

We have finally arrived at the section that will elucidate the meaning of the epigraph to this chapter.

The TRIZ inventive principle of Russian dolls (nested dolls, or *matreshka*) can be applied not only to engineering, but to many other areas, including science. A rather spectacular example is the construction of a high energy physics detector, where many different sub-detectors are inserted into one another, like a nested doll, in order to enhance the accuracy of detecting elusive particles (see Fig. 1.20).

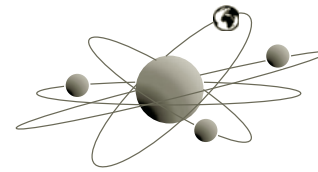


**FIGURE 1.19**
Valery Bryusov's electron as an analogy to the TRIZ inventive principle of nested dolls.

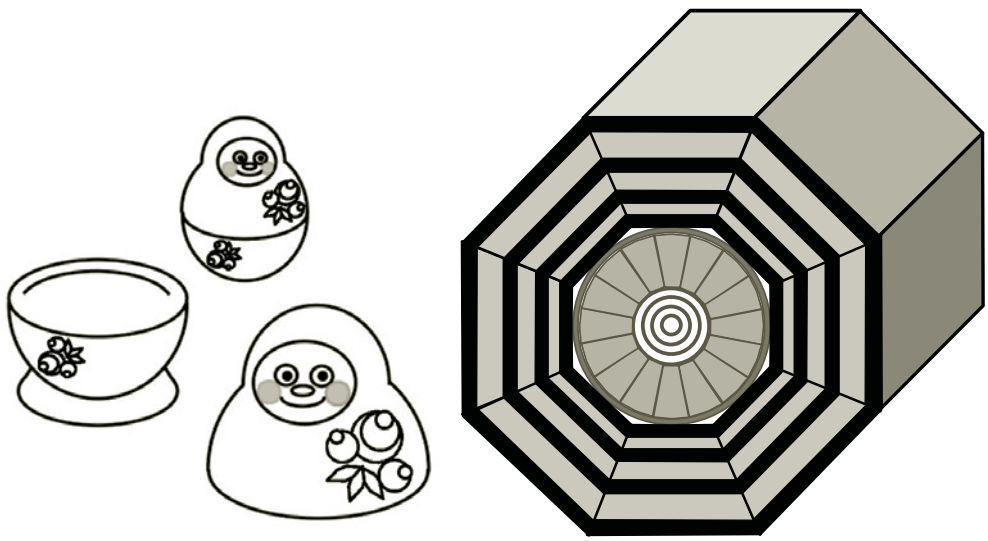


**FIGURE 1.20**
High energy physics detectors, which have a layered "nested" structure, reflecting the TRIZ inventive principle of Russian dolls.

The reader can now see that the 1920 poem by Valery Bryusov, which describes an electron as a planet in its own world (see Fig. 1.19), can also be seen as reflection of the nested doll inventive principle, this time in poetic science fiction.

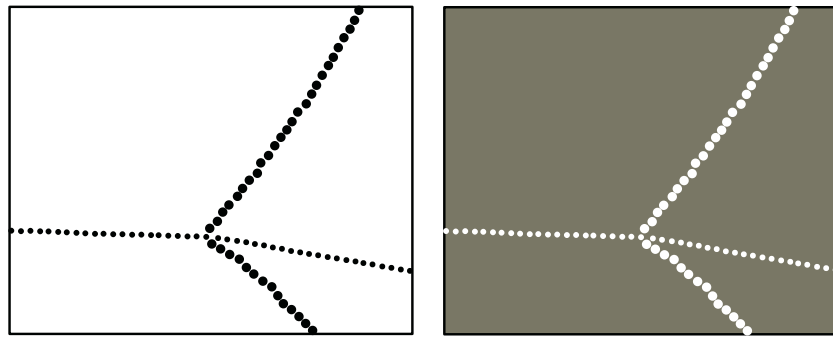


**FIGURE 1.21**
Particle interaction event observed in a cloud chamber invented by Wilson in 1911 (left), and in a bubble chamber invented by Glaser in 1952 (right).

TRIZ textbooks also often cite Wilson's cloud chamber (invented in 1911) and Glaser's bubble chamber (invented in 1952) as examples, in the terminology of TRIZ, of a system and anti-system. Indeed, the cloud chamber works on the principle of bubbles of liquid created in gas, whereas the bubble chamber uses bubbles of gas created in liquid (see Fig. 1.21).

It is thus reasonable to ask: would it have taken almost half a century to invent the bubble chamber had the TRIZ anti-system principle been applied?

The systematic application of TRIZ to science is indeed a valid question, and moreover, it can give us new insights. We will now discuss TRIZ in connection with accelerator science.

## 1.7 AS-TRIZ

The TRIZ method was originally created for engineers. However, the methodology is universal and, as the previous section demonstrated, can largely be applied to science, and in particular to accelerator science.

*AS-TRIZ is currently being developed. More examples will appear in the following chapters.*

Many of the parameters from the TRIZ contradiction matrix, as well as TRIZ inventive principles, can be directly applied to problems arising in accelerator science. Still, accelerator science is a distinct discipline, and it is only natural to add accelerator science-related parameters and inventive

principles to TRIZ. We call this extension *Accelerating Science TRIZ* or *AS-TRIZ*, highlighting via its name a wide applicability of the method to *various* areas of science, even those beyond the field of accelerators.

Below are just a few suggested additional parameters for the AS-TRIZ matrix of contradictions (which are based on manipulated beams and corresponding characteristics discussed in Section 1.2.1), and just a couple of inventive principles to start with — see Table 1.4 and Table 1.5. We will populate this table as we move along the chapters.

**TABLE 1.4**
Emerging AS-TRIZ parameters

| No. | Parameter |
|---|---|
| 1. | Energy |
| 2. | Rate of energy change |
| 3. | Emittance |
| 4. | Luminosity |
| 5. | Brightness |
| 6. | Intensity |
| 7. | Efficiency |
| 8. | Power |
| 9. | Integrity of materials |
| 10. | Time duration or length |
| 11. | Spatial extent |
| 12. | Sensitivity to imperfections |
| 13. | Cooling rate |
| 14. | ... |
| ... | ... |
| 21. | ... |

**TABLE 1.5**
Emerging AS-TRIZ principles

| No. | Principle |
|---|---|
| 1. | ... |
| 2. | ... |
| 3. | Undamageable or already damaged |
| 4. | Volume-to-surface ratio |
| 5. | ... |
| ... | ... |
| 21. | ... |

The first inventive principle shown in Table 1.5, number 3: "Undamageable or already damaged," is illustrated by the three examples below, while 4, "Volume-to-surface ratio," naturally arises, for example, from Maxwell's equation, where an integral on a surface is connected to the integral over volume $\mathbf{E}d\mathbf{S} \propto \rho dV$, or from other similar equations, e.g., from thermodynamics. We will discuss examples corresponding to this principle later on in the book.

Let us illustrate an application of AS-TRIZ to a couple of accelerator science examples.

The first example is the beam profile monitor with its tungsten or carbon wire, shown in Fig. 1.22. In order to measure the beam's profile, the wired frame needs to cross the beam very quickly. The problem is that, as beam intensity increases, the beam and energy losses in the wire also increase, and the wire can get damaged after a single use. The physical contradiction in this case is between the parameters *intensity* (to be improved) and *integrity* (what degrades).

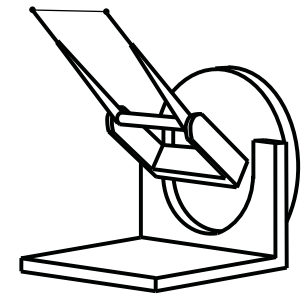
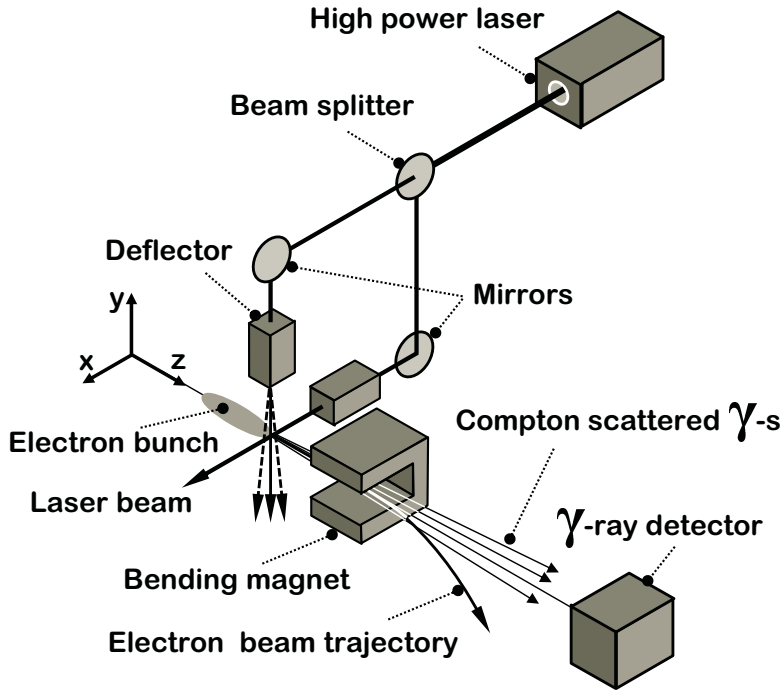


**FIGURE 1.22**
Carbon wire beam profile monitor.

In order to solve the beam profile monitor problem, let

us apply emerging AS-TRIZ principle 3: "Undamageable or already damaged." That is, we must replace the material that can be damaged with another media which either cannot be damaged (light) or is already "damaged" (e.g., plasma).

Selecting light as a medium that cannot be damaged naturally arrives at an already well-known solution — a beam profile monitor based on laser wire, shown in Fig. 1.23.




*In the laser wire the bunch interacts with optical light that acts like the physical wire of the mechanical beam profile monitor.*

**FIGURE 1.23**
Laser wire beam profile monitor.

The next example is a standard glass mirror. The problem here is that, as the intensity of the laser increases, the mirror can get damaged. The contradiction is again between *intensity* and *integrity*. Let us apply the same principle 3 of AS-TRIZ, and replace the mirror with something that is already damaged — plasma. Indeed, such a solution is already known in laser and plasma science — the so-called plasma mirror. A plasma mirror is created by focusing a laser beam onto a piece of glass or gas. It will then be transmitted until the intensity reaches a certain value sufficient to produce plasma on the surface. At that moment, the reflective index of the surface changes, and the laser beam is then reflected from the plasma's free electrons.

The final example for this chapter, one that we will return to for further details later on, is the dilemma of accelerating structures. Made from metal, the acceleration structures (either normal conducting or super-conducting) are suscepti-

ble to breakdowns that limit the useful accelerating gradient, typically to about 100 MeV/m.

The problem of an accelerating cavity can therefore be formulated as follows: as the rate of energy E change (accelerating gradient) increases, the surface of cavities gets damaged with occasional breakdowns. The contradiction is therefore between *rate of E change* (to be improved) and *integrity* (what degrades).

Applying again principle 3 of AS-TRIZ, and replacing accelerating structure with plasma, we come to the well-known concept of plasma acceleration, when an "accelerating structure" is temporarily produced in plasma by a driving laser pulse.

By taking note of these three examples, we already can start to fill in the AS-TRIZ contradiction matrix with inventive principles — see Table.1.6.

**TABLE 1.6**
Emerging AS-TRIZ contradiction matrix and indexes of inventive principles

| Improving parameter | Parameter that deteriorates | | | |
|---|---|---|---|---|
| | ... | ... | 9.Integrity | ... |
| ... | ... | ... | ... | ... |
| 2.Rate of E change | ... | ... | 3,... | ... |
| ... | ... | ... | ... | ... |
| 6.Intensity | ... | ... | 3,... | ... |
| ... | ... | ... | ... | ... |

Many more examples related to TRIZ and AS-TRIZ will be discussed in the following chapters. However, the style of the book from this point on will change — the main text will follow the standard style of scientific textbooks, while the occasional notes (primarily in the margins) will point out relevant connections to TRIZ. We will also use TRIZ in order to highlight similarities between the three discussed subjects — physics of accelerators, lasers and plasma.

These two parallel narratives, about science and about inventions, will eventually merge again in the final chapter, where we will summarize our discussions about the art of inventiveness and accelerator science.

## 1.8 Creativity

Use of TRIZ does not take away your creativity! It instead strives to encourage and inspire you.

It helps to find potential solutions, narrowing down the choices, but does not find the final answer for you. By arming you with a considerably deep breadth of knowledge, TRIZ

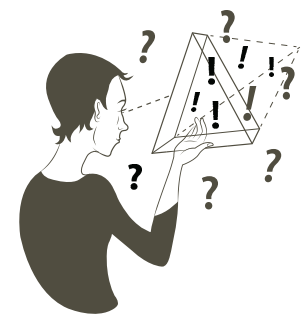


**FIGURE 1.24**
Looking at the world through the prism of TRIZ. Illustration by Sasha Seraia.

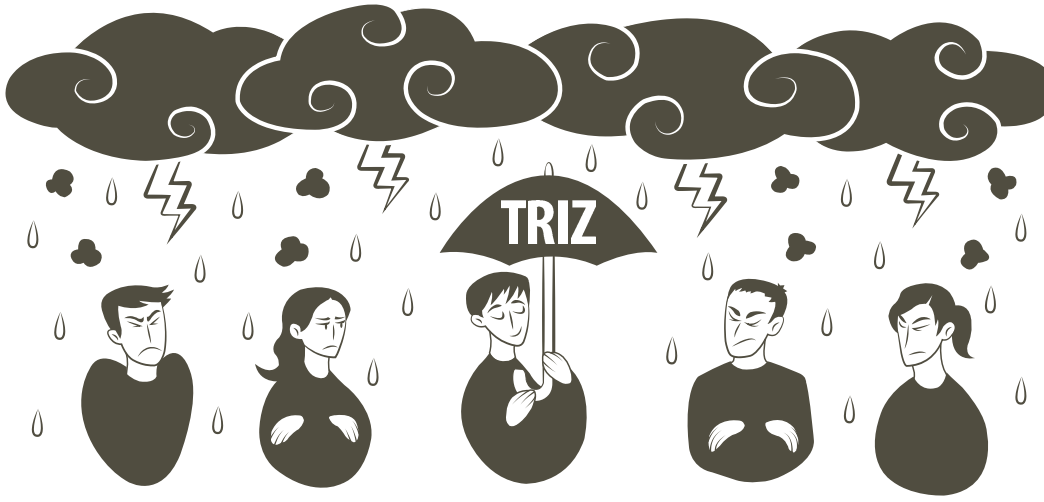



**FIGURE 1.25**
TRIZ vs. brainstorming. Illustration by Sasha Seraia.

can help you to be successful if you choose to apply its methods (see Fig. 1.25 and Fig. 1.24).

Moreover, use of TRIZ will help us to build bridges between accelerators, lasers and plasma, and indeed even possibly other currently unrelated sciences, technology or engineering. Remember, one of the most important principles of TRIZ is that "the same problems and solutions appear again and again but in different disciplines."

**EXERCISES**

1.1 *Analyze the evolution of technical or scientific systems.* Discuss the evolution of any scientific or technical (or accelerator-science-related) area, identifying successive technologies, which arise, saturate and get replaced by new approaches and solutions.

1.2 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.* A plasma mirror is often used in the situation when a standard metal mirror cannot withstand the power density of the laser. Analyze this technology in terms of the TRIZ and AS-TRIZ approach, identifying a contradiction and a general inventive principle that were used (could have been used) in this invention.

1.3 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.* Analyze and describe any scientific or technical (possibly, related to accelerator science) invention/discovery in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) in this invention or discovery.

1.4 *Developing AS-TRIZ parameters and inventive principles.* Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 2 Transverse Dynamics



Let's begin our discourse about the basics of accelerator physics with the topic of the transverse dynamics of charged particles. This will lead into a discussion of the basics of synchrotron radiation and of acceleration in the following chapters, intermediated by a dialogue on the synergies between accelerators, lasers and plasma.

## 2.1 Maxwell equations and units

We start by recalling the Maxwell equations with an emphasis on their systems of units, focusing in particular on SI and Gaussian-cgs systems. The microscopic Maxwell equations (i.e., equations in vacuum) expressed in SI units, in both differential and integral form, are

*This Maxwell equation explains the universality of the inventive principle of changing the volume-to-surface ratio.*

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0} \quad \text{or} \quad \oint_{\partial\Omega} \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\varepsilon_0} \int_{\Omega} \rho dV \tag{2.1}$$

$$\nabla \cdot \mathbf{B} = 0 \quad \text{or} \quad \oint_{\partial\Omega} \mathbf{B} \cdot d\mathbf{S} = 0 \tag{2.2}$$

*Permittivity of free space* $\varepsilon_0 \approx 8.85\, F/m$ *or* $A^2 s^4 kg^{-1} m^{-3}$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad \text{or} \quad \oint_{\partial\Sigma} \mathbf{E} \cdot d\boldsymbol{\ell} = -\frac{d}{dt} \int_{\Sigma} \mathbf{B} \cdot d\mathbf{S} \tag{2.3}$$

*Vacuum permeability* $\mu_0 = 1/(c^2 \varepsilon_0) \approx 1.26 \times 10^{-6}\ N \cdot A^{-2}$

$$\nabla \times \mathbf{B} = \mu_0 \left( \mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \quad \text{or}$$

$$\oint_{\partial\Sigma} \mathbf{B} \cdot d\boldsymbol{\ell} = \mu_0 \int_{\Sigma} \mathbf{J} \cdot d\mathbf{S} + \mu_0 \varepsilon_0 \frac{d}{dt} \int_{\Sigma} \mathbf{E} \cdot d\mathbf{S} \tag{2.4}$$

*Speed of light in vacuum* $c \approx 2.99 \times 10^8\ m/s$

The Lorentz force acting on a charged particle in an electric and magnetic field in SI units is expressed as

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.5}$$

While SI is the standard, the Gaussian system is more natural for electromagnetism. The differential Maxwell equations expressed in Gaussian-cgs units are

DOI: 10.1201/b18696-2

$$\nabla \cdot \mathbf{E} = 4\pi\rho \quad , \quad \nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \quad , \quad \nabla \times \mathbf{B} = \frac{1}{c}\left(4\pi\mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}\right) \tag{2.6}$$

*Charge of electron in SI:* $\approx 1.6 \times 10^{-19}$ *C or* $A \cdot s$

And the Lorentz force is

$$\mathbf{F} = q\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right) \tag{2.7}$$

*Charge of electron in Gaussian units:* $\approx 4.8 \times 10^{-10}$ *cgs units*

The above equation highlights why the Gaussian system is so useful: the electric and magnetic fields are expressed in the same units, which stresses that these fields have a similar nature.

*Deriving a formula, instead of using quantities such as e or h, express the end result via more natural quantities* ($m_e c^2, r_e, \lambda_e, \alpha$, *etc.).*

Throughout this text we will use both SI and Gaussian units. We will, however, plan to construct equations or at least the end results in such a way that allows for easy conversion between different units. As a result, we will avoid having quantities such as electric charges or Planck's constant in our equations, and instead will use only quantities of length, speed and energy.

Let's write down some of the equations that will be useful for such conversions into the natural and unit-independent form.

The most useful one is for the classical radius of an electron:

$$\text{SI}: \; r_e = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{m_e c^2} \quad , \quad \text{Gauss}: \; r_e = \frac{e^2}{m_e c^2} \tag{2.8}$$

$r_e \approx 2.82 \cdot 10^{-15} m$

And the one for the fine structure constant:

$$\text{SI}: \; \alpha = \frac{e^2}{(4\pi\varepsilon_0)\hbar c} \quad , \quad \text{Gauss}: \; \alpha = \frac{e^2}{\hbar c} \tag{2.9}$$

$\alpha \approx 1/137$

The latter also gives us the reduced Compton wavelength for the electron ($\lambda\!\!\!^{-}_e = r_e/\alpha \approx 3.86 \cdot 10^{-13} m$), which is very useful whenever the Planck's constant needs to be hidden in an equation.

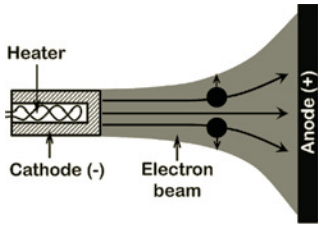



**FIGURE 2.1**
Simple electron gun.

## 2.2 Simplest accelerator

An example of a simple accelerator is a thermionic gun, in which thermionic emission from the cathode generates electrons that are accelerated across a high voltage gap to the anode. A variety of materials can be used for these cathodes. Pure metals are the most robust choice but require extremely high temperatures. On the other hand, the oxides cathodes are covered with alkaline earth metal oxides (e.g., the BaO *impregnated* tungsten cathode), and provide a higher emission at lower temperatures. However, they are more delicate

DOI: 10.1201/b18696-1

as they require a better vacuum for their operation and are vulnerable to poisoning by heavier elements present in the residual gas.

In thermionic guns, a grid can be installed between the anode and cathode. Applying pulsed voltage to the grid allows for the generation of a train of pulses suitable for consequent RF acceleration.

In an electron gun, the electrons generated by thermionic emission tend to repel from each other as illustrated in Fig. 2.1, resulting in reduced beam quality. The additional focusing electrodes can help to maintain the quality of the beam. In particular, it was shown by J.R. Pierce in 1954 that in planar geometry, an electrode at the potential of the cathode inclined at 67.5° (called the Pierce angle), as shown in Fig. 2.2, will help to maintain a parallel flow of electrons. Any intermediate accelerating electrodes or the anode would need to be placed along the equipotential lines, as shown in Fig. 2.2, to maintain the parallel flow.

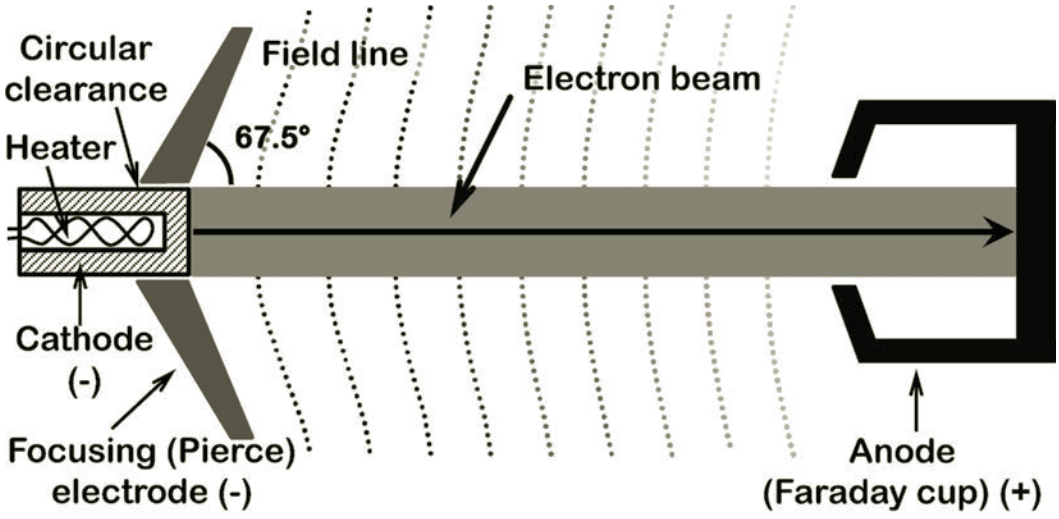



**FIGURE 2.2**
Electron gun with Pierce electrode and collector made in the form of a Faraday cup.

The same Fig. 2.2 also illustrates the concept of the beam anode (collector) made in the shape of a Faraday cup — a useful device for the accurate measurement of electron current. Any secondary charged particles emitted from the walls are eventually absorbed and do not affect the measurements of the current.

Controlling the beam shape and beam quality in high density electron guns often requires the use of an accompanying solenoid magnetic field. Let us now consider equations for the motion of charged particles in electromagnetic fields.

## 2.3 Equations of motion

### 2.3.1 Motion of charged particles in EM fields

The motion of a particle with charge $q$ in an electric **E** and magnetic **B** fields is given by the following equations:

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad , \quad \frac{d\mathcal{E}}{dt} = \mathbf{F} \cdot \mathbf{v} \tag{2.10}$$

where the momentum **p** and energy $\mathcal{E}$ of the particle are

$$\mathbf{p} = m_0 \gamma \mathbf{v} \quad , \quad \mathcal{E} = m_0 \gamma c^2$$

and $m_0$ is the particle mass, and $\gamma$ is the relativistic factor.

Let's consider the case of the uniform magnetic field when the equation of motion simplifies to

$$m_0 \frac{d\gamma \mathbf{v}}{dt} = q\mathbf{v} \times \mathbf{B} \quad \text{or} \quad m_0 \gamma \dot{v}_x = q\, v_y B \quad \text{and} \quad m_0 \gamma \dot{v}_y = -q\, v_x B$$

which then can be rewritten as

$$\ddot{v}_x = \frac{qB}{m_0 \gamma} \dot{v}_y = -\left(\frac{qB}{m_0 \gamma}\right)^2 v_x$$

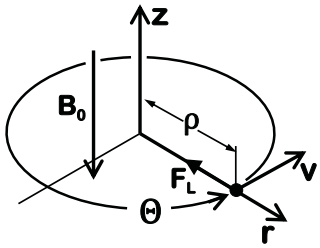



**FIGURE 2.3**
Motion of charged particles in a uniform magnetic field.

which has a solution of

$$v_x = v_0 \cos(\omega t) \quad \text{or} \quad x = \frac{v_0}{\omega} \sin(\omega t)$$

where

$$\omega = \frac{qB}{m_0 \gamma} \tag{2.11}$$

This solution describes motion (see Fig. 2.3) with a radius of

$$\rho = \frac{v_0}{\omega} = \frac{v_0 m_0 \gamma}{qB} \tag{2.12}$$

Let's rewrite this radius (called the Larmor radius) in both systems of units:

$$\text{SI}: \ \rho = \frac{p}{qB} \qquad \text{Gaussian}: \ \rho = \frac{pc}{qB} \tag{2.13}$$

This kind of motion is observed, for example, in dipoles — magnets intended primarily for bending the trajectories of charged particles.

*Magnetic rigidity* $B\rho\ [Tesla \cdot m] \approx 3.3356\ p\ [GeV/c]$

A quantity called *magnetic rigidity* $B\rho$ is often used to describe motion in magnetic fields. It is defined as

$$\text{SI}: \ B\rho = \frac{p}{q} \qquad \text{Gaussian}: \ B\rho = \frac{pc}{q} \tag{2.14}$$

and for a particle with the elementary charge and momentum $p$ given in $GeV/c$ is equal to $B\rho[Tesla \cdot m] \approx 3.3356\ p[GeV/c]$ or $B\rho[kGs \cdot cm] \approx 3335.6\ p[GeV/c]$.

### 2.3.2 Drift in crossed $\mathbf{E} \times \mathbf{B}$ fields

While we are on this topic, let's consider the case of uniform **E** and **B** fields that are perpendicular — a situation often encountered when dealing with plasma and beams.

Qualitatively, if a particle is initially at rest in these crossed fields, it is initially pulled by the electric field, and then, due to emerging velocity, the magnetic field turns it around. When the direction of the particle's motion reverses, the electric field slows it down and eventually stops the particle some distance away from its initial position. After that, the aforementioned motion begins anew. The resulting trajectory of the particle resembles Fig. 2.4. As a result, these equations of motion predict a particle drift with constant velocity, which is perpendicular to both **E** and **B** and with its value given by

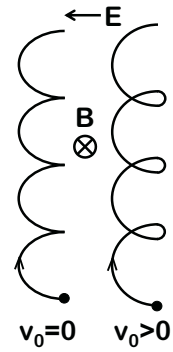



**FIGURE 2.4**
Drift in crossed E×B fields.

$$\text{SI}: \quad \mathbf{v}_d = \frac{\mathbf{E} \times \mathbf{B}}{B^2} \qquad \text{Gaussian}: \quad \mathbf{v}_d = c\frac{\mathbf{E} \times \mathbf{B}}{B^2} \tag{2.15}$$

The efficiency of the Gaussian system of units continues to astound us! Not only does it give us an intuitively clear and beautiful formula, but it also immediately shows that the above equation is valid only in the assumption that the electric field is much smaller than the magnetic field $E \ll B$.

### 2.3.3 Motion in quadrupole fields

Let's consider motion in a quadrupole magnet (see Fig. 2.5) where, ideally, the fields depend linearly on the distance from the center of the magnet:

$$B_x = Gz \quad \text{and} \qquad B_z = Gx$$

where $G$ is the gradient of the quadrupole (and we use $z \equiv y$ in this section). We will rewrite the equation of motion as

$$m_0 \frac{d\gamma \mathbf{v}}{dt} = q\,\mathbf{v} \times \mathbf{B}$$

and using Cartesian coordinates:

$$\ddot{x} = -\frac{q}{m_0\gamma} G\dot{s}x \quad , \quad \ddot{z} = \frac{q}{m_0\gamma} G\dot{s}z$$

$$\text{and} \quad \ddot{s} = \frac{q}{m_0\gamma} G(\dot{x}x - \dot{z}z)$$

Changing the independent variable from time to path length, and considering only the case of small deviations from the axis, reduces these equations to

$$x'' - Kx = 0 \quad \text{and} \quad z'' + Kz = 0 \tag{2.16}$$

$$\text{where} \quad K = \frac{q}{p}\frac{\partial B_z}{\partial x} = \frac{q}{p} G \tag{2.17}$$

This brings us to the discussion of linear betatron motion.

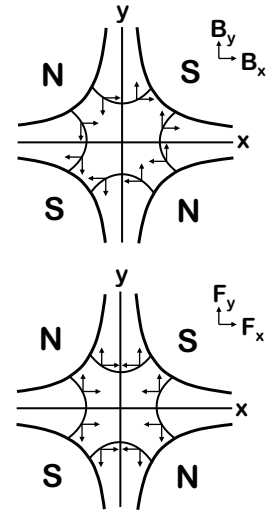



**FIGURE 2.5**
Magnetic fields and forces acting on a particle in a quadrupole.

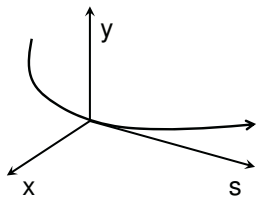



**FIGURE 2.6**
Frenet–Serret curvilinear coordinate system.

### 2.3.4 Linear betatron equations of motion

The particle trajectory in an accelerator is initially defined by dipole magnets, therefore a *curvilinear coordinate system* is best used to describe the motion of the particles; see Fig. 2.6. A *reference orbit* is usually selected, corresponding to an *ideal particle*, which typically has a nominal energy and zero transverse offsets and angles.

The focusing elements, quadrupoles and higher-order elements are placed in space so that their centers correspond to the reference orbit.

The motion of a charged particle with respect to the *reference orbit* and along the *curvilinear abscissa* ($s$ in Fig. 2.6), and influenced by the the magnetic fields of dipole magnets and quadrupole magnets, is given by the linear Hill's equations

$$\frac{d^2y}{ds^2} + K_y(s)y = 0 \tag{2.18}$$

where transverse coordinate $y$ stands for either a horizontal or vertical axis ($y = x, z$).

Let's write these equations down for the horizontal

$$K_x(s) = \frac{1}{\rho^2(s)} - \frac{1}{B\rho}\frac{\partial B_z(s)}{\partial x} \tag{2.19}$$

and the vertical

$$K_z(s) = \frac{1}{B\rho}\frac{\partial B_z(s)}{\partial x} \tag{2.20}$$

planes. We see that the equation almost exactly resembles those derived in the previous section (see Eq. 2.16 and Eq. 2.17) except for an additional term $1/\rho^2(s)$, which corresponds to a weak focusing of a dipole.

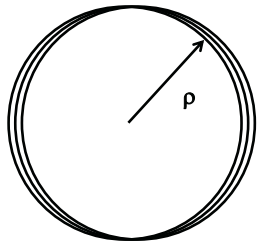



**FIGURE 2.7**
Shifted circles cross.

The origin of the term corresponding to the weak focusing in a dipole can be illustrated by the following example. Consider the shifted circles in Fig. 2.7. They cross in two points (we will ignore second-order effects). Translating this example of shifted circles to a dipole magnet whose field fills a half plane as shown in Fig. 2.8, we conclude by observation that the trajectories of the particles in this dipole exhibit an equivalent "focusing" with the wavelength of motion (along the curvilinear coordinate $s$) given by $2\pi\rho$, corresponding to

$$x = x_0 \sin(s/\rho)$$

or to the following equation

$$\frac{d^2x}{ds^2} + \frac{x}{\rho^2} = 0 \tag{2.21}$$

which has the same term $1/\rho^2(s)$ as the Hill's equations above.

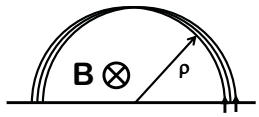



**FIGURE 2.8**
Illustration of the origin of weak focusing in dipoles.

The above derivations generally assumed no periodicity; however, in a circular machine, $K_x, K_z$ and $\rho$ are periodic. These are linear equations and can be integrated, which will be discussed in the next section.

## 2.4 Matrix formalism

### 2.4.1 Pseudo-harmonic oscillations

Let's look for the solution of the Hill's Eq. 2.18 in following form

$$y(s) = \sqrt{\varepsilon_y \beta_y(s)} \cos\left[\phi_y(s) - \phi\right] \tag{2.22}$$

where

$$\phi_y(s) = \int_{s_0}^{s} \frac{ds'}{\beta_y(s')} \tag{2.23}$$

which describes *pseudo-harmonic oscillations*. Here the beta functions $\beta$ (in $x$ and $z$) are proportional to the square of the *envelope* of the oscillations. The functions $\phi(s)$ (also in $x$ and $z$) describe the phase of the oscillations.

*Beta functions $\beta$ are proportional to the square of the envelope of the oscillations*

We can find the differential equation for the beta functions, by substituting the form Eq. 2.22 into the Hill's equation. We use the following equation first

$$y'(s) = \frac{\beta'(s)}{2}\sqrt{\frac{\varepsilon}{\beta(s)}}\cos(\phi(s)-\phi) - \phi'(s)\sqrt{\varepsilon\beta(s)}\sin(\phi(s)-\phi)$$

and the second derivatives

$$y''(s) = \left[\frac{\beta''(s)}{2\sqrt{\beta(s)}} - \frac{\beta'^2(s)}{4\beta^{3/2}(s)} - \sqrt{\beta(s)}\phi'^2(s)\right]\sqrt{\varepsilon}\cos(\phi(s)-\phi) -$$

$$-\left[\phi''(s)\sqrt{\beta(s)} + \frac{\beta'(s)\phi'(s)}{\sqrt{\beta(s)}}\right]\sqrt{\varepsilon}\sin(\phi(s)-\phi)$$

which we substitute to Hill's equation and proceed by equating the coefficients to zero in front of $sin$ and $cos$ parts. We therefore obtain

*Alpha function is defined as $\alpha = -\beta'/2$*

$$\frac{1}{2}\beta\beta'' - \frac{1}{4}\beta'^2 + k(s)\beta^2 = 1 \tag{2.24}$$

and

$$\phi_y'(s) = \frac{1}{\beta_y(s)} \tag{2.25}$$

which represents the differential equations for beta function and the betatron phase.

### 2.4.2 Principal trajectories

The solutions of Hill's equation can be found in the form of *principal trajectories*. These are two particular solutions of the homogeneous Hill's equation

$$y'' + k(s)y = 0$$

which satisfy the following initial conditions

$$C(s_0) = 1;\ C'(s_0) = 0;\ \text{cosine–like solution}$$

$$S(s_0) = 1;\ S'(s_0) = 1;\ \text{sine–like solution}$$

The general solution can then be written as a linear combination of the principal trajectories

$$y(s) = y_0 C(s) + y_0' S(s)$$

The pseudo-harmonic oscillations, and the cosine-like and sine-like principal trajectories are illustrated in Fig. 2.9 for a FODO chain of quadrupole magnets.

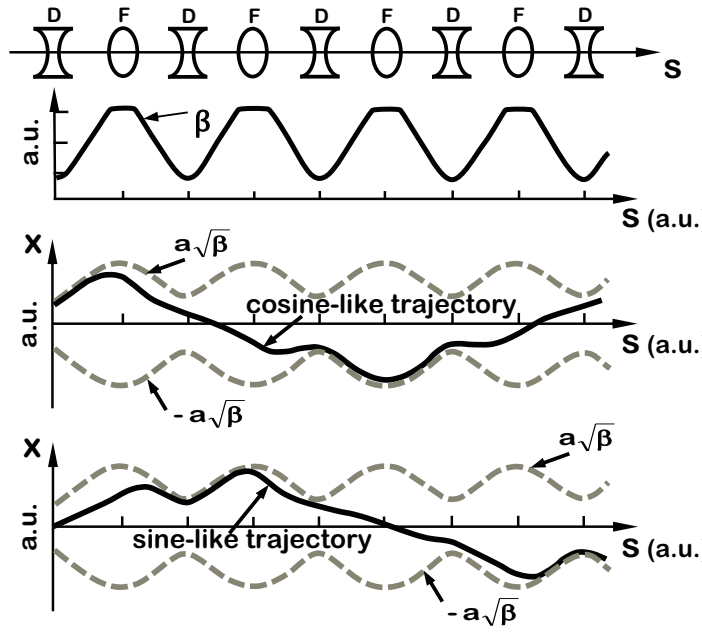



**FIGURE 2.9**
Illustration of pseudo-harmonic oscillations and cosine-like and sine-like principal trajectories.

Furthermore, we can derive the connection between the principal trajectories and pseudo harmonic oscillations. Let's express the amplitude and angle functions as

$$y(s) = \sqrt{\varepsilon\beta(s)}\,\cos\left(\phi(s) - \phi\right)$$

$$y'(s) = -\sqrt{\frac{\varepsilon}{\beta(s)}}\left[\sin\left(\phi(s) - \phi\right) + \alpha(s)\cos\left(\phi(s) - \phi\right)\right]$$

in terms of the principal trajectories

$$y(s) = y_0 C(s) + y_0' S(s)$$

Using simple algebraic manipulations, we find that

$$C(s) = \sqrt{\frac{\beta(s)}{\beta_0}}\left(\cos\phi(s) + \alpha_0 \sin\phi(s)\right)\ ,\quad S(s) = \sqrt{\beta(s)\beta_0}\,\sin\phi(s)$$

and correspondingly for the reverse formula

$$\phi(s) = \text{arctg}\frac{S(s)}{\beta_0 S(s) - \alpha_0 C(s)}$$

$$\beta(s) = \frac{1}{\beta_0}\left\{\frac{S^2(s) + [\beta_0 S(s) - \alpha_0 C(s)]^2}{\beta_0 S(s) - \alpha_0 C(s)}\right\}$$

or, in a simpler shape

$$\beta(s) = \frac{1}{\beta_0}\left[\frac{S(s)}{\sin\phi(s)}\right]^2$$

$$\alpha(s) = \frac{-S'(s)\sqrt{\frac{\beta(s)}{\beta_0}} + \cos\phi(s)}{\sin\phi(s)}$$

Having derived the above equations, we can now see that we can describe the evolution of the particle trajectories in a transfer line or in a circular accelerator by means of matrix formalism (see Fig. 2.10). In other words, linear transformations that are enabled by the linearity of the Hill's equations express as

$$\begin{pmatrix} y(s) \\ y'(s) \end{pmatrix} = \begin{pmatrix} C(s) & S(s) \\ C'(s) & S'(s) \end{pmatrix}\begin{pmatrix} y(s_0) \\ y'(s_0) \end{pmatrix} \qquad (2.26)$$

The matrix elements $C(s)$ and $S(s)$ depend only on the magnetic lattice and not on the initial conditions of the particle.

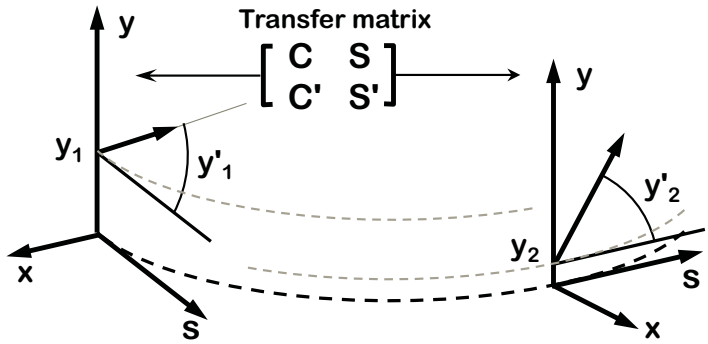



**FIGURE 2.10**
Linear matrix approach for evaluation of the evolution of the particle coordinates in a transfer line.

The transfer matrix is therefore given by

$$\mathbf{M}_{1\to 2} = \begin{pmatrix} C(s) & S(s) \\ C'(s) & S'(s) \end{pmatrix} \qquad (2.27)$$

The described approach allows for the possibility of using the matrix formalism to describe the evolution of the coordinates of a charged particle in a magnetic lattice.

### 2.4.3 Examples of transfer matrices

The most common elements in accelerators are *drifts*, *bending magnets*, which steer the trajectory, *quadrupoles*, which provide transverse focusing, and the *accelerating section*, which accelerates the beam. Each of these elements can be represented by a particular transfer matrix. Let's consider here a couple of the simplest examples. The transfer matrix of a *drift space* is simply

$$\mathbf{M} = \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \tag{2.28}$$

Bending magnets, especially in large accelerators where each individual bend provides only a tiny bit of bending, have transfer matrices very close to those of drift space except when *edge focusing* (illustrated in Fig. 2.11) needs to be taken into account. In the latter, the faces of the bending magnet are not perpendicular to the reference trajectory. As clearly seen from the figure, this can provide additional defocusing on the horizontal plane[1] due to edge effects.

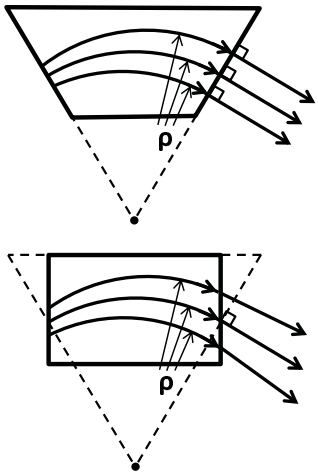



**FIGURE 2.11**
Mechanism of the edge focusing of bending magnet in a horizontal plane.

The matrix of a focusing quadrupole is

$$\mathbf{M} = \begin{pmatrix} \cos(\sqrt{|K|}L) & \frac{1}{\sqrt{|K|}}\sin(\sqrt{|K|}L) \\ -\sqrt{|K|}\sin(\sqrt{|K|}L) & \cos(\sqrt{|K|}L) \end{pmatrix} \tag{2.29}$$

and the matrix of a defocusing quadrupole is

$$\mathbf{M} = \begin{pmatrix} \cosh(\sqrt{K}L) & \frac{1}{\sqrt{K}}\sinh(\sqrt{K}L) \\ \sqrt{K}\sinh(\sqrt{K}L) & \cosh(\sqrt{K}L) \end{pmatrix} \tag{2.30}$$

For a thin lens, $L \to 0$ with $KL$ staying finite, these matrices correspondingly become

$$\mathbf{M}_F = \begin{pmatrix} 1 & 0 \\ -|K|L & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{M}_D = \begin{pmatrix} 1 & 0 \\ KL & 1 \end{pmatrix} \tag{2.31}$$

The higher-order elements such as sextupole and octupole magnets are often used in accelerators for nonlinear corrections. In linear approximation their transfer matrices are equivalent to the corresponding matrix of a drift space. They will have, however, higher-order terms and can be described by higher-order matrices (as will be discussed later in this chapter).

### 2.4.4 Matrix formalism for transfer lines

The matrix formalism is very practical for computing propagation through transfer lines, especially since the transfer matrix of each individual element of the beamline needs to be calculated only once. In a purely linear transfer matrix, the

[1] Vertical plane edge focusing can also occur but only in the case of a finite vertical gap of the bending magnet.

overall linear matrix of the beamline, computed as a step-by-step matrix multiplication of all individual elements, would satisfactorily describe the propagation of particles through the beamline.

As a practical example, let's consider a pair of thin quadrupoles of different polarity separated by a drift space as shown in Fig. 2.12. The overall horizontal transfer matrix of such a system is given by

$$\mathbf{M}_x^{1\to 2} = \begin{pmatrix} 1 & 0 \\ \frac{1}{f_2} & 1 \end{pmatrix}\begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\frac{1}{f_1} & 1 \end{pmatrix} = \begin{pmatrix} 1-\frac{L}{f_1} & L \\ -\frac{1}{f^*} & 1+\frac{L}{f_2} \end{pmatrix}$$

$$\text{where} \quad \frac{1}{f*} = \frac{1}{f_1} - \frac{1}{f_2} + \frac{L}{f_1 f_2} \tag{2.32}$$

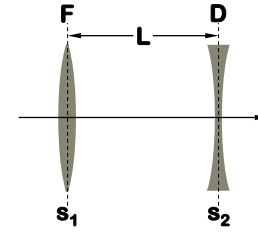



**FIGURE 2.12**
For illustration of transfer matrix of two quadrupoles separated by a drift.

The quantity $f^*$ can be considered as an effective focal distance of a system of two lenses — such a system is usually called a focusing doublet. The overall vertical transfer matrix $\mathbf{M}_y^{1\to 2}$ is obtained by reversing the signs of $f_1$ and $f_2$. By referring to Eq. 2.32, we can see that there is a region of parameters where the sign of $f^*$ is the same and positive for both the horizontal and vertical planes (for example, when $f_1 = f_2$), which corresponds to focusing in both planes.

### 2.4.5 Analogy with geometric optics

As we can see, the particle trajectories can be described using matrix formalism in a very similar way to ray propagation in an optical system. The magnetic quadrupoles play the role of focusing and defocusing lenses but, unlike an optical lens, a magnetic quadrupole focuses in one plane and defocuses in the other plane due to the nature of Maxwell's equations. However, as we have just shown in the previous section, a doublet of quadrupoles can focus simultaneously in both planes, as illustrated in Fig. 2.13.

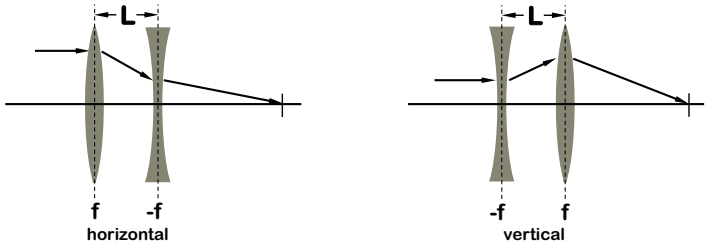



**FIGURE 2.13**
A doublet can focus simultaneously in both planes.

The similarity and difference with geometrical optics can be highlighted by the following example. Let's consider an optical telescope consisting of two lenses as shown in Fig. 2.14. It is intuitively clear and in fact can be proven that

*An optical telescope with two lenses can provide arbitrary demagnification.*

two lenses (in linear approximation), properly spaced and located, can provide an arbitrary *demagnification*.

In a direct analogy to geometrical optics, two focusing doublets are needed in order to create a telescope with arbitrary demagnification in the case of magnetic element optics (i.e., four quadrupoles appropriately placed and spaced).

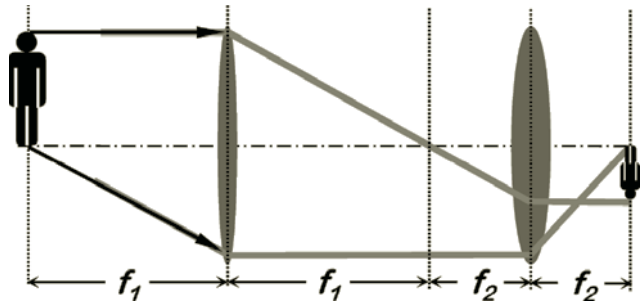



**FIGURE 2.14**
Optical telescope with two lenses.

The comparison of geometrical optics to magnetic element optics is a powerful method that often helps for back-of-the-envelope evaluations of various optical systems.

*Four quadrupoles are needed to create a telescope with arbitrary demagnification for charged particle optics.*

### 2.4.6 An example of a FODO lattice

Let's consider one more practical example — an alternating sequence of focusing (F) and defocusing (D) quadrupoles separated by a drift (O) — this is a so-called FODO lattice; see Fig. 2.15.

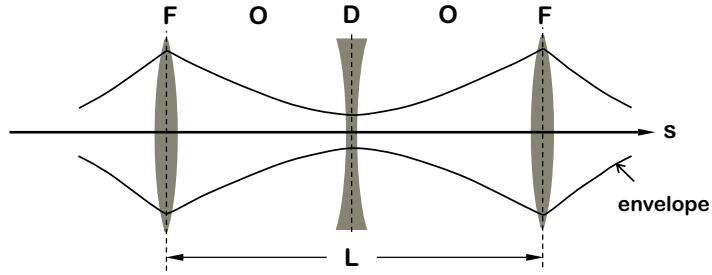



**FIGURE 2.15**
FODO lattice.

The transfer matrix of the FODO cell can be derived as

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}\begin{pmatrix} 1 & \frac{L}{2} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \frac{1}{f} & 1 \end{pmatrix}\begin{pmatrix} 1 & \frac{L}{2} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+\frac{L}{2f} & L\left(1+\frac{L}{4f}\right) \\ -\frac{L}{2f^2} & 1-\frac{L}{2f}-\frac{L^2}{4f^2} \end{pmatrix}$$

We will use this expression in the following section to evaluate beam stability.

### 2.4.7 Twiss functions and matrix formalism

The optical functions — beta, alpha and gamma (defined below) are called Twiss functions. Using the formulas defined in the previous sections, the matrix elements can be expressed via the optics functions at the beginning and end of the beamline:

$$\mathbf{M}_{s_0\to s} = \begin{pmatrix} C(s) & S(s) \\ -C'(s) & S'(s) \end{pmatrix} = \tag{2.33}$$

$$= \begin{pmatrix} \sqrt{\frac{\beta(s)}{\beta_0}}(\cos\Delta\phi + \alpha_0 \sin\Delta\phi) & \sqrt{\beta(s)\beta_0}\sin\Delta\phi \\ -\frac{(\alpha(s)-\alpha_0)\cos\Delta\phi+(1+\alpha(s)\alpha_0)\sin\Delta\phi}{\sqrt{\beta(s)\beta_0}} & \sqrt{\frac{\beta_0}{\beta(s)}}[\cos\Delta\phi - \alpha(s)\sin\Delta\phi] \end{pmatrix}$$

Here $\beta_0$, $\alpha_0$ and the phase $\phi_0$ (in $\Delta\phi = \phi - \phi_0$) correspond to the beginning of the transfer line. The expressions above are called *Twiss parameterization* of the transfer matrices.

So far, we haven't yet assumed any periodicity in the transfer line. However, if we now consider a periodic machine, then the transfer matrix over a single turn (single turn map) would reduce to

$$\mathbf{M}_{s_0\to s_0} = \begin{pmatrix} \cos\mu + \alpha_0 \sin\mu & \beta_0 \sin\mu \\ -\gamma_0 \sin\mu & \cos\mu - \alpha_0 \sin\mu \end{pmatrix} \tag{2.34}$$

where the gamma function is defined as

$$\gamma_0 = \frac{1+\alpha_0^2}{\beta_0} \tag{2.35}$$

and where we used $\mu = \Delta\phi$ to define the phase advance for one turn.

### 2.4.8 Stability of betatron motion

Having considered periodic transfer maps in the previous section, we are now ready to discuss stability of the multi-turn motion.

Consider a circular accelerator with a transfer matrix, which for one turn equals to $\mathbf{M}$. Let's rewrite the Twiss parameterization for $\mathbf{M}$ given by Eg.2.34 as

$$\mathbf{M} = \cos\mu \cdot \mathbf{I} + \sin\mu \cdot \mathbf{J} \tag{2.36}$$

$$\text{where } \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{J} = \begin{pmatrix} \alpha_0 & \beta_0 \\ -\gamma_0 & -\alpha_0 \end{pmatrix}$$

After $n$ turns, the particle coordinates will be given by the successive application of the one-turn transformation matrix $n$ times, as follows:

$$\mathbf{x}_1 = \mathbf{M}\mathbf{x}_0 \quad ... \quad \mathbf{x}_2 = \mathbf{M}^2\mathbf{x}_0 \quad ... \quad \mathbf{x}_n = \mathbf{M}^n\mathbf{x}_0$$

The beauty of the parametrization given in Eq. 2.36 is that, as

can be easily proven, $\mathbf{J}^2 = -\mathbf{I}$ and thus $\mathbf{M}^2 = \cos 2\mu \cdot \mathbf{I} + \sin 2\mu \cdot \mathbf{J}$. Similarly, one can show that

$$\mathbf{M}^n = \begin{pmatrix} \cos n\mu + \alpha_0 \sin n\mu & \beta_0 \sin n\mu \\ -\gamma_0 \sin n\mu & \cos n\mu - \alpha_0 \sin n\mu \end{pmatrix} \tag{2.37}$$

Observing the expression for this periodic transverse map, one can conclude that stability of the transverse motion necessarily requires the phase advance $\mu$ to be a real number, which ensures that the multi-turn motion represents stable oscillations. The condition of $\mu$ being real can be re-written as $|\cos\mu| < 1$ or as a more general expression involving the *trace* or *spur* (sum of its diagonal elements) of the transfer matrix

$$|\cos\mu| = \frac{1}{2}|tr\,\mathbf{M}| < 1 \tag{2.38}$$

The criteria defined above is a necessary condition for a transfer line to be suitable for multi-turn stable dynamics. We are now ready to apply this criteria to a practical example.

### 2.4.9 Stability of a FODO lattice

Let's apply the stability criteria expressed as Eq. 2.38 to the Twiss parameterization of the matrix or the FODO cell derived in Section 2.4.6

$$\mathbf{M} = \begin{pmatrix} 1 + \frac{L}{2f} & L\ \ 1 + \frac{L}{4f} \\ -\frac{L}{2f^2} & 1 - \frac{L}{2f} - \frac{L^2}{4f^2} \end{pmatrix} = \begin{pmatrix} \cos\mu + \alpha \sin\mu & \beta \sin\mu \\ -\gamma \sin\mu & \cos\mu - \alpha \sin\mu \end{pmatrix}$$

The trace of this transfer matrix is given by

$$tr\,\mathbf{M} = 2 - \frac{L^2}{4f^2}$$

And the stability criteria thus requires

$$|\cos\mu| = \left|1 - \frac{L^2}{8f^2}\right| < 1 \quad \text{or} \quad f > \frac{L}{4} \tag{2.39}$$

The resulting criteria $f > L/4$ is intuitively very clear and it can also be understood from the analogy with geometric optics and from considerations of the behavior of the beam envelope. Looking at Fig. 2.15, one can observe that $f = L/4$ would correspond to the situation when the size of the envelope in the defocusing quadrupole approaches zero, and even stronger quadrupoles (i.e. lower $f$) would make it impossible to sketch a repeatable finite envelope.

### 2.4.10 Propagation of optics functions

As we have discussed above, the coordinates of the particles can be propagated via a transfer line using the matrices of the

transfer line defined by the principal trajectories

$$\mathbf{M}_{1\to 2} = \begin{pmatrix} C(s) & S(s) \\ C'(s) & S'(s) \end{pmatrix} \tag{2.40}$$

Similarly, we can write down the expression that computes the propagation of the optics function along the transfer lines using a matrix based on the principal trajectories

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} C^2 & -2CS & S^2 \\ -CC' & CS'+SC' & -SS' \\ C'^2 & -2C'S' & S'^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix} \tag{2.41}$$

The initial values of the optical functions in this equation are either determined by the periodicity conditions as in the case of a circular machine, or correspond to the initial values at the entrance of the system as in the case of a transfer line.

Let's consider two cases as examples. Drift space expresses as

$$\mathbf{M} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} 1 & -2s & s^2 \\ 0 & 1 & -s \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix}$$

We can see that the $\beta$ function has a parabolic behavior in correlation to the drift length.

A thin focusing quadrupole of focal length $f = 1/KL$

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ KL & 1 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ KL & 1 & 0 \\ (KL)^2 & 2KL & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \alpha_0 \\ \gamma_0 \end{pmatrix}$$

In this case, it is the $\gamma$ function that has a parabolic behavior in correlation to the inverse focal length of the quadrupole.

## 2.5 Phase space

### 2.5.1 Phase space ellipse and Courant–Snyder invariant

To summarize our discussion about betatron motion, let's discuss the evolution of a phase space ellipse.

The solutions of the Hill's equation

$$\frac{d^2y}{ds^2} + K_y(s)y = 0 \tag{2.42}$$

discussed in Section 2.4.1 is reproduced below

$$y(s) = \sqrt{\varepsilon\beta(s)}\cos\left(\phi(s) - \phi\right) \tag{2.43}$$

$$y'(s) = -\sqrt{\frac{\varepsilon}{\beta(s)}}\left[\sin\left(\phi(s) - \phi\right) + \alpha(s)\cos\left(\phi(s) - \phi\right)\right]$$

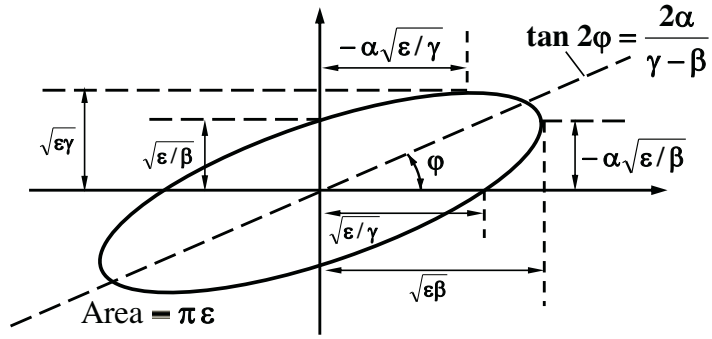



**FIGURE 2.16**
Betatron motion in phase space.

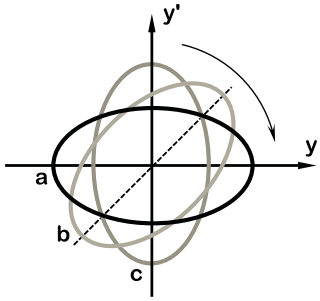



**FIGURE 2.17**
Evolution of phase-space ellipse. Locations: (a) in D and (c) in F quadrupoles, and (b) in between.

This solution describes an evolution of an ellipse in phase space $(y, y')$. The parameters of the ellipse are described in Fig. 2.16.

Hill's equations have a remarkable property — they have an invariant:

$$A(s) = \beta y'^2 + 2\alpha y y' + \gamma^2 y^2 = const. = \varepsilon \qquad (2.44)$$

This can be proven by substituting the solutions of Hill's equations Eq. 2.43 into Eq. 2.44 for $A(s)$.

The quantity $A(s)$ is called the *Courant–Snyder invariant* and is connected to the area of the ellipse in phase space with a factor of $\pi$ as $Area = \pi\varepsilon$.

The Courant–Snyder invariant — and thus the area of the ellipse — stay constant independent of the optics of the beamline. As illustrated in Fig. 2.17, the ellipse rotates and its shape may change while its area remains invariant.

## 2.6 Dispersion and tunes

### 2.6.1 Dispersion

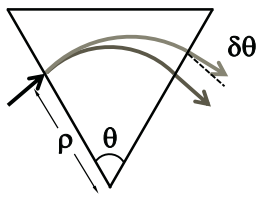



**FIGURE 2.18**
Bending magnet creates dispersion.

We have so far assumed that the particles of the beam have a nominal energy equal to that of the energy of the reference particle. In practice, however, there is always some energy offset or energy spread within the beam. The function that characterizes the orbit of an off-energy particle in an accelerator is called the *dispersion function*.

The primary effect of the energy offset is the difference of the trajectory in bending magnets — as illustrated in Fig. 2.18. After the propagation of an off-energy particle in a magnet, both an offset and angle of the orbit are created. In linear approximation, the radius of curvature of the trajectory for an off-energy particle expresses as

$$\frac{1}{\rho(s)} = \frac{eB}{p} = \frac{eB}{p_0\ 1+\frac{\Delta p}{p_0}} \approx \frac{1}{\rho_0(s)}\left(1-\frac{\Delta p}{p_0}\right) \qquad (2.45)$$

Substituting this to Eq. 2.19 gives us the Hill's equation for an off-energy particle:

$$\frac{d^2x}{ds^2} - \left(K(s) - \frac{1}{\rho^2}(s)\right)x = \frac{1}{\rho(s)}\frac{\Delta p}{p_0} \tag{2.46}$$

The answer to the above equation can be found using the following form:

$$x = x_0 + \frac{dp}{p}D \tag{2.47}$$

where $D$ is the dispersion function and $x_0$ describes betatron oscillation around the dispersive orbit as shown in Fig. 2.19.

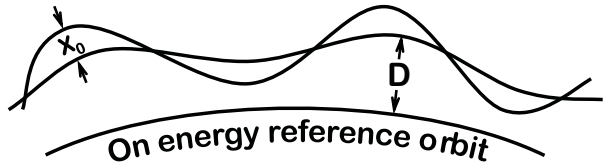



**FIGURE 2.19**
Dispersion.

Substituting this to Eq. 2.46 will yield the following equation, which governs the evolution of the dispersion function:

$$\frac{d^2D}{ds^2} - \left(K(s) - \frac{1}{\rho^2(s)}\right)D = \frac{1}{\rho(s)} \tag{2.48}$$

Dispersion function $D$ can also be expressed in terms of the principal trajectories as

$$D(s) = S(s)\int_{s_0}^{s}\frac{C(t)}{\rho(t)}dt - C(s)\int_{s_0}^{s}\frac{S(t)}{\rho(t)}dt \tag{2.49}$$

$$D'(s) = S'(s)\int_{s_0}^{s}\frac{C(t)}{\rho(t)}dt - C'(s)\int_{s_0}^{s}\frac{S(t)}{\rho(t)}dt \tag{2.50}$$

We have assumed in this section that all bending occurs in a horizontal plane and therefore only a horizontal dispersion function is nonzero. This may not be the case in instances of vertically bending magnets or coupling, discussed below.

As dispersion affects the space location of the reference orbit for off-energy particles, it thus also affects the orbit's path length. Defining C as circumference or orbit path length in curvilinear coordinates, the deviation of the path length can be shown to be given by

$$\Delta C = \frac{dp}{p}\int\frac{D(s)}{\rho(s)}ds \tag{2.51}$$

We can also compute the so-called *momentum compaction factor* with

$$\alpha_c = \frac{dC/C}{dp/p} = \frac{1}{C} \quad \frac{D(s)}{\rho(s)} ds \tag{2.52}$$

which will be discussed further in Chapter 5, which deals with longitudinal dynamics.

### 2.6.2 Betatron tunes and resonances

Taking into account the definition of the betatron phase in Eq. 2.23, we can write the phase advance over one turn of a circular machine as

$$\Delta\phi_C = 2\pi Q = \quad \frac{ds}{\beta(s)} \tag{2.53}$$

The quantity $Q$ is called the *betatron tune*. The betatron tunes are essential quantities used to analyze the stability of a circular accelerator. In particular, if $Q$ is an integer number, *resonance* conditions occur, as a tiny disturbance at some place along the orbit would repeat for many turns, accumulating into a large disruption of the particle's motion.

The general equations for resonance conditions of the betatron tunes can be written as follows:

$$m\, Q_x + n\, Q_z = k \tag{2.54}$$

where $m$, $n$ and $k$ are integer numbers and where $|m| + |n|$ is called the order of the resonance. Resonances of the lowest orders are the most dangerous for the stability of the particle motion and thus must be carefully avoided by proper machine optics design.

## 2.7 Aberrations and coupling

### 2.7.1 Chromaticity

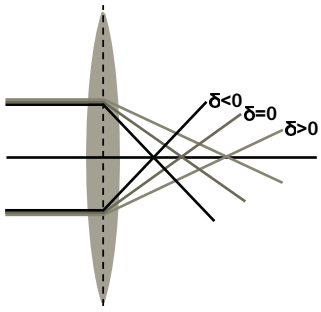



**FIGURE 2.20**
Chromaticity of a focusing quadrupole.

Offsets of energy in the particles cause not only dispersion but also result in different focusing strengths of the magnetic elements (as illustrated in Fig. 2.20).

The quadrupole strength for off-energy particles in the first order can be approximated as follows

$$k_1 = \frac{e}{p}\frac{\partial B_y}{\partial x} = \frac{e}{p_0(1+\delta)}\frac{\partial B_y}{\partial x} = \frac{k_0}{1+\delta} \approx k_0(1-\delta) \tag{2.55}$$

Taking into account the equations that describe beta function evolution, we can show that the betatron tunes shift in correlation to changes in focusing strength:

$$\Delta Q = \frac{1}{4\pi} \quad \beta(s)\Delta K(s) ds \tag{2.56}$$

Using the above approximation for the off-energy quadrupole strength, we thus write the expression for *chromaticity* $Q'$ as describing the dependence of the betatron tune on the energy offset of the particle and define it as the derivative of the betatron tunes with respect to the relative energy change:

$$Q' = \frac{dQ}{d\delta} = -\frac{1}{4\pi} \quad \beta(s)k_0(s)ds \tag{2.57}$$

### 2.7.2 Coupling

Throughout this chapter we have assumed that the motions of particles in horizontal and vertical planes are independent. This could indeed be the case if machine optics consist of bending magnets and quadrupoles that are perfectly placed in space. However, any rotational misalignments of these elements can create *coupling* of the horizontal and vertical motions.

Coupling can also be created by other magnetic elements such as solenoids (especially strong coupling can occur when a solenoid overlaps with quadrupole field, mixing different types of symmetry), or by misaligned nonlinear magnets such as sextupoles or octupoles, etc.

Some amount of coupling is unavoidable in a real machine and it usually needs to be corrected. A standard way to correct coupling (or to create it on purpose if needed) is to use skew quadrupoles — these are standard quadrupoles rotated by 45° as shown in Fig. 2.21.

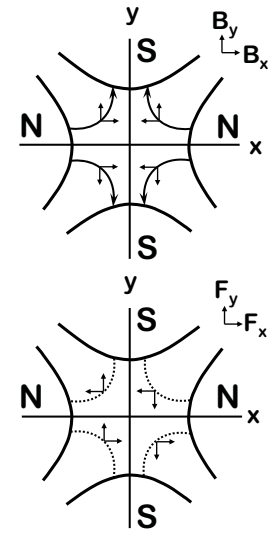



**FIGURE 2.21**
Skew quad fields and forces.

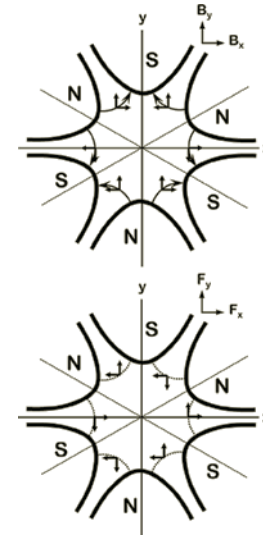



**FIGURE 2.22**
Sextupole fields and forces.

### 2.7.3 Higher orders

In storage rings, chromaticity is defined as a dependence of the betatron tunes on energy.

In single-path beamlines, it is more convenient to use other definitions. Let us first recall the linear matrix approach

$$x_i^{out} = R_{ij}\ x_j^{in} \tag{2.58}$$

this time noting all six components of the vector of interest, adding to the two coordinates and their angles the longitudinal offset $\Delta l$ as well as the energy offset $\delta$

$$x_i = (x, x', y, y', \Delta l, \delta)' \tag{2.59}$$

The second, third and other higher terms that can result from nonlinear elements such as a sextupole shown in Fig. 2.22 or octupole (Fig. 2.23) can be included in the matrix formalism in a similar manner:

$$x_i^{out} = R_{ij}\ x_j^{in} + T_{ijk}\, x_j^{in}\, x_k^{in} + U_{ijkn}\, x_j^{in}\, x_k^{in}\, x_n^{in} + \ldots \tag{2.60}$$

where $T$ and $U$ are the second and third-order matrices.

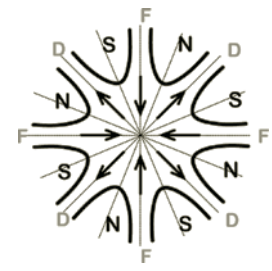



**FIGURE 2.23**
Octupole magnet forces.

Unlike in the storage rings, in transfer line design, one usually calls chromaticity the second-order elements $T_{126}$ and $T_{346}$. All other high-order terms are just *aberrations*, purely chromatic (like $T_{166}$, which is second-order dispersion), or chromo-geometric (like $U_{32446}$).

## EXERCISES

2.1 *Chapter materials review*.
Define the region of parameters where a pair of thin quadrupoles will focus the beam in both planes.

2.2 *Chapter materials review*.
Prove the Eq. 2.39, which defines the stability of a FODO beamline, geometrically, using the analogy with a traditional geometrical optics.

2.3 *Chapter materials review*.
A parallel proton beam of E=200 MeV enters a beamline. It is necessary to focus this beam into a point at a 3 m distance from the entrance. Estimate the necessary parameters of a quadrupole system (gradients, lengths) that can perform this task.

2.4 *Mini-project*.
Consider the same proton beam as in the previous exercise, as well as the same focusing requirements. Assume that the focusing is performed by a continuous, cylindrical electron beam. Estimate the necessary electron density which can perform the focusing task. Select electron beam energy, determine the electron current and discuss and select an optimal design of the electron beam system, as well as its feasibility.

2.5 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this chapter in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

2.6 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 3 Synchrotron Radiation



In this chapter we will consider one of the most important phenomena that governs the behavior of accelerators — synchrotron radiation (SR).

SR can be both helpful, as it yields the creation of high brightness radiation sources, and harmful, as it can deteriorate the beam by creating additional energy spread and beam emittance growth.

Traditional derivations of SR equations are rather mathematically involved. However, in this chapter we will use simplified back-of-the-envelope style derivations, which nevertheless obtain all of the important characteristics of SR with high accuracy.

## 3.1 SR on the back of an envelope

In our simple picture, the SR is the result of the charged particle leaving part of its fields behind when it is moving on a curved trajectory. The part of the field that is left behind (or radiated) cannot catch up with the motion of the particle, as it cannot move faster than the speed of light.

Armed with this concept, let's estimate the power loss due to SR, the typical energy of the emitted photons, and other parameters of synchrotron radiation as well as the most important effects that SR inflicts on the beam.

### 3.1.1 SR power loss

The straightforward concept of SR described above is represented in Fig. 3.1. In this instance, the particle moving with velocity $v$ (which is close to the speed of light) on a radius $R$ has its field lines pointing mostly transversely, and the part of field moving further away (on the radius $R + r$) would be left behind, as it cannot move faster than $c$.

The radius $r$ can be evaluated as

$$r = R\left(\frac{c}{v} - 1\right) \approx \frac{R}{2\gamma^2} \tag{3.1}$$

where we assumed that $\gamma \gg 1$ or $\beta \approx 1$ and thus $(1 - v/c) = (1-\beta)(1+\beta)/(1+\beta) \approx (1-\beta^2)/2 = 1/(2\gamma^2)$.

DOI: 10.1201/b18696-3

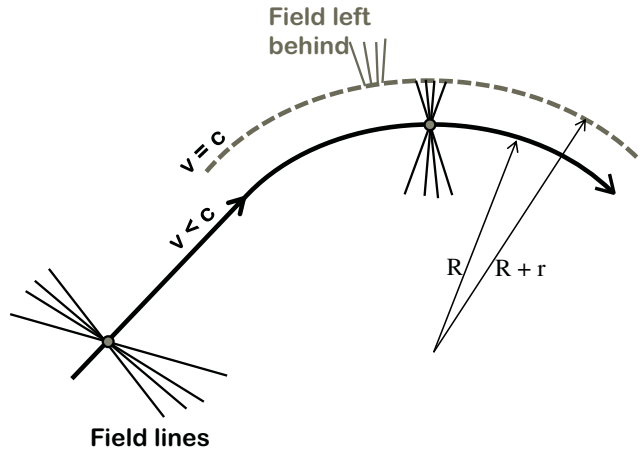



**FIGURE 3.1**
Synchrotron radiation — conceptual explanation.

The energy in the field that is left behind (radiated) can be estimated (we use Gaussian units in this section) as a volume integral of the field squared:

$$W \approx \int E^2 \, dV \tag{3.2}$$

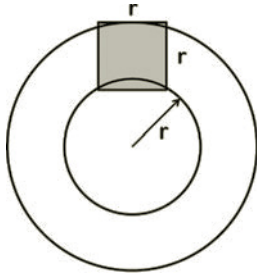



**FIGURE 3.2**
Illustrating the characteristic volume used in Eq. 3.3.

The field $E$ can be estimated as the field on the radius $r$ from the particle

$$E \approx \frac{e}{r^2}$$

and the characteristic volume can be estimated (see Fig. 3.2) as

$$V \approx r^2 \, ds \tag{3.3}$$

where $ds$ is an element of the path along the orbit, as we will eventually wish to find the power lost per unit of length.

The energy loss per unit length can thus be written as:

$$\frac{dW}{ds} \approx E^2 r^2 \approx \left(\frac{e}{r^2}\right)^2 r^2 \tag{3.4}$$

and after substituting

$$r \approx \frac{R}{2\gamma^2}$$

we get an estimate of:

$$\frac{dW}{ds} \approx \frac{e^2 \gamma^4}{R^2} \tag{3.5}$$

which we can compare with the exact formula:

$$\frac{dW}{ds} = \frac{2}{3} \frac{e^2 \gamma^4}{R^2} \tag{3.6}$$

and conclude that our rough estimations give a very reasonable result.

### 3.1.2 Cooling time

Knowledge of the SR-caused power losses immediately allow us to estimate the SR cooling time of the beam in a storage ring.

First of all, let's — for convenience's sake — rewrite the exact formula for power lost per unit length

$$\frac{dW}{ds} = \frac{2}{3}\frac{e^2\gamma^4}{R^2}$$

in the way that does not depend on the systems of units:

$$\frac{dW}{ds} = \frac{2}{3}\frac{r_e\gamma^4}{R^2}mc^2 \tag{3.7}$$

The SR energy loss of a particle per turn is therefore:

$$U_0 = \frac{4\pi}{3}\frac{r_e\gamma^4}{R}mc^2 \tag{3.8}$$

The effect of particle cooling due to SR is based on the fact that when an electron radiates a photon, its momentum decreases. Taking into account that while the beam of particles can have a range of angles within the beam, the accelerating RF cavity would restore only the longitudinal part of momentum, whereas the transverse degrees of freedom of the particles will be cooled down as illustrated in Fig. 3.3.

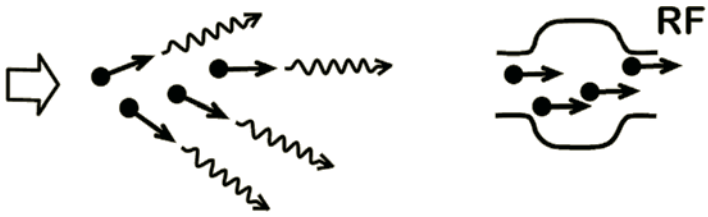



**FIGURE 3.3**
RF cavity restores only longitudinal momentum, thus other degrees of freedom are cooled due to synchrotron radiation.

Let's estimate the cooling time $\tau$ of a particle with energy $E_0$ cycling with revolution period $T_0$ in a circular machine as $\tau = E_0 T_0/U_0$, which yields

$$\tau \approx \frac{2\pi R}{c}\frac{\gamma mc^2}{U_0} \tag{3.9}$$

After substitution, the inverse cooling time can be written as

$$\tau^{-1} \approx \frac{2}{3}\frac{c\,r_e\gamma^3}{R^2} \tag{3.10}$$

### 3.1.3 Cooling time and partition

In the previous section we estimated the inverse cooling time as $\tau^{-1} \approx 2 c r_e \gamma^3/(3R^2)$. Traditionally, there is a factor of 2 in the definition in the cooling time:

$$\tau = 2E_0 T_0/U_0 \quad \Rightarrow \quad \tau^{-1} = \frac{1}{3}\,\frac{c r_e \gamma^3}{R^2} \tag{3.11}$$

We will use this latter definition in this section.

We can express the evolution of the beam emittance under the influence of an SR damping as

$$\varepsilon(t) = \varepsilon_0 \exp(-2t/\tau) \tag{3.12}$$

Both transverse planes, as well as the longitudinal motion in rings, are usually coupled. Thus we can expect that the damping will be distributed between these degrees of freedom in some proportion depending on details of the optics.

Distribution of cooling between the degrees of freedom is defined by the so-called partition numbers $J_x$, $J_y$ and $J_E$, which we mention here without derivations. The cooling time of a degree of freedom is correspondingly

$$\tau_i = \frac{\tau}{J_i} \tag{3.13}$$

The total radiated power due to SR is fixed and constant, therefore

$$\sum \tau_i^{-1} = const. \tag{3.14}$$

which corresponds to the *partition theorem*

$$\sum J_i = 4 \tag{3.15}$$

for a typical accelerator

$$J_x \approx 1\,, \quad J_y \approx 1\,, \quad J_E \approx 2 \tag{3.16}$$

and adjusting the optics of the machine changes the distribution of the partition numbers.

### 3.1.4 SR photon energy

*For $\gamma \gg 1$ the emitted photons go into $1/\gamma$ cone.*

In order to estimate the typical energy of the SR photons, we need to make an assumption that is based on relativistic kinematics: the radiation of relativistic particles is emitted into a cone with angular spread of $1/\gamma$.

Let's take this assumption into account when examining the radiation emitted during motion along the curved trajectory shown in Fig. 3.4 and ask a question — during what time interval $\Delta t$ would the remote observer see the emitted fields?

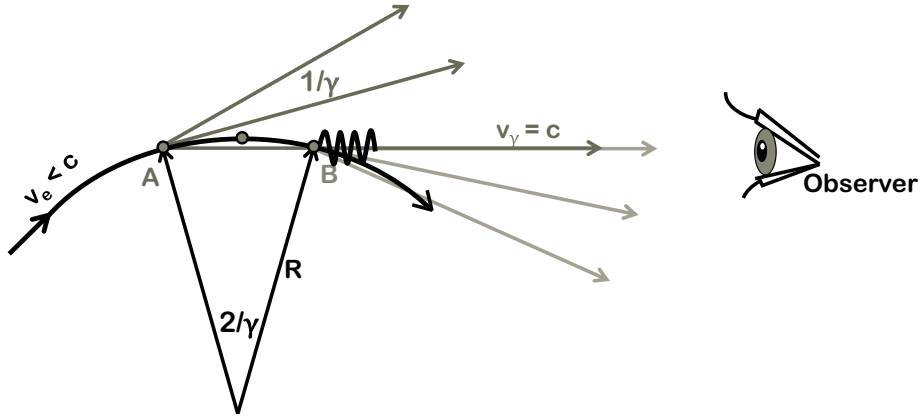



**FIGURE 3.4**
SR and remote observer.

Looking at this figure we conclude that the emitted radiation is observed during particle travel along the $2R/\gamma$ arc.

Let's now keep in mind that the radiation travels at speed $c$, while particles travel at $v$. At point B shown in Fig. 3.4 the separation between radiation and the particles is given by

$$ds \approx \frac{2R}{\gamma} \quad 1 - \frac{v}{c} \tag{3.17}$$

Therefore, the observer will see radiation during the following time interval:

$$\Delta t \approx \frac{ds}{c} \approx \frac{2R}{c\gamma}(1-\beta) \approx \frac{R}{c\gamma^3} \tag{3.18}$$

We can thus proceed to estimate the characteristic frequency of emitted photons as the inverse of the time duration of the flash, as seen by the observer:

$$\omega_c \approx \frac{1}{\Delta t} \approx \frac{c\gamma^3}{R} \tag{3.19}$$

Comparing the above with the exact formula that we reproduce here without derivation

$$\omega_c = \frac{3}{2}\frac{c\gamma^3}{R} \tag{3.20}$$

we can again see that our back-of-the-envelope estimations give pretty accurate results while not hiding the physics of the phenomena behind heavy math.

### 3.1.5 SR — number of photons

Having estimated the characteristic energy of the SR photons, we can now estimate the number of photons emitted per unit length by a single electron.

Let's use our estimation for the rate of energy loss $dW/ds \approx e^2\gamma^4/R^2$ and the estimation of the characteristic frequency of photons $\omega_c \approx c\gamma^3/R$, and rewrite the latter in terms of the photon energy:

$$\varepsilon_c = \hbar\omega_c \approx \frac{\gamma^3 \hbar c}{R} = \frac{\gamma^3}{R}\lambda_e mc^2 \tag{3.21}$$

where

$$r_e = \frac{e^2}{mc^2} \qquad \alpha = \frac{e^2}{\hbar c} \qquad \lambda_e = \frac{r_e}{\alpha}$$

The number of photons emitted per unit length can be obtained by dividing the energy loss per unit length by the energy of the photons

$$\frac{dN}{ds} \approx \frac{1}{\varepsilon_c}\frac{dW}{ds} \approx \frac{\alpha\gamma}{R} \tag{3.22}$$

It is also practical to derive an expression for the number of photons emitted per unit of the bending angle $\theta$

$$N \approx \alpha\gamma\theta \tag{3.23}$$

which is given by a remarkably simple and clear formula.

## 3.2 SR effects on the beam

The derived characteristics of SR allow evaluation of the effects of SR on the beam.

### 3.2.1 SR-induced energy spread

The energy spread $\Delta E/E$ will grow due to statistical fluctuations ($\sqrt{N}$) of the number of emitted SR photons and therefore can be estimated as

$$\frac{d\ (\Delta E/E)^2}{ds} \approx \varepsilon_c^2 \frac{dN}{ds}\frac{1}{(\gamma mc^2)^2} \tag{3.24}$$

which gives the following estimation

$$\frac{d\ (\Delta E/E)^2}{ds} \approx \frac{r_e\,\lambda_e\,\gamma^5}{R^3} \tag{3.25}$$

Comparing this with the exact formula

$$\frac{d\ (\Delta E/E)^2}{ds} = \frac{55}{24\sqrt{3}}\frac{r_e\,\lambda_e\,\gamma^5}{R^3} \tag{3.26}$$

confirms good accuracy of the estimation as the numerical factor $55/(24\sqrt{3}) \approx 1.32$.

### 3.2.2 SR-induced emittance growth

Let's estimate the beam emittance growth rate due to synchrotron radiation. The qualitative picture of the phenomenon is shown in Fig. 3.5.

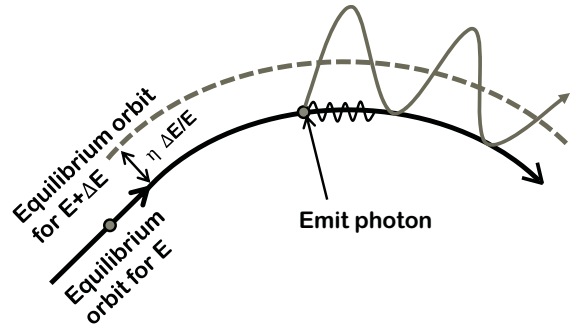



**FIGURE 3.5**
SR can cause excitation of oscillation of particles and corresponding emittance growth.

In this diagram, the dispersion function $\eta$ shows how the equilibrium orbit shifts when particle energy changes due to photon emission. Correspondingly, when a photon is emitted and the energy of the particle becomes equal to $E+\Delta E$ (where $\Delta E$ is negative), the particle starts to oscillate around a new equilibrium orbit. The amplitude of oscillation will be equal to

$$\Delta x \approx \eta\, \Delta E/E$$

Let's compare this with the betatron beam size given by

$$\sigma_x = (\varepsilon_x\, \beta_x)^{1/2}$$

and write an estimate for the emittance growth as

$$\Delta\varepsilon_x \approx \Delta x^2/\beta$$

By expanding the equation, we obtain an estimation for the emittance growth:

$$\frac{d\varepsilon_x}{ds} \approx \frac{\eta^2}{\beta_x}\,\frac{d\ (\Delta E/E)^2}{ds} \approx \frac{\eta^2}{\beta_x}\,\frac{r_e\,\lambda_e\,\gamma^5}{R^3} \qquad (3.27)$$

In the above estimation we ignored the dependence of $\beta$ and $\eta$ on $s$; however, these dependences can alter the results. The exact formula, which takes into account the derivatives of the Twiss functions, is as follows:

$$\frac{d\varepsilon_x}{ds} = \underbrace{\frac{\eta^2 + \left(\beta_x\eta' - \beta_x'\eta/2\right)^2}{\beta_x}}_{=H}\,\frac{55}{24\sqrt{3}}\,\frac{r_e\,\lambda_e\,\gamma^5}{R^3} \qquad (3.28)$$

where the parenthesis with Twiss functions in front of the numerical coefficient is usually called $H$. As we see, the back-of-the-envelope estimation correctly captures the most important features of the phenomenon and also produces a useful and simple expression.

### 3.2.3 Equilibrium emittance

The SR-induced cooling of the beam emittance and SR-induced emittance growth would naturally balance, so that the beam emittance would eventually reach equilibrium value.

Let's look at the estimated rate of emittance growth:

$$\frac{d\varepsilon_x}{ds} \approx \frac{\eta^2}{\beta_x} \frac{d\ (\Delta E/E)^2}{ds} \approx \frac{\eta^2}{\beta_x} \frac{r_e\, \lambda_e\, \gamma^5}{R^3}$$

and the SR cooling rate

$$\frac{d\varepsilon}{ds} = -\frac{2}{c\,\tau}\,\varepsilon \qquad \text{with} \qquad \tau^{-1} = \frac{1}{3}\frac{c\, r_e\, \gamma^3}{R^2}$$

and equate them to obtain an expression for the horizontal equilibrium emittance:

$$\varepsilon_{x0} \approx \frac{c\,\tau}{2} \frac{\eta^2}{\beta_x} \frac{r_e\, \lambda_e\, \gamma^5}{R^3} \tag{3.29}$$

or, after substitution:

$$\varepsilon_{x0} \approx \frac{3}{2} \frac{\eta^2}{\beta_x} \frac{\lambda_e\, \gamma^2}{R} \tag{3.30}$$

In the equations above, we ignored dependence of $R$ on longitudinal coordinate $s$. In order to obtain more accurate formulas from these equations, one needs to use the values $\langle 1/R^2 \rangle$ and $\langle 1/R^3 \rangle$, which are averaged over the orbit period.

In the vertical plane, SR's contribution to emittance is only due to $1/\gamma$ angles of emitted photons, but usually the impact on highly relativistic beams is negligibly small.

The vertical equilibrium emittance is therefore usually defined not by SR directly, but by the coupling coefficient $k$ (which is $\ll 1$) of x-y planes:

$$\varepsilon_{y0} \approx k\, \varepsilon_{x0}$$

In the above, we ignored partition numbers, but they can be taken into account in accurate calculations. The equilibrium energy spread of the beam can also be calculated in a similar manner.

## 3.3 SR features

Performance of SR-based light sources depends on spatial and spectral characteristics of synchrotron radiation. Below we will introduce basic relevant characteristics, leaving detailed discussion for Chapter 7.

### 3.3.1 Emittance of single radiated photon

In order to discuss the ultimate brightness of the SR-based light sources, we require knowledge of the emittance of a single photon radiated from the curved beamline.

To answer this question we must first evaluate the size of the emitting region seen by the remote observer in the case of a single electron, as illustrated in Fig. 3.6.

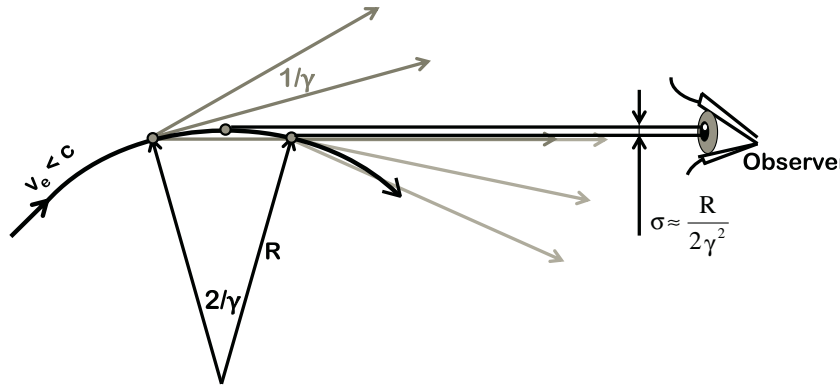



**FIGURE 3.6**
For illustration of emittance of a single photon.

We first note that the angles of photons coming from the emitting region are spread over $\sigma' \approx 1/\gamma$. The size of the emitting region is given by the height of the arc segment and is thus equal to $\sigma \approx R/(2\gamma^2)$.

The estimate for the emittance of an SR photon emitted by a single electron can therefore be written as

$$\varepsilon_{ph} = \sigma\,\sigma' \quad \Rightarrow \quad \varepsilon_{ph} \approx \frac{R}{2\gamma^3} \tag{3.31}$$

(In a similar way as above, one can estimate beta function of photons as $\beta = \sigma/\sigma'$.)

Let's rewrite the photon emittance equation using the expression for photon wavelength

$$\omega_c = \frac{2\pi c}{\lambda_c} \approx \frac{c\gamma^3}{R} \quad \Rightarrow \quad \varepsilon_{ph} \approx \frac{\lambda_c}{4\pi} \tag{3.32}$$

We can see here that the emittance of synchrotron radiation is directly related to its wavelength. This is not a coincidence

in a single example, but is actually an inherent property of photon radiation:

$$\varepsilon_{ph} = \frac{\lambda}{4\pi} \tag{3.33}$$

This, together with information about the SR spectrum in the next section, will bring us to discuss the brightness of SR light sources.

### 3.3.2 SR spectrum

In all the estimations above, we assumed that the photons emitted are monoenergetic. It is not exactly the case, and in reality the energy of the photons will be distributed around the *characteristic frequency* of the SR photons $\omega_c$.

Accurate mathematics, which we do not show here, predicts that the SR spectrum looks like the one shown in Fig. 3.7.

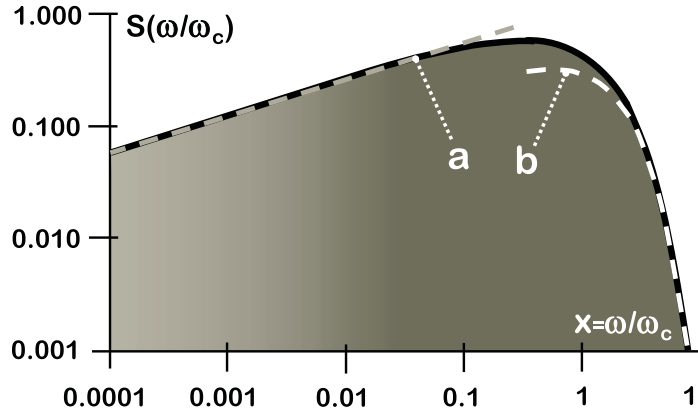



**FIGURE 3.7**
SR spectrum and its approximations for low (curve **a** — behaves as $4/3 \cdot x^{1/3}$) and high (curve **b** — behaves as $7/9 \cdot x^{1/2}e^{-x}$) energies.

We can indeed see that a large fraction of the photons will have energies close to $\omega_c$. However, there is also a lower energy tail, as well as some fraction of higher energy photons.

It is also natural to expect that the photons' angular distribution will deviate from the $1/\gamma$ rule, and indeed, the lower energy photons typically have larger angular spread.

### 3.3.3 Brightness or brilliance

Following discussion of the SR spectrum, we can introduce the notion of *bandwidth* — the interval of interest in the spectrum of photon frequencies. This bandwidth is denoted here $BW$ and is expressed, typically, in %.

Let's assume that our photon beam is emitted from the

area $A_s$ and the emitted radiation has angular opening angles of $\Delta\Phi$ and $\Delta\psi$, as shown in Fig. 3.8.

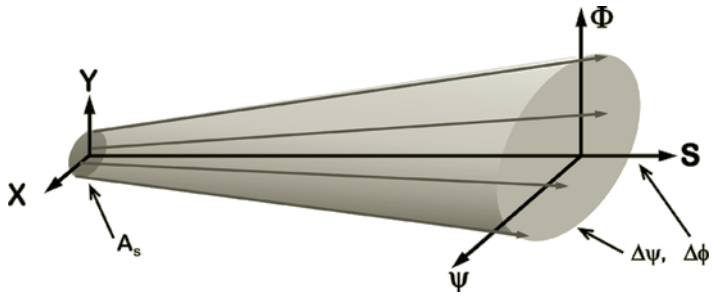



**FIGURE 3.8**
For illustration of brilliance or brightness. Here $A_s$ is the emitting area and $\Delta\Psi$ and $\Delta\Phi$ are opening angles of emitted photons.

The first concept to make note of is *flux*, which is expressed as the number of photons emitted per units of time and per unit of bandwidth:

$$\text{Flux} = \text{Photons}/(s \cdot BW) \tag{3.34}$$

The brilliance (or brightness) is then defined as flux per unit of emitting area and product of angles:

$$\text{Brilliance} = \text{Flux}/(A_s \cdot \Delta\Phi \cdot \Delta\psi)$$

and is usually expressed in units of [Photons/$(s \cdot mm \cdot mrad^2 \cdot BW)$].

In a typical case of Gaussian distributions, the definition of brilliance is based on the total effective sizes and divergences

$$\text{brilliance} = \frac{\text{flux}}{4\pi^2 \Sigma_x \Sigma_{x'} \Sigma_y \Sigma_{y'}} \tag{3.35}$$

where the total effective sizes include contributions from electrons as well as photons:

$$\Sigma_x = \sqrt{\sigma_{x,e}^2 + \sigma_{ph,e}^2} \qquad \sigma_x = \sqrt{\varepsilon_x \beta_x + (D_x \sigma_\varepsilon)^2} \tag{3.36}$$

$$\Sigma_{x'} = \sqrt{\sigma_{x',e}^2 + \sigma'_{ph,e}{}^2} \qquad \sigma_{x'} = \sqrt{\varepsilon_x \beta_x + (D'_x \sigma_\varepsilon)^2}$$

and similarly for the other plane.

### 3.3.4 Ultimate brightness

As we have seen, brilliance is defined by the overall effective emittance, which convolves electron and photon distributions:

$$\varepsilon_{eff} = \sqrt{\sigma_e^2 + \sigma_{ph}^2}\ \sqrt{\sigma_{e'}^2 + \sigma_{ph'}^2} \tag{3.37}$$

Since the lowest photon emittance depends on the photon wavelength (Eq. 3.33), the smallest overall emittance will be obtained when:

$$\varepsilon_e = \sigma_e \sigma_{e'} \leq \varepsilon_{ph} \tag{3.38}$$

which corresponds to a *diffraction-limited* source.

*Wavelength $\lambda \approx 1\,\mathring{\mathrm{A}}$ corresponds to 12.4 keV photons.*

In modern SR sources, the typical radiation of interest spans from 100 eV to 100 keV in terms of photon energy. Let's take an example of 12.4 keV of photon energy which corresponds to $\varepsilon_{ph} \approx 8$ pm. For typical third-generation SR light sources, the horizontal emittance $\varepsilon_x$ is usually between 1 and 5 nm and the vertical emittance $\varepsilon_y$ between 1 and 40 pm. Thus, such rings are close, in terms of its emittance, to the ultimate performance in the vertical plane. However, they are many orders of magnitude away from the diffraction-limited emittance in the horizontal plane. We will discuss the questions of brilliance in closer detail in Chapter 7, which is dedicated to light sources.

### 3.3.5 Wiggler and undulator radiation

Let's consider radiation from a sequence of bends, and in particular, let's assume that the bends are arranged in a sequence with $+-+-+-$ polarity with a period of $\lambda_u$, so that the trajectory of the particle *wiggles* as shown in Fig. 3.9.

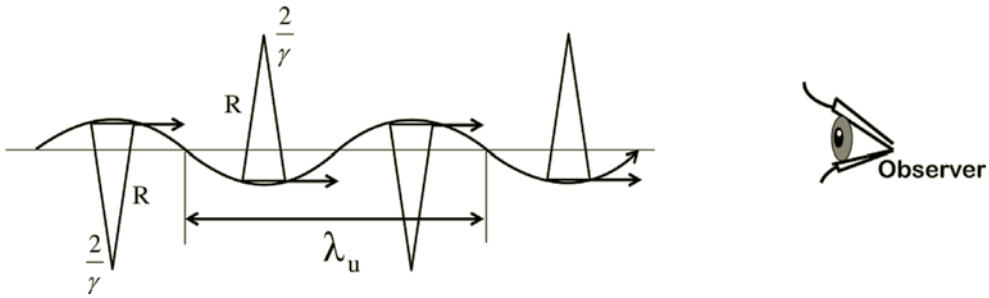



**FIGURE 3.9**
Radiation from sequence of bends.

An external observer will see photons emitted by the particle during its travel along the arc $2R/\gamma$. Let's now define the parameter $K$ as the ratio of the wiggling period to the length of this arc:

$$K \sim \gamma\, \lambda_u / R \tag{3.39}$$

We can qualitatively see that if $2R/\gamma \ll \lambda_u/2$, then the radiation emitted at each wiggle is independent. This corresponds to $K \gg 1$ and is called the *wiggler regime*.

On the other hand, if $2R/\gamma \gg \lambda_u/2$, then we are in a regime where the entire wiggling trajectory contributes to radiation (therefore *interference* leads to *coherence* of radiation,

which is explored in detail in further chapters). This corresponds to $K \ll 1$ and is called *the undulator regime*.

As we see, the *undulator parameter* $K \sim \gamma\ \lambda_u/R$ defines different regimes of synchrotron radiation: $K \gg 1$ is the wiggler regime, $K \ll 1$ is the undulator regime.

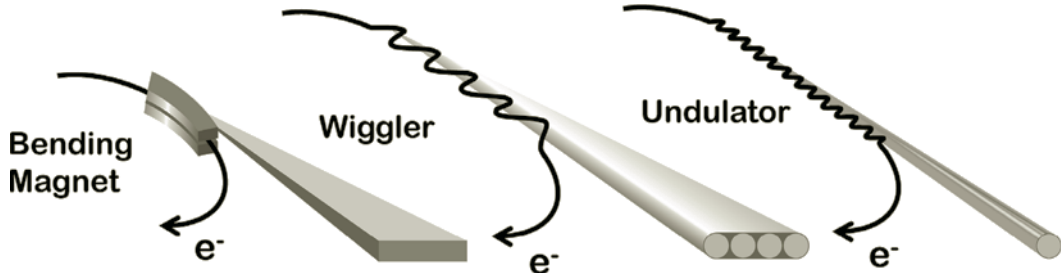



**FIGURE 3.10**
Wiggler and undulator radiation.

Fig. 3.10 shows the differences between radiation from a single bend and from a sequence of bends in wiggler and undulator regimes, respectively. We will consider this topic in more detail in the Chapters dedicated to light sources and to FELs, which are respectively 7 and 8.

### 3.3.6 SR quantum regime

Let's define the parameter "*Upsilon*"as $\Upsilon = \hbar\omega_c/E$. The meaning of this parameter depends on the regimes of SR.

When the parameter $\Upsilon \ll 1$, its physical meaning is the ratio of the characteristic photon energy to the energy of a single electron in the beam.

However, when $\Upsilon \sim 1$ and higher, the classical regime of synchrotron radiation is not applicable, and the quantum SR formulae of Sokolov–Ternov should be used. Such a situation may happen in particular in collision or highly relativistic focused beam, e.g., in linear colliders.

In a quantum regime, the shape of the SR spectrum changes, as there should not be a photon emitted that has energy larger than the energy of the initial particle.

The qualitative dependence of the SR spectrum in classical and quantum regimes is shown in Fig. 3.11.

Though the quantum SR is unlikely to occur in radiation from bends, it can happen in SR during beam collisions, as beams focused to tiny spots can produce enormous fields that cause the oncoming particles to radiate in a quantum regime.

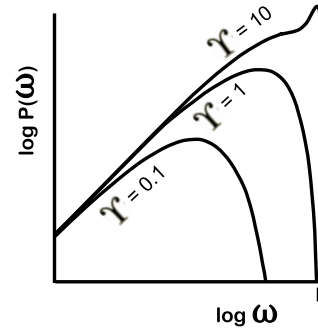



**FIGURE 3.11**
SR spectrum in classical and quantum regimes.

## EXERCISES

3.1 *Chapter materials review*.
A proton beam of E=50 TeV circulates in a 100 km perimeter ring. Estimate the synchrotron radiation energy loss per turn, the characteristic energy of the emitted photons and the cooling time.

3.2 *Chapter materials review*.
Describe how one needs to change the optics of third-generation SR sources in order to approach the diffraction-limited SR source, particularly in the horizontal plane.

3.3 *Chapter materials review*.
In a manner similar to how the equilibrium emittance was estimated in this chapter, derive the equilibrium energy spread of the beam.

3.4 *Mini-project*.
Define the approximate parameters (energy, sizes, fields in bending magnets) of a second-generation SR source aiming to achieve 10 keV of X-rays.

3.5 *Analyze inventions or discoveries using TRIZ and AS-TRIZ*.
Analyze and describe scientific or technical inventions described in this chapter in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

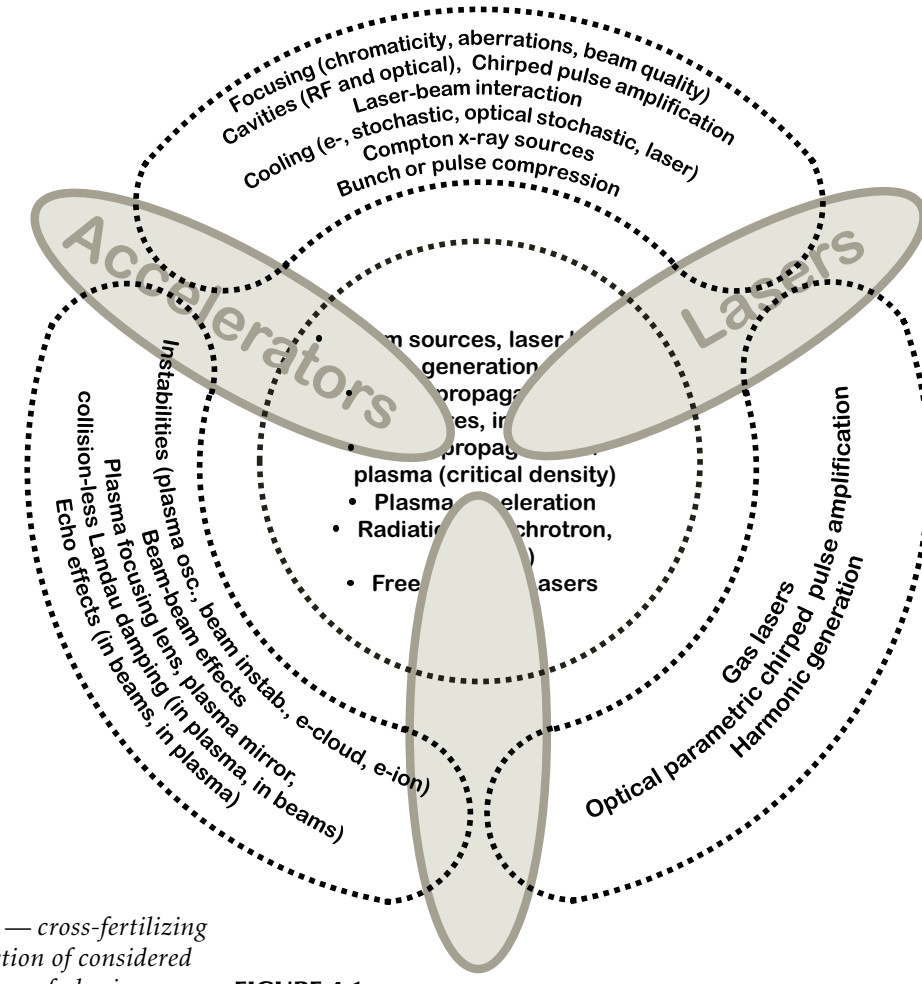
3.6 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

# 4 Synergies between Accelerators, Lasers and Plasma



In this chapter we will discuss the synergies between accelerators, lasers and plasma.

Within the scope of our interests, the themes linking all three areas *(accelerators, lasers, plasma)*, include, among others, beam sources, laser beam generation, wave propagation (in structures and in plasma), laser propagation in plasma, plasma acceleration, radiation (synchrotron, betatron) and free electron lasers, as illustrated in Fig. 4.1.




*Synergy — cross-fertilizing interaction of considered areas of physics.*

**FIGURE 4.1**
Synergies between accelerators, lasers and plasma.

DOI: 10.1201/b18696-4

Some of the themes that connect both accelerators and lasers are: focusing (chromaticity, aberrations, beam quality), cavities (RF and optical), laser-beam interaction (ponderomotive force), laser imprints (on e-beam in wigglers), cooling (e-, stochastic, optical stochastic, laser), Compton X-ray sources, chirped pulse amplification and bunch or pulse compression.

Accelerators' and plasma's connecting themes include: instabilities (plasma oscillation, beam instability, e-cloud, e-ion), beam–beam effects, plasma-focusing lenses, plasma mirrors, collision-less Landau damping (in plasma, in beams) and echo effects (in beams, in plasma).

Lastly, lasers' and plasma's themes include: gas lasers, optical parametric chirped pulse amplification and harmonic generation. In this chapter and in this entire text, we will touch on only some of the topics mentioned above.

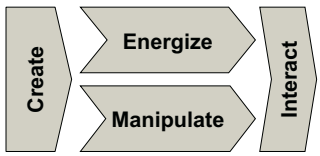



**FIGURE 4.2**
Discussion of synergies will follow this sequence.

In the following sections we will discuss accelerators, lasers and plasma in the order of how we interact with them in the real world: First, we *create* the beam/light pulse/plasma wave; then we prepare them for use, i.e., *energize* (accelerate, amplify, excite in plasma) them or *manipulate* (focus, compress, stretch, etc.) them; and, lastly, use or *interact* with them. This sequence is illustrated in Fig. 4.2.

## 4.1 Create

Let's discuss the topic of beam, light pulse and plasma wave creation in the *Create* — Energize — Manipulate — Interact sequence.

### 4.1.1 Beam sources

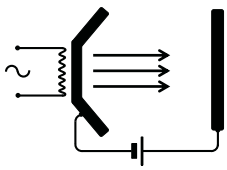


**FIGURE 4.3**
Thermal cathode e-gun.

We will begin our dialogue on particle sources by starting with *leptons* (electrons, positrons, muons, etc.) and later moving on to *hadrons* (protons, antiprotons, etc.) and ions.

The simplest form of an electron source is the thermal cathode gun (Fig. 4.3). According to the "three-halves power" law (or the Child-Langmuir law), the space charge effects of the non-relativistic accelerated electron beam often limit the current in electron guns.

$$I = P \cdot U^{3/2} \tag{4.1}$$

where $I$ is the current, $U$ is the cathode to anode voltage and the coefficient $P$ is called *perveance*.

A much more modern source of electrons is the laser-driven photocathode gun (illustrated in Fig. 4.4). Photo electron guns contain electrons that are generated by a laser field via the photoelectric effect. Such a gun usually consists of one-and-a-half RF cavities and a photocathode made either out of metal (e.g., copper) or from a special alloy. While a pure metal photocathode is robust, its *quantum efficiency* (the

number of electrons per number of incident photons) is usually quite low, of the order of $10^{-4}$. Photocathodes with alkali metals such as cesium can have a *quantum efficiency* of 10% or higher.

The photocathode guns driven by a laser pulse are also very suitable for production of short electron beams. The laser pulse is usually sent to the cathode at a small angle as shown in Fig. 4.4 to avoid interference with the accelerating electron beam.

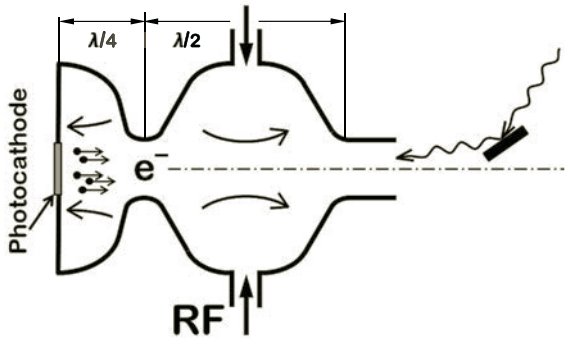



**FIGURE 4.4**
One-and-a-half-cell RF photocathode electron gun.

As seen from Eq. 4.1, photoguns can provide a much higher pulsed electron current — due to a higher accelerating voltage — produced at the cathode by the RF structure. The typical accelerating voltage in guns ranges from around 50 MV/m in an L-band, 100 MV/m in an S-band and 200 MV/m in an X-band.

*RF frequencies bands: L: 1-2 GHz, S: 2-4 GHz, C: 4-8 GHz, X: 8-12 GHz.*

Producing positrons usually requires creating $e^+e^-$ pairs followed by separating the positrons. An electron or photon beam with a sufficient amount of energy is sent onto a target where the $e^+e^-$ pairs will be produced. Separated positrons are then accelerated and sent to a damping ring, where their emittance will decrease due to radiation damping.

Ion or proton beams are produced by plasma-based ion sources — a large variety of source types exists. An example shown in Fig. 4.5 depicts a *Penning source* in which a magnetic trap is arranged in the cathode-anode area of an electron beam, ionizing gas via discharge. Ions of certain charges are extracted with the help of an electrode, and are then separated and sent for further acceleration and miscellaneous use. In particular for negative ion sources, *cesiation* (developed by V. Dudnikov; when a small amount of Cs atoms is added into the gas) is often used to significantly enhance the emission of negative ions.

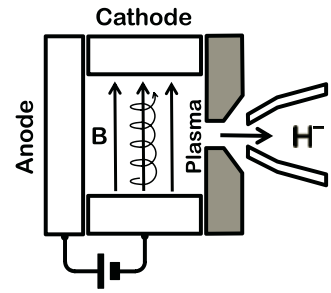



**FIGURE 4.5**
Surface-plasma Penning $H^-$ ion source.

Plasma in ion sources can be created by various means such as ionization via laser (which was realized in the ion source developed at CERN, shown in Fig. 4.6). In this exam-

*Consider whether evolution from discharge ion source to laser-driven ion source can suggest a new inventive principle.*

ple, the laser beam is focused by a metal mirror on a target, ionizing the target and producing plasma. The ions of plasma are then directed to the extraction electrodes via a hole in the mirror.

This brings us to the discussion of lasers.

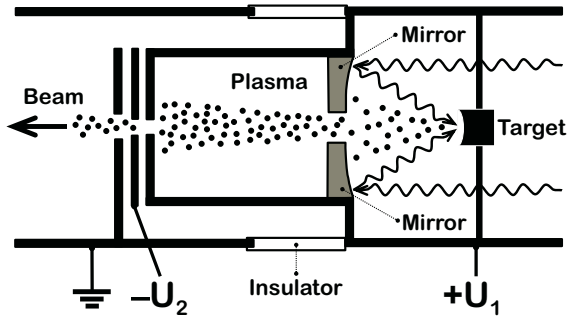



**FIGURE 4.6**
Conceptual diagram of the laser-driven ion source.

### 4.1.2 Lasers

Lasers are a source of coherent light, and a laser diagram is shown in Fig. 4.7. The main laser components are the *gain medium* (which amplifies the light), the resonator (which gives optical feedback) and the pump source (which creates *population inversion*).

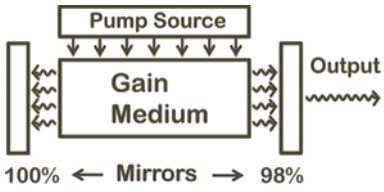



**FIGURE 4.7**
Laser diagram.

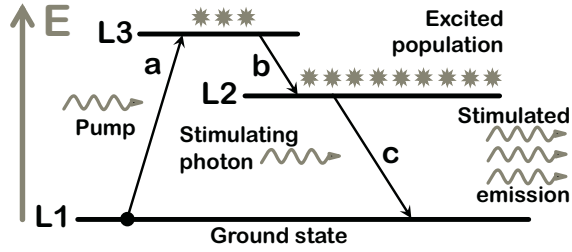



**FIGURE 4.8**
Three-level laser.

The gain medium contains atoms with specific conditions for the energy levels and for the lifetime of an excited state of the atom, suitable for arranging a *three-level laser* as described in the following paragraph. The pump gets the atom population from the ground state $L_1$ to the higher energy level $L_3$. The excited population gets from $L_3$ to $L_2$ through non-radiative decay as the lifetime of the $L_3$ state is very short and all the population in state $L_3$ decays to state $L_2$. The lifetime of energy state $L_2$ is, in contrast, very long and therefore, a *population inversion* occurs with respect to the state $L_1$. Once the population inversion is obtained, stimulated emission can occur when an incident photon forces the population to drop

from level $L_2$ to $L_1$ (as shown in Fig. 4.8), resulting in an optical gain.

As you can see from the above description, a minimum of three levels are needed to arrange a laser, but it is possible to create four and higher-level laser systems as well.

*An advantage of four-level laser systems is that much less pumping power is needed for the creation of population inversion.*

The first laser built was a three-level ruby laser emitting a 694 nm wavelength (shown conceptually in Fig. 4.9). The first mass-produced laser was a four-level He-Ne gas laser emitting a 694 nm wavelength. These first lasers were characterized by low emission power and very low *efficiency* — typically around 0.01-0.1%.

A $CO_2$ gas laser emitting a 10.6 $\mu$m wavelength, on the other hand, has an efficiency close to 30% and a high power level measuring up to kW in CW. It is also interesting to note that the quantization of the $CO_2$ molecule's vibrational and rotational states enables this laser's system levels.

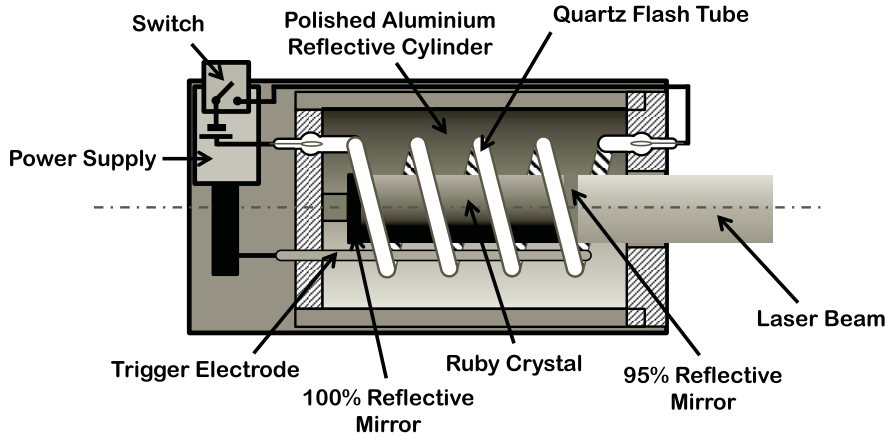



**FIGURE 4.9**
Conceptual diagram of a ruby laser. Quartz flash tube serves as the pump source and ruby crystal as the gain medium.

One more example of an efficient laser (at around 40%) is the *diode laser*. In this laser, the levels of the system are enabled by the quantization of energies of holes and electrons in the semiconductor diode. Due to the compactness of the diode laser, its output light has a very large divergence, low coherence and usually low power. However, the versatility of wavelength output and high efficiency makes this type of laser ideal for *pumping* — the excitation of gain medium in high-power laser amplifiers. In the latter, the low power, low coherence and large divergence output of the laser diode is amplified in a gain medium such as Nd:YAG, resulting in a high-power, high-coherence, high-efficiency laser beam as illustrated in Fig. 4.10. Note that pumping by the diode laser occurs at a shorter wavelength than the output radiation, in agreement with Fig. 4.8.

*Nd:YAG — Neodymium-doped yttrium aluminum garnet: $Nd:Y_3Al_5O_{12}$.*
*Yb:YAG — Ytterbium-doped YAG.*

*When used for laser pumping, many low-power diode lasers are usually assembled into arrays called "diode bars."*

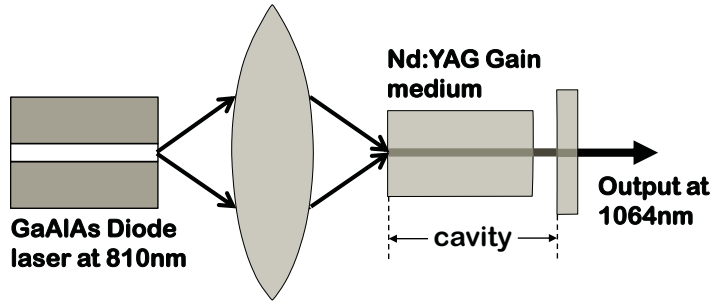



**FIGURE 4.10**
Diode laser in application for pumping of Nd:YAG laser.

The laser diode output can also be applied to fiber optics and fiber amplifiers, as we will discuss in the next section.

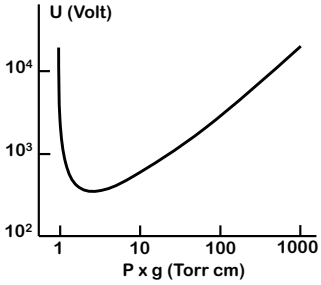



**FIGURE 4.11**
Paschen discharge curve for hydrogen gas.

### 4.1.3 Plasma generation

To conclude our discussion regarding creation, let's briefly consider the process of how plasma is made. The most well-known method is discharge ionization in which plasma is produced by DC voltage or a pulse typically applied to a low pressure gas. The voltage $U$ that creates discharge depends on the product of the gap $g$ to gas pressure $P$, i.e. $(P \times g)$ — the behavior of this voltage is represented by the *Paschen curve* as shown in Fig. 4.11.

*Ionization energy of hydrogen is* ~ 13.6 *eV.*

Several factors are important in sustaining conditions appropriate for a breakdown. On one side, the energy that the electron gains during its acceleration over the *mean free path* (before the next collision) should be larger than the first *ionization energy* of the gas molecules. A lower pressure (longer mean free path) would therefore create preferable conditions for a breakdown. Multiple collisions and multiple electron ionizations are required in order to create suitable conditions for an *avalanche* and thus a breakdown. On the other hand, overly low $P \times g$ would mean that electrons would not have enough collisions while traveling through the gap, making an avalanche less probable. The balance of these effects is reflected in the behavior of the Paschen's curve.

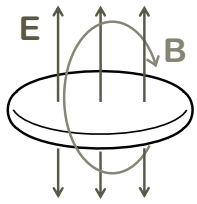



**FIGURE 4.12**
Fields of a relativistic electron bunch can produce field ionization of gas.

Plasma can also be created by *field ionization*. Both a well-focused relativistic electron beam (see Fig. 4.12) and a laser can have a sufficient field to ionize gas. A gas or target can then be instantaneously ionized if the field level reaches an *atomic field* scale:

$$E_{Beam} \gg E_{Atomic} \quad \text{or} \quad E_{Laser} \gg E_{Atomic}$$

As we will discuss in Chapter 6, due to the *multi-photon ionization* and *tunneling* effects, much lower-level fields than atomic ones are often sufficient for ionization.

## 4.2 Energize

Next comes the process of *energizing* the beam, laser or plasma. In the case of a charged particle beam, this involves acceleration (electrostatic, betatron, acceleration in RF cavities and structures, or plasma acceleration). In the case of a laser, it involves amplification (standard, chirped pulse amplification, optical parametric chirped pulse amplification, etc.), and for plasma, excitation of waves by various means such as short pulse of a laser or beam.

### 4.2.1 Beam acceleration

The simplest and exceedingly widespread acceleration method is electrostatic acceleration (Fig. 4.13). Not so widely used now is betatron acceleration principle, which is based on Maxwell's Eq. 2.4.

Perhaps the most versatile acceleration method involves the use of resonators — also known as RF cavities. Cavities can be arranged in structures and made to be suitable for the acceleration of any types of particles from electrons (which almost immediately become relativistic) to protons or ions.

Often, acceleration is combined with focusing, either via magnetic quadrupoles inserted between acceleration sections, or via an EM wave with a quadrupole component as in a *RFQ accelerating structure*.

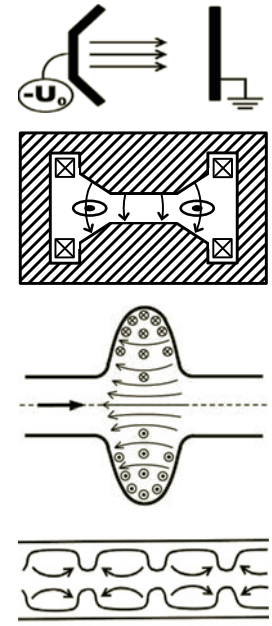



**FIGURE 4.13**
Electrostatic and betatron acceleration. RF cavity and RF structure.

### 4.2.2 Laser amplifiers

One of the possible principles of laser amplification is that a pumped gain medium of the *amplifier* amplifies light at the wavelength of the *oscillator* laser, which is made with the same material as the pumped gain medium of the amplifier. The diagram of such a laser amplifier is shown in Fig. 4.14.

The technical challenges caused by laser light amplification actually resulted in numerous inventions and breakthroughs, as discussed in Chapter 1.

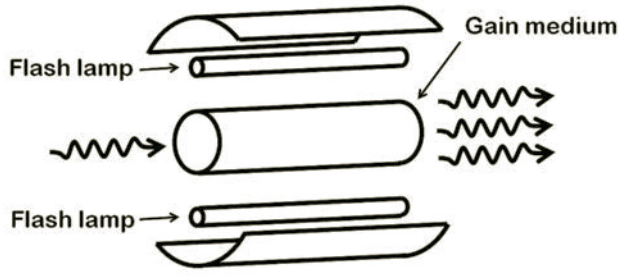



**FIGURE 4.14**
Laser amplifier. Flash lamp emits in broad spectrum. Gain medium amplifies selected wavelength.

In particular, lasers with ultra-short and ultra-high power pulse have had their share of challenges. Ultra-short intense pulses can cause nonlinear effects in the medium, while the high-power pulses cause heating up of the amplifier medium. These issues limit the repetition rate, power and efficiency of laser systems. Some of the most powerful lasers fire just once every few hours!

### 4.2.3 Laser repetition rate and efficiency

Let's take a look at the challenges of high average power lasers from the point of view of AS-TRIZ. The TRIZ-formulated problem is as follows: as the intensity of the pulsed laser light increases, it takes much more time for the medium to cool down and be ready for its next use. The identified *contradiction* is between the parameter that needs to be improved, INTENSITY, and the parameter that gets worse, REPETITION RATE.

A general principle that can solve this contradiction can be looked up in nature (see Fig. 4.15) or taken from AS-TRIZ, where Principle 4, *Volume-to-surface ratio*, suggests changing this ratio in order to alter the object's characteristics (such as its cooling rate or its fields).

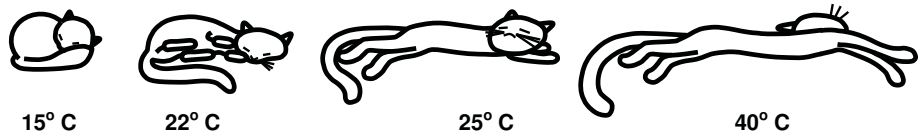



**FIGURE 4.15**
The cat intuitively knows the inventive principle of changing her surface-to-volume ratio depending on the external temperature.

### 4.2.4 Fiber lasers and slab lasers

Fiber laser technology is a perfect illustration of the inventive principle of changing the surface-to-volume ratio.

Indeed, if the main technical issue with high average power laser amplifiers is cooling the active medium, then one would try to increase its surface, without changing its volume, to enhance thermal exchange with the environment. Fig. 4.16 clearly demonstrates that the typical geometry of the active medium (when its length $L$ is similar to its radius $R$) is the most disadvantageous, in terms of cooling.

On the other hand, either increasing $L/R$ towards fiber geometry of the active medium, or decreasing $L/R$ toward slab geometry, would both have advantages in terms of the surface-to-volume ratio $S/V$, thus making it easier to manage their temperature regimes.

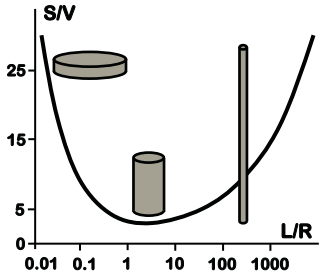



**FIGURE 4.16**
Volume-to-surface ratio $S/V$ in units of $(2\pi V)^{1/3}$ vs $L/R$.

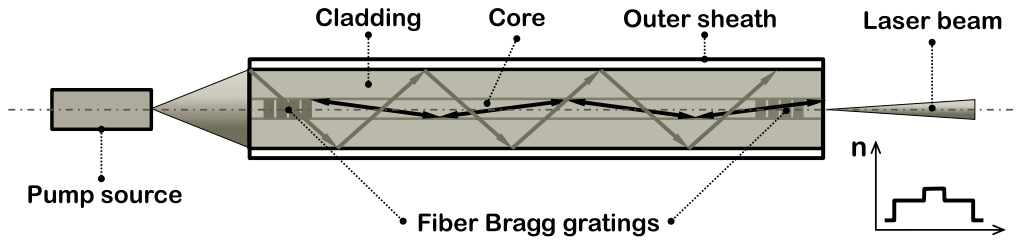



**FIGURE 4.17**
Schematic of a fiber laser and cross section of refractive index.

Correspondingly, the fiber laser and slab lasers are the technologies most suitable for achieving higher average power laser pulses, higher repetition rate and higher wall-plug efficiency. The main components of a fiber laser are shown in Fig. 4.17.

*Fiber laser technology uses the principle of larger surface-to-volume ratio.*

### 4.2.5 CPA — chirped pulse amplification

Invented for lasers by G. Mourou and D. Strickland, *chirped pulse amplification* (or CPA) is a process that stretches the short pulse in time, amplifies the now-longer laser pulse and then compresses the amplified pulse. This multi-step process relieves the optical component of laser amplifiers from pulsed power, thus reducing nonlinear effects and avoiding material damage. The stretching and compressing of the pulses are based on the time–energy correlation property of the laser pulse.

*CPA technique was originally developed for radar, while chirped pulses could also be observed in nature, e.g., in the voices of bats.*

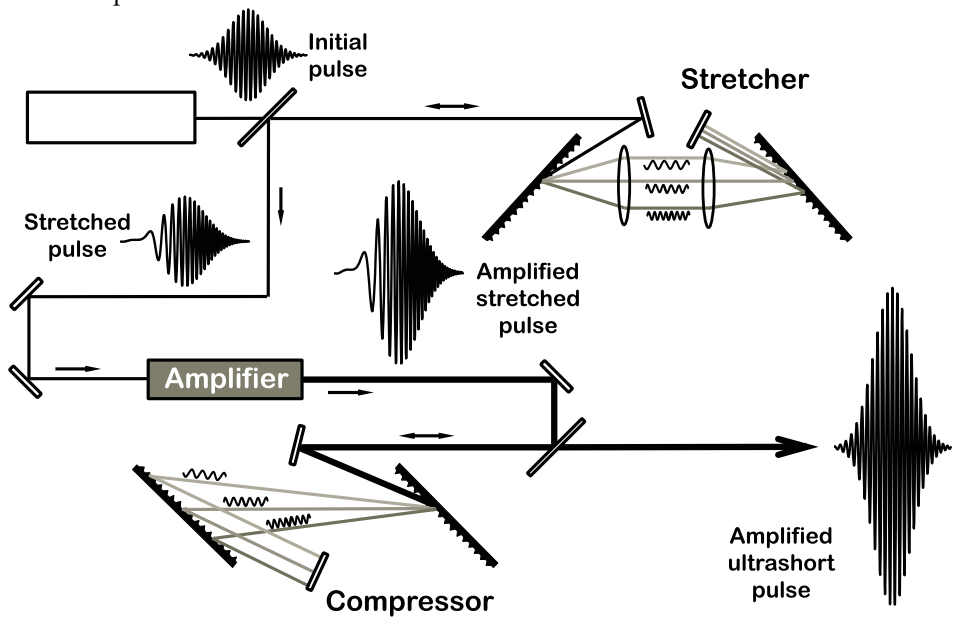



**FIGURE 4.18**
Schematics of CPA — chirped pulse amplification.

The schematics of a laser stretcher and compressor are presented in Section 4.3.4. The stretcher and compressor use a pair of diffraction gratings and rely on the fact that the angle of reflection from the grating depends on the wavelength, which sends different colors along different paths. The spectral width of the short laser pulse with duration $\tau$ is approximately given by $\Delta f \sim 1/\tau$. This finite spread of the spectrum around the carrier frequency makes such laser pulses suitable for spatial–spectral manipulation.

The schematics of CPA are shown in Fig. 4.18. The initial short pulse is provided by a short-pulse *oscillator*. The first pair of gratings disperses the spectrum and stretches the pulse by a factor of about a thousand (for visibility, the longitudinal extent of the pulses in Fig. 4.18 is shown qualitatively). After stretching, the pulse is long and has a low peak power, which is thus safer for amplification. After passing the power amplifier, the pulse is sent to the second pair of gratings, which reverses the dispersion of the first pair and compresses the pulse, producing a high-energy ultra-short laser pulse.

*Amplification of chirped pulses was used in radar and is now used in lasers — this trend from microwave to optical range can be taken as one of the generic principles of AS-TRIZ.*

The invention of CPA was one of the factors that ultimately pushed laser technology to such peak power levels that these lasers became a possible competitor alongside particle accelerators.

### 4.2.6 OPCPA — optical parametric CPA

Another method of laser amplification is called OPCPA — *optical parametric* CPA. Its principle is based on nonlinear properties of crystals (typically barium borate or *BBO crystals*) which, being subjected to radiation of wavelength $\omega_s$, generate radiation at two frequencies — $\omega_1$ and $\omega_2$ where, as energy is conserved, $\omega_1 + \omega_2 = \omega_s$ (as shown in Fig. 4.19).

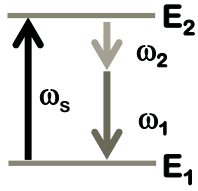



**FIGURE 4.19**
Optical parametric generation in nonlinear crystals.

In *optical parametric amplification*, the input consists of two beams: the *pump* at $\omega_s$ and the *signal* at $\omega_1$. The OPA output is the amplified $\omega_1$ beam and weakened $\omega_s$ beam, plus an additional *idler* beam at $\omega_2$.

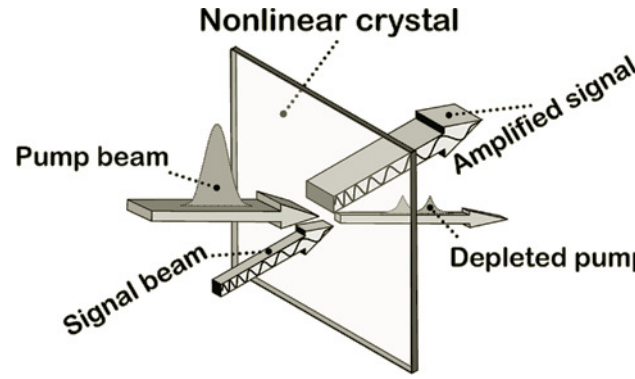



**FIGURE 4.20**
Optical parametric chirped pulse amplification — OPCPA.

The optical parametric amplifier system can be fed with a frequency-stretched *signal* pulse, as illustrated in Fig. 4.20. This makes the OPA system into a *chirped pulse* method known as OPCPA.

The OPCPA method is versatile; it can work from CW to femtosecond range in terms of pulse length, from UV to TeraHz in terms of light wavelength, and from mW to TW and PW in terms of the peak power.

*The main advantage of OPCPA is that it works via a parametric process, i.e., no energy is left in the nonlinear crystal and everything comes out in the form of light. This is beneficial for high energy or high mean power systems since the thermal issues are eliminated.*

### 4.2.7 Plasma oscillations

Jumping from lasers back to plasma topics, let's discuss the process of energizing the plasma, i.e., creating oscillations in plasma.

Imagine that there is a region in plasma where electrons of density $n$ shift with respect to the ions by a distance of $x$ as shown in Fig. 4.21. Applying Gauss's law

$$\oint_{\partial\Omega} \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\varepsilon_0} \int_{\Omega} \rho dV$$

will yield the value of an electric field produced by displaced charges

$$E = \frac{nex}{\varepsilon_0} \tag{4.2}$$

Writing an equation for the electrons' non-relativistic motion

$$F = m\frac{d^2x}{dt^2} = -eE = -\frac{ne^2x}{\varepsilon_0} \tag{4.3}$$

will then give us the expression for the oscillation frequency:

$$\omega_p^2 = \frac{ne^2}{\varepsilon_0 m} \tag{4.4}$$

Recalling the advice to express the end result in a form independent of the systems of units, we use

$$r_e = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{m_e c^2}$$

to rewrite the oscillating frequency or the *plasma frequency* as:

$$\omega_p^2 = 4\pi n c^2 r_e \tag{4.5}$$

We can also write a practical formula for $f_p = \omega_p/(2\pi)$:

$$f_p \approx 9000\, n^{1/2} \quad \text{(Hz)} \tag{4.6}$$

where $n$ is expressed in $(\text{cm}^{-3})$.

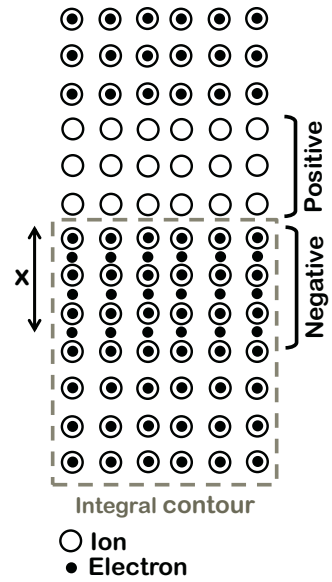



**FIGURE 4.21**
Plasma oscillations.

### 4.2.8 Critical density and surface

Let's take this moment to qualitatively consider the process of laser plasma penetration and the corresponding factors of *critical density* and *critical surface*.

When a laser hits a target or a dense gas, the target surface or gas is heated and ionized, which forms plasma. Hot plasma then starts expanding into the vacuum, creating a gradient of plasma density as illustrated in Fig. 4.22, with a respective gradient of plasma frequency.

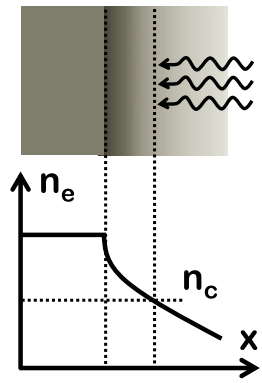



**FIGURE 4.22**
Laser penetration to plasma and critical density.

Qualitatively, if the plasma frequency $\omega_p$ in a particular layer of plasma is larger than the laser frequency $\omega$, then the plasma electrons can move fast enough and can thus create electric currents that will screen the fields of the laser EM wave. Therefore, lasers can penetrate plasma only to the point when

$$\omega_p < \omega$$

The critical density is therefore

$$n_c = \frac{\omega^2}{4\pi c^2 r_e} \tag{4.7}$$

In other words, the laser beam of frequency $\omega$ cannot penetrate areas with $n > n_c$.

## 4.3 Manipulate

Let's consider some of the ways we can *manipulate* beams, laser pulses or plasma.

For particle beams the topics of interest include focusing (weak, strong, chromaticity, aberrations); compressing; cooling (electron, stochastic, optical stochastic, laser); phase plane exchange; transverse stability, etc. For laser beams we are interested in focusing; compression; phase locking; harmonic generation, etc. Plasma-manipulation topics include plasma focusing; Landau damping; and laser self-initiated focusing in the plasma channel.

We will touch on some of these in this section, and several other topics will be discussed in Chapter 10.

### 4.3.1 Beam and laser focusing

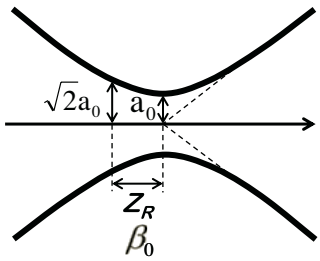



**FIGURE 4.23**
Beam or light in the focus.

Optics of light and optics of charged particles have a lot of similarities. However, different names are sometimes used for the same quantities in beam and light optics. To illustrate this, consider a situation wherein the laser or particle beam is focused into a point so that its minimum size at the waist is equal to $a_0$ (also called $w_0$ in light optics). The distance from the waist point to the point where the beam size increases by $\sqrt{2}$ is called the *Rayleigh length* $Z_R$ in light optics and is equivalent to the Twiss beta function $\beta_0$ at the location of the waist point in optics of charged particles (see Fig. 4.23).

### 4.3.2 Weak and strong focusing

Analogies between light and beam optics — as well as mechanical analogies (in the spirit of TRIZ or synectics) — can help us build up an intuitive understanding of complex phenomena. Let's illustrate this in an example of weak versus strong focusing.

Weak focusing can be brought about by bending dipoles with the same gradient all along the perimeter of an accelerator (refer to Fig. 4.24). In this figure, the weak focusing is compared to the motion of a ball in a gutter. Looking at Eq. 2.19 and Eq. 2.20, one can conclude that there are conditions when the beam is focused simultaneously in the $x$ and $y$ planes (due to the presence of an additional focusing term in bends in the $x$ plane). However, as follows from these equations or as Fig. 2.8 suggests, such focusing is weak — the spatial period of particle oscillation in this focusing field is of the order of the orbital circumference.

Weak focusing accelerators were mainly built in the early days. One of the disadvantages of weak focusing accelerators was the large transverse oscillations of particles, leading to wide apertures and correspondingly large and heavy magnets. For example, the 10 GeV weak focusing Synchrophasotron built in Dubna in 1957 (the biggest and the most powerful in its time), was registered in the Guinness Book of World Records for housing the heaviest magnet system, weighing 36,000 tons.

On the other hand, strong focusing can be generated via a sequence of focusing–defocusing quadrupoles, with the overall effect being equivalent to focusing, if certain conditions are satisfied. These conditions can be understood from the gutter analogy shown in Fig. 4.25 where the gutter is now bent, first of all, much more strongly, and second, it is bent up and down (see also Fig. 1.11). As clearly seen from the picture, stable motion is possible in this event only if the particle passes the areas of the downward-bent gutter near the center, in a way similar to what it would do in a FODO lattice. CERN's Proton Synchrotron, the first operating strong focusing proton accelerator, reached 24 GeV in 1959. It is constructed as a 200-m diameter ring, with a magnet weight of 3,800 tons — weighing ten times less for twice the energy of Dubna's Synchrophasotron — resulting in a clear demonstration of the advantages of the strong focusing approach.

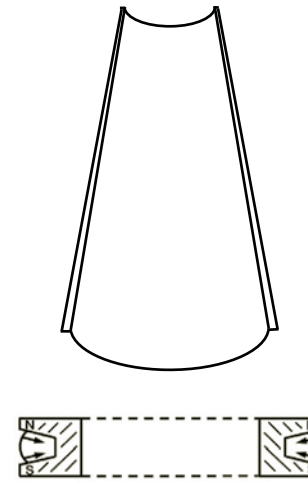



**FIGURE 4.24**
Weak focusing.

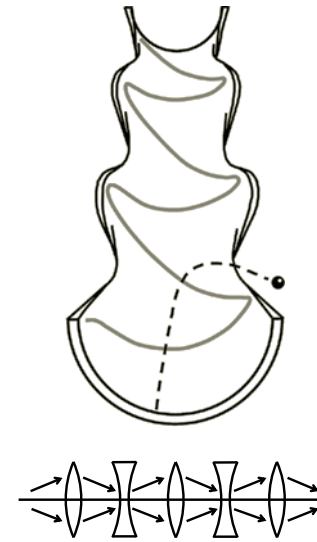


**FIGURE 4.25**
Strong focusing.

### 4.3.3 Aberrations for light and beam

Following Newton, we know that sunlight consists of a spectrum of colors and that each different color focuses differently — these are called chromatic aberrations. In precise optical devices, such as photo cameras, the chromatic aberrations need to be compensated.

*A curved mirror, in contrast to a lens, does not produce chromatic aberrations.*

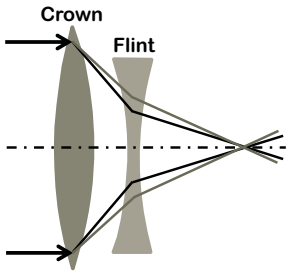



**FIGURE 4.26**
Lenses made from different glasses compensate chromatic aberrations.

For light, one uses lenses made from different materials to compensate for chromatic aberrations, as illustrated in Fig. 4.26. In this example, using a strong focusing lens made from a low-dispersion *crown* glass coupled with a weak defocusing high-dispersion *flint* glass can correct the chromatic aberrations.

For particle beams, chromatic aberrations are caused by the dependence of the focusing strength on the beam energy. There are no different materials and no specific Maxwell's equations in particle optics, therefore, other means have to be used for chromatic compensation in accelerators.

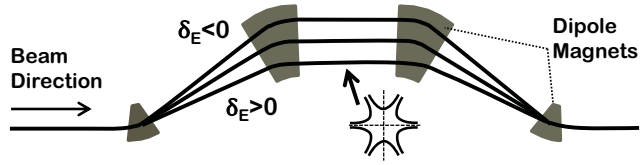



**FIGURE 4.27**
Compensation of chromatic aberration by inserting nonlinear sextupole magnets in a dispersive region.

Let's consider *chicane* arrangements of four bending dipoles as shown in Fig. 4.27. Such an arrangement creates a *dispersive* region, where orbit position depends on the energy. Let's insert a sextupole magnet with normalized strength $S$ (or $K'$) into the dispersive area

$$S = \frac{1}{2!B\rho}\frac{\delta^2 B}{\delta x^2} \tag{4.8}$$

The sextupole magnet produces the following effect (*kick*) on the angle of the beam trajectory

$$x' = x' + S\ \ x^2 - y^2 \qquad \text{and} \qquad y' = y' - S\ 2xy \tag{4.9}$$

In the dispersive region, one needs to replace $x$ with $x+\eta\delta$ where $\eta$ is the dispersion, thus the sextupole kick will contain the energy-dependent focusing since the terms in the equation below correspond to a quadrupole with an effective gradient equal to $S \cdot \eta$:

$$x' \Rightarrow\ S(x+\eta\delta)^2 \Rightarrow 2S\ \eta\ x\delta + .. \tag{4.10}$$

$$y' \Rightarrow\ -S\ 2(x+\eta\delta)y \Rightarrow\ -2S\ \eta\ y\delta + ..$$

Therefore, such an arrangement of nonlinear magnets placed in a dispersive region can be used for *chromatic correction* in the event of beam optics. One should also note that the terms that are not shown in Eq. 4.3.3, containing $x^2$ and $\delta^2$, are unwanted additional aberrations — geometrical aberrations and higher-order chromatic aberrations. They often need to be

corrected, either by special optic arrangements and appropriate placement of *sextupole pairs*, or by higher-order magnets.

### 4.3.4 Compression of beam and laser pulses

Compression of laser pulses relies on the dependence of the light reflection angle from the grating on the light wavelength, in the case when the grating spacing is comparable to the wavelength. This produces an optical system where the path length through the system depends on the light's wavelength.

*Energy-z correlation is used to compress/stretch either laser pulse or particle bunch — this is one more general principle of AS-TRIZ.*

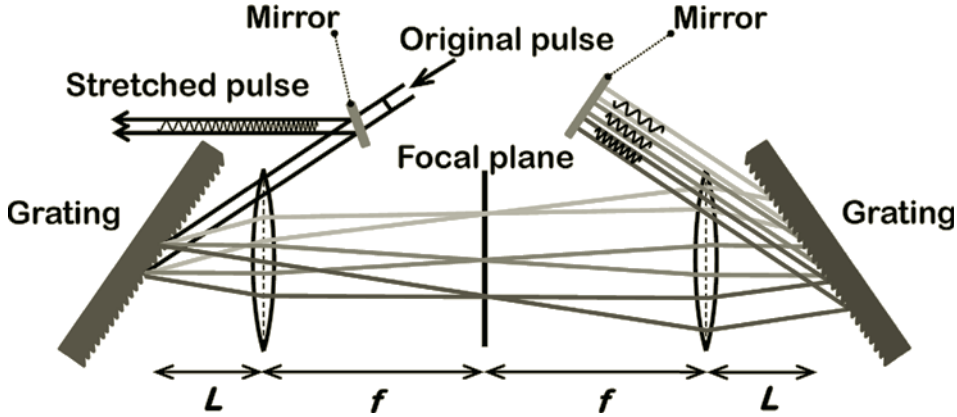



**FIGURE 4.28**
Laser pulse stretcher.

The combination of a pair of grated plates, lenses and mirrors can create a laser pulse stretcher — as realized in Fig. 4.28.

Equally, a laser pulse compressor can be arranged with a pair of gratings, as shown in Fig. 4.29.

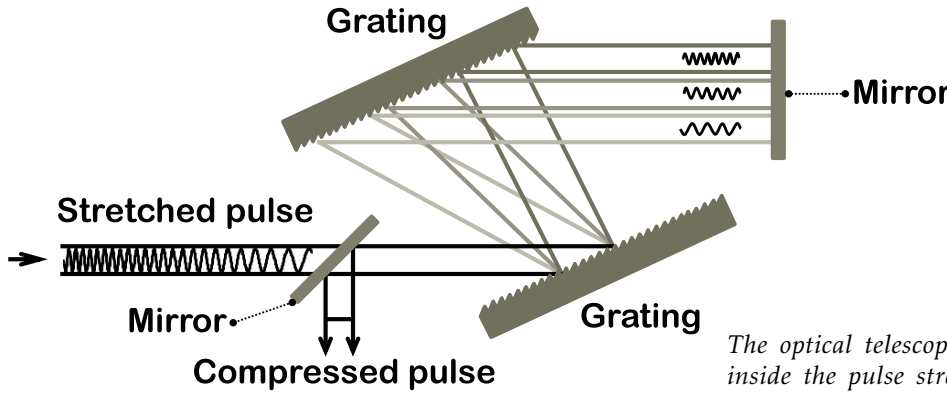



**FIGURE 4.29**
Laser pulse compressor.

*The optical telescope placed inside the pulse stretcher is needed in order to provide a "negative distance" between the gratings.*

Laser pulse compressor gratings are technological marvels. In order to stay below the damage threshold of the grating material, the size of the plates has to be around a meter in diameter for some of the highest power lasers. Taking into account that the typical space between grooves is around a micrometer, we can imagine the high levels of accuracy needed to produce such plates.

Charged particle beam compression is based on the same principle of the path length dependence on the particle's energy. Dispersion is created using dipole bending magnets and the initial beam is arranged to have an $E - z$ correlation, via its acceleration *off-crest* of RF voltage.

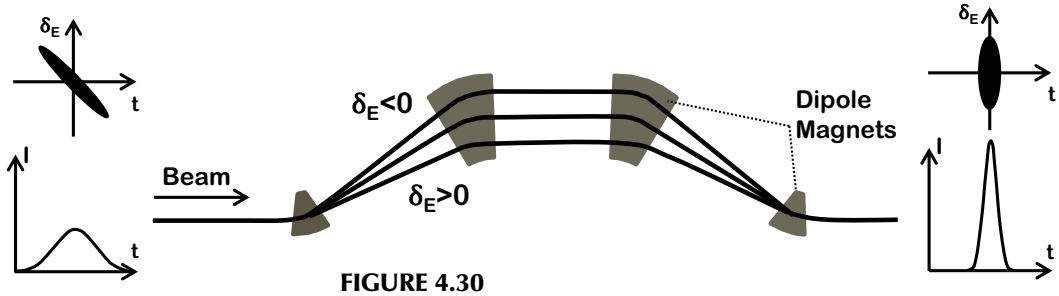



**FIGURE 4.30**
Bunch compressor.

Fig. 4.30 shows a bunch compressor and the phase space of the beam before and after compression.

### 4.3.5 Beam cooling

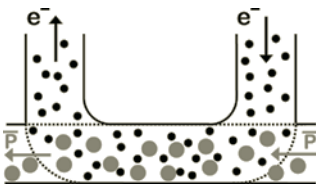



**FIGURE 4.31**
Electron cooling.

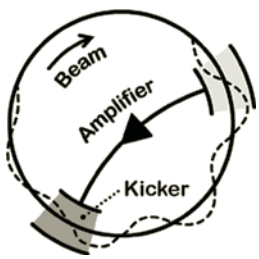



**FIGURE 4.32**
Stochastic cooling.

Beam cooling methods are usually necessary for antiparticles, e.g., antiprotons $\bar{p}$, as they are produced on a target very "hot" — with large emittance and energy spread. The cooling methods are intended to decrease beam emittances.

*Electron cooling*, invented by G.I. Budker and realized at Novosibirsk, consists of creating a region along the orbit of the beam where a "cold" electron beam will co-propagate together with a "hot" antiproton beam (see Fig. 4.31) at the same velocity. Energy exchange between the beams will eventually result in cooling of the $\bar{p}$ beam.

Another method of cooling the antiproton beams was proposed by S. van der Meer and realized at CERN. In this approach, a beam particle's betatron oscillation is detected by a pick-up electrode, and then the signal is amplified and sent via a short path across the ring onto a kicker (see Fig. 4.32). This will apply a kick to the same particle, resulting in a reduction of its oscillations. This method is called *stochastic cooling* and was indispensable for ensuring the discovery of W and Z bosons (Carlo Rubbia and Simon van der Meer, Nobel Prize in Physics, 1984).

As it should be apparent from the description of stochastic cooling, this method can clearly have drawbacks in the

case of a large number of particles in the beam, as the pick-up electrode would see signals from many particles, which would then smear individual contributions.

### 4.3.6 Optical stochastic cooling

"Standard" stochastic cooling entails sampling of the beam by electrostatic pick-up electrodes (see Fig. 4.33), and therefore its cooling rate is limited by the system bandwidth, which is defined by pick-up length. An extension of the method, *optical stochastic cooling*, uses optical pick-ups and optical amplifiers as shown in Fig. 4.34, resulting in potential increases of bandwidth and of the cooling rate by several orders of magnitude.

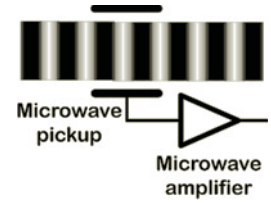



**FIGURE 4.33**
Standard stochastic cooling.

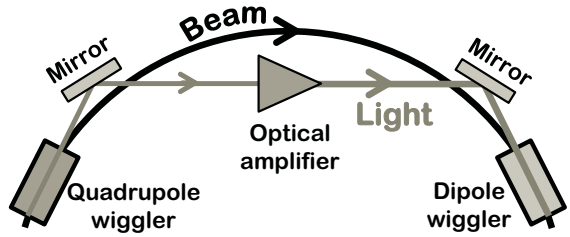



**FIGURE 4.35**
Layout of optical stochastic cooling system in an accelerator.

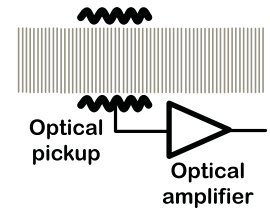



**FIGURE 4.34**
Optical stochastic cooling.

In optical stochastic cooling, quadrupole wigglers and dipole wigglers play the role of pick-up electrodes and kickers as shown in Fig. 4.35.

Beam cooling methods, and stochastic cooling in particular, are excellent examples that demonstrate the discussed AS-TRIZ principle — the evolution of technical systems from microwave frequencies into optical range.

## 4.4 Interact

We have reached the final step of the sequence of Create — Energize — Manipulate — *Interact*.

Once the beam is accelerated, it can be used in a variety of ways — from creating a radiation (synchrotron, betatron) source, a free electron laser, a collider, a spallation neutron source, to using beams for particle therapy, industrial or security applications, or energy applications in *accelerator-driven systems — ADS*. Equally, amplified and compressed laser pulses can be used in a variety of ways; in the context of accelerator physics they can be used as a driver for particle acceleration, as the main component of a Compton X-ray source or of a Photon collider, among other things.

We will touch on some of these further on in this text.

## EXERCISES

4.1 *Chapter materials review*.
The plasma density is $10^{17}$ cm$^{-3}$. A laser light of which wavelength could still penetrate such a plasma? Also, estimate the corresponding plasma frequency.

4.2 *Chapter materials review*.
Taking into account a) Fermat's principle of the least amount of time for light propagation through an optical system, and b) the observation that the diffraction angle should approach the law for reflection from a mirror as the wavelength becomes very short (and hence diffraction becomes less important), explain qualitatively why an optical telescope is needed inside the laser pulse stretcher, but not required in the laser pulse compressor.

4.3 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this chapter in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

4.4 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

# 5 Conventional Acceleration



In this chapter we will discuss conventional acceleration. Beginning with a historical introduction, we will then touch on waveguides, resonant cavities fed by RF power generators, and conclude with an overview of the basics of linacs and longitudinal dynamics.

## 5.1 Historical introduction

The "Livingston plot" presented in Fig.1.6 of Chapter 1 oversees the development of accelerators over the past several decades. These days, we refer to them as conventional accelerators. These accelerators are also human-made, in contrast with, for example, a cosmic accelerator, which has the capacity to produce particles with energies exceeding $10^{20}$ $eV$, or lightning, which produces fairly minimal acceleration.

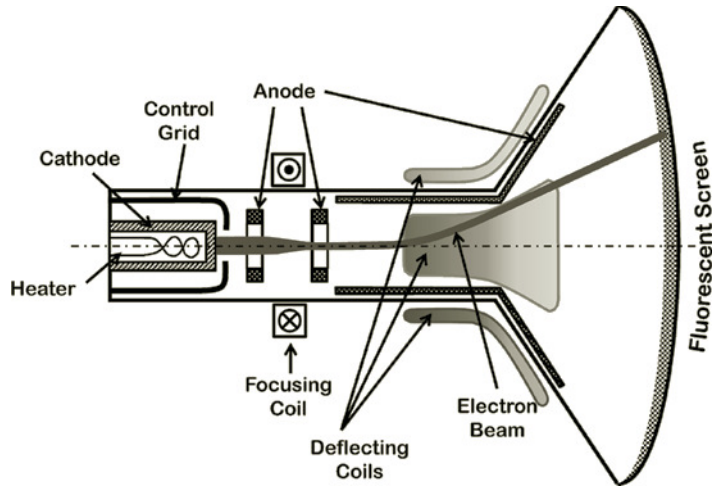



**FIGURE 5.1**
A cathode ray tube TV as an example of an accelerator.

To contrast the composition of older accelerators with that of a contemporary one, we will first look at an accelerator familiar to everyone — a cathode ray tube TV (Fig.5.1), which is a common example of an accelerator with all subsystems present: a beam source, focusing and steering, an accelerating region and an interaction or a target area (i.e., a fluorescent screen).

DOI: 10.1201/b18696-5

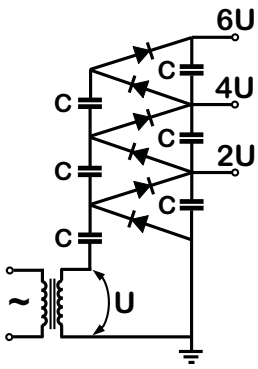



**FIGURE 5.2**
Cockcroft–Walton generator.

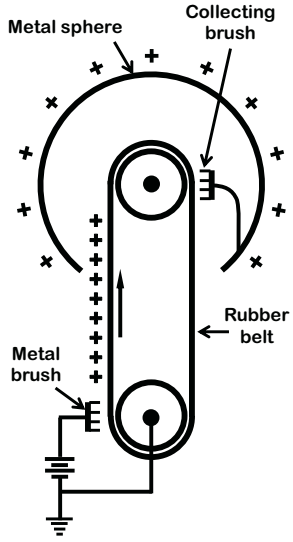



**FIGURE 5.3**
Van der Graaf accelerator.

### 5.1.1 Electrostatic accelerators

In an inspirational push in favor of accelerator development, Rutherford spoke at The Royal Society in 1928, lamenting, "I have long hoped for a source of positive particles more energetic than those emitted from natural radioactive substances."

The first accelerator developed was the Cockcroft–Walton generator which is based on a system of multiple rectifiers (Fig.5.2). In this generator, voltage generated by the cascade circuit

$$U_{tot} = 2Un - \frac{2\pi I}{\omega C}\left(\frac{2}{3}n^3 + \frac{1}{4}n^2 + \frac{1}{12}n\right) \tag{5.1}$$

depends on the number of cells $n$, as well as generated current $I$, capacity $C$ and frequency $\omega$.

The Cockcroft–Walton accelerator helped to make Rutherford's dream a reality. In the first-ever transmutation experiment, the accelerated 700 keV protons were sent onto a lithium target, resulting in the production of helium.

Practical reasons (i.e., size of the device, performance of capacitors and diodes) limited the generated voltages and corresponding energies of the accelerated particles up to about $\sim 4\ MV$. Cockcroft–Walton accelerators can typically generate beam currents of several hundred $mA$ with CW or pulsed particle beams of few $\mu$s pulse lengths.

Another example of an electrostatic accelerators is the "Van der Graaf" (see Fig.5.3), where a metal brush deposits charges onto a rubber belt, carrying the charges into a metal sphere where they are later collected by another brush. The exchange of charges is performed with help from a discharge between the sharp tips of the needles of the brushes and the belt.

The ion source in the "Van der Graaf" accelerator is located inside of the metal sphere, and is charged to a high voltage. An electrical generator (mechanically connected to

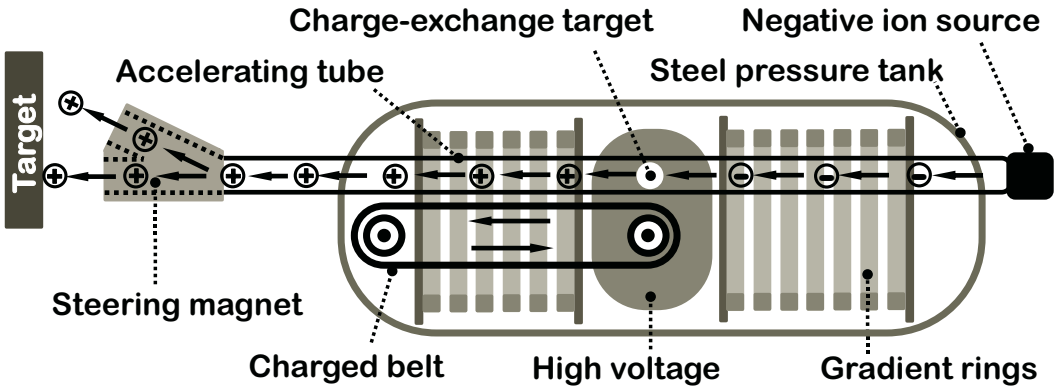



**FIGURE 5.4**
Tandem electrostatic accelerator.

the upper axis of the rubber belt pulley) produces electrical power needed to operate the ion source or any other internal control electronics. The ions produced by the source are accelerated on their way down from the sphere.

With any electrostatic accelerator, it is difficult to achieve high energies due to limitations determined by the size of the vessels. In particular, the highest recorded energy created via the Van der Graaf accelerator was around ~25 MeV.

A version of the Van der Graaf accelerator that can produce twice higher energy of the particles is called a *tandem accelerator* (Fig.5.4). In this case, negative ions are accelerated and, when they reach the charged sphere, they pass through a foil or gas target to perform their *charge exchange*. Now positive, ions continue to accelerate towards ground potential and thus reach twice the voltage of the charged sphere. An additional advantage of the tandem accelerator is having its ion source located at ground potential, which significantly simplifies its operation and maintenance.

Another version of the charge-carrying mechanism is realized in a *pelletron*, where instead of the rubber belt, metal pellets are connected by non-conductive links into a chain (see Fig.5.5).

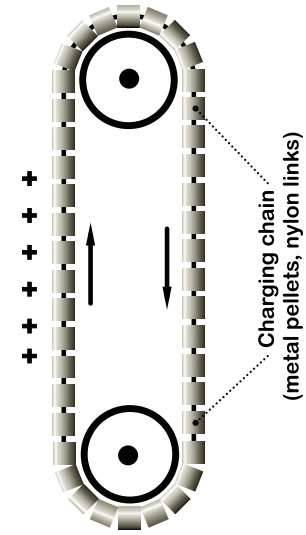



**FIGURE 5.5**
Pelletron charging mechanism.

### 5.1.2 Synchrotrons and linacs

*Synchrotrons* can accelerate particles to much higher energies than electrostatic devices can. Their name is derived from the process of *synchronous* change of the magnetic field of bending magnets according to the growing energy of the accelerated beam (see the quote at right).

*"Particles should be constrained to move in a circle of constant radius thus enabling the use of an annular ring of magnetic field...which would be varied in such a way that the radius of curvature remains constant as the particle gains energy through successive accelerations by an alternating electric field applied between coaxial hollow electrodes." Mark Oliphant, Oak Ridge, TN, 1943.*

Many modern accelerators are synchrotrons (e.g., the Large Hadron Collider at CERN is a synchrotron). These types of accelerators can reach very high energies, limited only (especially for electrons) by synchrotron radiation and the cost of their construction.

As we will momentarily explain, time-varying fields are the necessary conditions for acceleration of charged particles to high energies, and in particular for overcoming the limitations of the maximum achieved energy in the electrostatic accelerators.

Both linear and circular accelerators use EM fields oscillating in resonant cavities to achieve acceleration. In circular accelerators, particles follow an orbit guided by a magnetic field and return to the same accelerating cavity on every turn, while in linac accelerators the particles follow a straight path through a sequence of cavities, as shown in Fig.5.6.

Powerful radio-frequency (RF) systems produce the required powerful electric fields in the resonant cavities. Accelerators of this kind progressed in large part due to the telecommunications industry, which drove the development

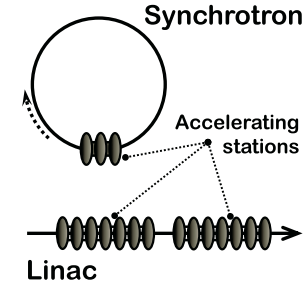



**FIGURE 5.6**
Synchrotron and linac.

of power systems with frequencies ranging from a few MHz to several GHz.

Recalling Maxwell's equation and its integral form

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{or} \qquad \oint_{\partial\Sigma} \mathbf{E} \cdot d\boldsymbol{\ell} = -\frac{d}{dt} \int_{\Sigma} \mathbf{B} \cdot d\mathbf{S}$$

we see that if the particle moves on an enclosed orbit, like in a synchrotron, finite acceleration (i.e., nonzero contour integral of **E**) would not be possible without a time-dependent magnetic field **B**. On the other hand, time-dependent magnetic flux can provide acceleration when — either in a linac or a synchrotron — the EM field oscillates in the resonant cavity and the particles receive a finite energy increment at each pass through the cavities.

### 5.1.3 Wideröe linear accelerator

In 1925, Ising inspired the new technological branch of accelerators by realizing that limitations imposed by corona formation and discharge in electrostatic accelerators can be overcome with the use of alternating voltages. This was carried further by Wideröe who, in 1928, performed the first successful test of the linac based on this principle.

The Wideröe linac has a series of drift tubes arranged along its beam axis and connected with an alternating high frequency RF voltage: $V(t) = V_{\max} \sin(\omega t)$ as shown in Fig.5.7. During the first half of the RF period in the acceleration process, voltage applied to the first drift tube accelerates charged particles emerging from the ion source. Particles then enter the first drift tube and pass through it. Meanwhile, the direction of the RF field reverses without affecting the particles — seeing as the drift tube acts as a Faraday cage and shields the particles from external fields. When the particles reach the gap between the first and the second drift tubes, they accelerate and the process repeats in the following gaps as well.

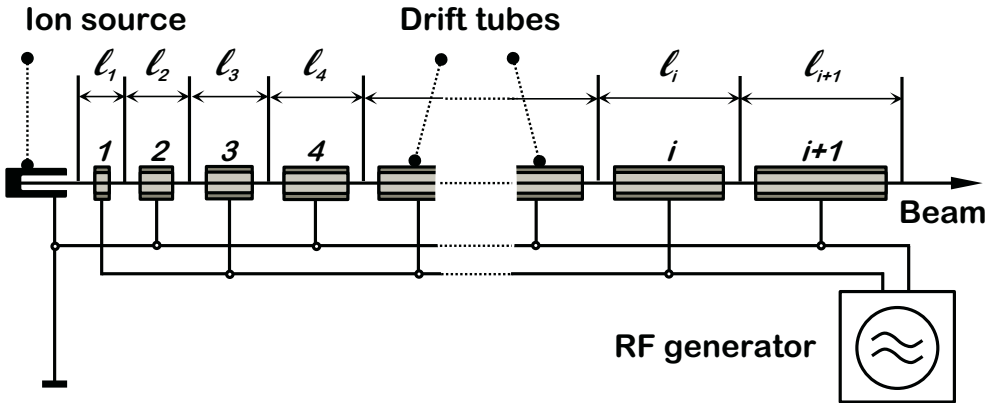



**FIGURE 5.7**
Wideröe linear accelerator.

The energy of the particle in Wideröe linac after passing the i-th drift tube is

$$E_i = iqV_{\max} \sin \Psi_0 \tag{5.2}$$

where $q$ is the charge of the particle and $\Psi_0$ is average phase of the RF voltage that particles feel as they cross the gaps (see Fig.5.8). As we can see, the energy is proportional to the number of stages $i$ passed by the particle. Furthermore, the largest voltage to ground in the entire system never exceeds $V_{\max}$. This lets us reach high energies without using voltage levels, which can cause electrical breakdown.

The accelerating gaps between drift tubes in the Wideröe linac must increase in sync with the monotonically increasing velocity of the particle. Taking into account that the half-period of RF $\tau_{RF}/2$ should correspond to a particle passing with velocity $v_i$ through one drift section, we write the distance between $i$-th and $(i+1)$-th gaps as

$$\ell_i = \frac{v_i \tau_{RF}}{2} = \frac{1}{f_{RF}} \quad \overline{\frac{iqV_{\max} \sin \Psi_0}{2m}}$$

which we expanded using Eq.5.2.

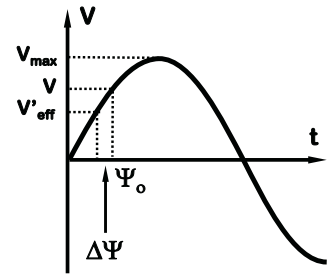



**FIGURE 5.8**
Voltage in Wideröe linac.

### 5.1.4 Alvarez drift tube linac

The Alvarez linac is conceptually quite similar to the Wideröe linac. It differs in that, in an Alvarez linac, the accelerating voltage at individual drift tubes is created by an RF wave in a container (a tank made of a good conductor such as copper), in which the drift tubes are located (see Fig.5.9). The drift tubes may have magnets installed inside to focus the beam during acceleration.

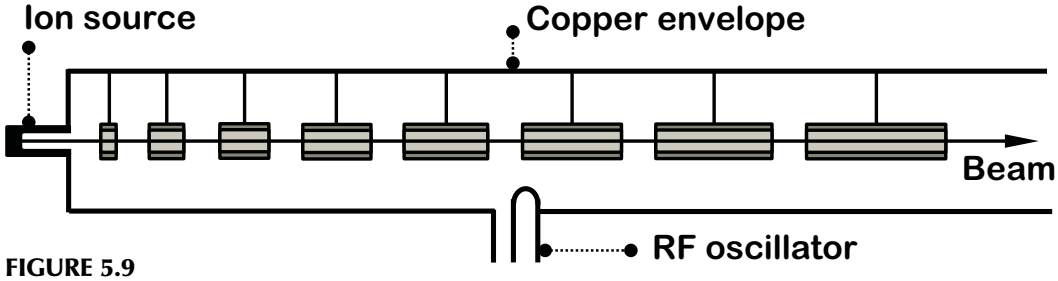



**FIGURE 5.9**
Alvarez drift tube linac.

The drift tube linacs are still used (particularly in hadrons), but they are being replaced by better-performing *RFQ-structures* (Fig.5.10). Due to periodic transverse variations of their shape, such structures allow for the creation of not only accelerating fields, but also focusing fields. Recent

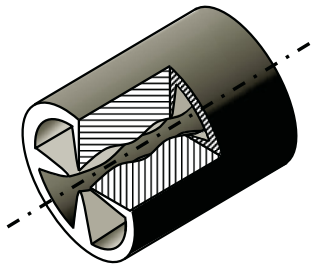


**FIGURE 5.10**
RFQ structure.

progress via 3D computer simulations of RF fields and beam dynamics resulted in a widespread use of RFQs.

Most of the complicated linac designs discussed above are applicable to hadrons, which (in practically achievable accelerating gradients) become relativistic rather slowly, after a hundred or so meters of acceleration. Correspondingly, the size and shape of accelerating cavities need to vary along the hadron linac, so as to match their increasing velocity. For electrons, which have already become highly relativistic after a few tens of cm of acceleration, relatively simpler linac design is possible, in which cavities of the same size and shape are regularly placed along the acceleration path.

### 5.1.5 Phase focusing

As we have seen in the previous section, the energy transferred to particles in a drift tube linac depends on the voltage amplitude $V_{\max}$ and phase $\Psi_0$. We should note that a small deviation of the nominal voltage $V_{\max}$ would result in a particle velocity that no longer matches the design velocity fixed by the length of drift sections. In this case, the particles would undergo a phase shift relative to the RF voltage and the synchronization of particle motion with respect to RF field will be eventually lost.

The system can be made to self-adjust to such small deviations if we use $\Psi_0 < \pi/2$ so that the effective accelerating voltage is $V_{\text{eff}} < V_{\max}$, as shown in Fig.5.8. In this instance, if a particle gains too much energy in the preceding stage and is travelling faster than the ideal particle and arrives at the next acceleration stage earlier, it will then feel the average RF phase $\Psi = \Psi_0 - \Delta\Psi$ and will be accelerated by the voltage

$$V'_{\text{eff}} = V_{\max} \sin(\Psi_0 - \Delta\Psi) < V_{\max} \sin\Psi_0 \qquad (5.3)$$

which is below the ideal voltage. The particle will thus gain less energy and will slow down and return to the nominal velocity. The process will then repeat and the particles will continuously oscillate about the nominal phase $\Psi_0$, exhibiting the phenomena of *phase focusing*. If dissipation is present in the system (such as SR damping), these oscillations would eventually be forced to decay and the particles would gather around the synchronous phase. A similar effect of bunching can happen without damping, due to acceleration.

### 5.1.6 Synchrotron oscillations

The phase focusing of particles in accelerators manifests itself as the periodic longitudinal particle motion about the nominal phase, and is called *synchrotron oscillation*.

In circular accelerators, as the ideal particle encounters the RF voltage at exactly the nominal phase on each revolution, the RF frequency $\omega_{RF}$ must be an integer multiple of the

revolution frequency $\omega_{rev}$

$$h = \omega_{RF}/\omega_{rev} \tag{5.4}$$

where $h$ is called the *harmonic number* of the ring: an integer chosen to allow matching the frequency of available RF generators.

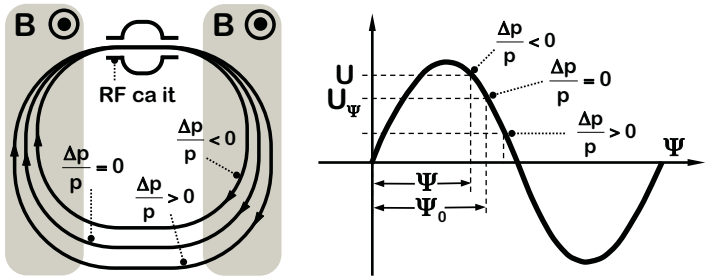



**FIGURE 5.11**
Synchrotron oscillations.

For *relativistic* particles, the phase focusing is enabled by the dependence of the circular path on energy, as shown in Fig.5.11. The stable point in this case is on the declining slope of the sine wave, as particles with $\Delta p/p > 0$ will travel on the longer path, come to the RF cavity in the next turn somewhat later, and ultimately lower their acceleration. As the reader can already guess, for a particle that is not yet relativistic and whose velocity depends on energy, the situation is quite different — we will discuss this further later on in this chapter.

## 5.2 Waveguides

Prior to engaging in detailed discussion of RF accelerating cavities, let's introduce — via simple considerations and analogies — the basic properties of waveguides, as they provide important and intuitive understanding applicable for accelerating structures.

### 5.2.1 Waves in free space

The velocity of an EM wave in a vacuum and in a medium is

$$\text{vacuum}: \quad v = c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \quad , \quad \text{medium}: \quad v = \frac{1}{\sqrt{\varepsilon_0 \varepsilon_r \mu_0 \mu_r}} \tag{5.5}$$

where $\varepsilon_r$ is the dielectric constant and $\mu_r$ is the magnetic permeability of the medium.

The amplitudes of electric and magnetic fields in an EM wave in a vacuum are exactly the same if expressed in Gaussian-cgs units (demonstrating the naturalness of this system of units) and relates as $E = cB$ in SI units.

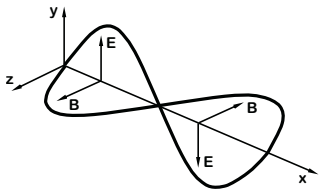



**FIGURE 5.12**
TEM wave in free space.

A plane EM wave with transverse electric and magnetic fields (TEM wave) propagating in free space in x-direction is shown in Fig.5.12. The quantity

$$\mathbf{P} = (\mathbf{E} \times \mathbf{B})/\mu_0 \tag{5.6}$$

is called the *Poynting vector* and equals to the local power flux.

### 5.2.2 Conducting surfaces

In the derivations presented in this section, we discuss the behavior of EM waves bounded in metal boxes; therefore, we need to recall the boundary conditions of a wave at a perfectly conducting metallic surface.

On the surface of a perfect conductor, the tangential component of an electric field $E_{\parallel}$ and the normal component of a magnetic field $B_{\perp}$ will vanish, as illustrated in Fig.5.13.

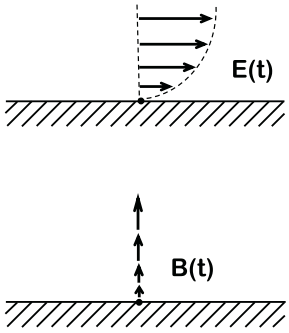



**FIGURE 5.13**
Boundary conditions on perfectly conducting surfaces.

A non-ideal surface with conductivity $\sigma$ is characterized by *skin depth*; an EM wave entering a conductor is dampened to $1/e$ of its initial amplitude at the depth

$$\delta_S = \frac{1}{\sqrt{\pi f \ \mu_0 \mu_r \sigma}} \tag{5.7}$$

This allows us to introduce the notion of *surface resistance*

$$R_{surf} = \frac{1}{\sigma \, \delta_S} \tag{5.8}$$

which plays an important role in determining performance of accelerating cavities.

### 5.2.3 Group velocity

In preparation to discuss dispersion properties of waveguides and RF structures, let us recall the derivation of *group velocity* — the propagation velocity of energy (and information) in an EM wave.

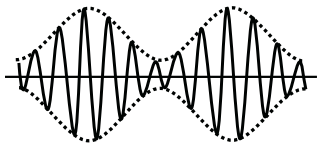


**FIGURE 5.14**
Two-wave interference.

Consider the interference between two continuous waves of slightly different frequencies $\omega \pm d\omega$ and wavenumbers $k \pm dk$ (see Fig.5.14):

$$\begin{aligned} E &= E_0 \sin[(k+dk)x - (\omega + d\omega)t] \\ &\quad + E_0 \sin[(k-dk)x - (\omega - d\omega)t] \\ &= 2E_0 \sin[kx - \omega t] \cos[dk\, x - d\omega\, t] \\ &= 2E_0 f_1(x,t)\, f_2(x,t) \end{aligned} \tag{5.9}$$

The last line of Eq.5.9 contains two functions, $f_1$ and $f_2$. The first function corresponds to a continuous wave with the mean wavenumber and frequency: $f_1(x,t) = \sin[kx - \omega t]$. In this wave, any given phase is propagated such that $kx - \omega t$ remains constant, which gives us the equation for the *phase velocity* of the wave:

$$v_p = \frac{\omega}{k} \tag{5.10}$$

The same can be obtained by requesting the *convective derivative* $(\partial/\partial t + v_p \partial/\partial x)$ of $f_1$ to be equal to zero, which again results in

$$v_p = -\frac{\partial f_1(x,t)/\partial t}{\partial f_1(x,t)/\partial x} = \frac{\omega}{k}$$

*Convective derivative — the term originates from fluid mechanics — is the derivative taken with respect to a moving coordinate system.*

The second function in Eq.5.9 describes the evolution of the envelope of the pattern: $f_2(x,t) = \cos[dk\,x - d\omega\,t]$. Again, any point in the envelope propagates such that the quantity $dk\,x - d\omega\,t$ remains constant and therefore its velocity, i.e., the group velocity, is given by

$$v_g = -\frac{\partial f_2(x,t)/\partial t}{\partial f_2(x,t)/\partial x} = \frac{d\omega}{dk} \tag{5.11}$$

This prepares us for discussion of the notions of *dispersion* and *group* or *phase velocity* of a waveguide.

### 5.2.4 Dispersion diagram for a waveguide

We will start our conversation regarding wave propagation down a waveguide using two extreme cases.

First of all, if the wavelength $\lambda$ of an EM wave in free space is much shorter than the transverse size $a$ of the waveguide $\lambda \ll a$ (as shown in Fig.5.15, case (i)), then the waveguide does not matter, and we expect the dispersion at large $\omega$ to approach the equation for free space (i.e. $\omega/c = k$). With the goal of deriving the dependence of frequency $\omega$ against wavenumber $k = 2\pi/\lambda$ in a waveguide, let us place a corresponding segment on a waveguide dispersion curve at a high frequency (Fig.5.16, case (i)).

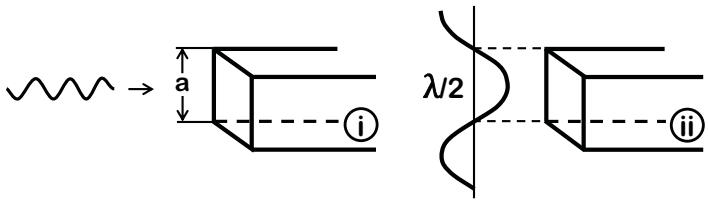



**FIGURE 5.15**
Waves in a waveguide, two extreme cases.

Another extreme case is shown in Fig.5.15, (ii), when half of a wavelength in free space equals the waveguide transverse size. As can be seen from this diagram, the longest wavelength for which the boundary conditions at a perfectly conducting surface of the waveguide can still be satisfied, is given by $\lambda/2 \leq a$. This defines the cut-off parameters $\lambda_c = 2a$ or $\omega_c = \pi c/a$; that is, waves with wavelengths longer than $\lambda_c$ cannot propagate in the waveguide.

As Fig.5.15 (ii) suggests, the case of $\omega = \omega_c$ corresponds to an infinite wavelength in in the waveguide, or $k = 0$. We thus plot a corresponding point in Fig.5.16.

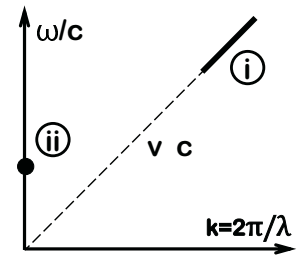



**FIGURE 5.16**
Dispersion of a waveguide, two extreme cases.

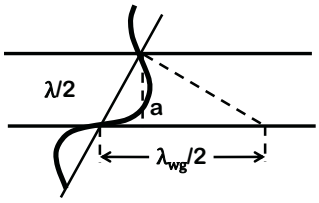



**FIGURE 5.17**
Intermediate case.

In the intermediate region of frequencies, the dispersion curve should connect the point (ii) and region (i) in Fig.5.16. Looking at the wave in a waveguide from a simple geometrical point of view (Fig.5.17) and considering corresponding similar triangles, one can write

$$\frac{\omega}{c}^2 = k^2 + \frac{\omega_c}{c}^2 \tag{5.12}$$

thus describing dispersion of the wave in a rectangular waveguide, graphically shown in Fig.5.18.

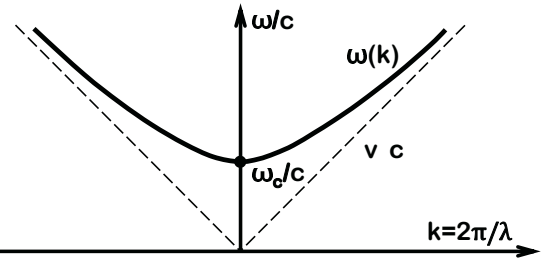



**FIGURE 5.18**
Dispersion of a waveguide.

Eq.5.12 suggests that, for any wavenumber $k$, the frequency is always greater than the cut-off frequency. Looking at the slope (i.e. derivative $d\omega/dk = v_g$) of the curve in Fig.5.18 we can also observe that the longer the wavelength or lower the frequency, the slower the group velocity, and at the cut-off frequency no energy flows along the waveguide.

Eq.5.12 and Eq.5.10 also help find that

$$v_p v_g = c^2 \tag{5.13}$$

which tells us that, in a waveguide, the phase velocity is always *larger* than the speed of light.

### 5.2.5 Iris-loaded structures

We can conclude from the previous section that acceleration in a waveguide is not possible because the phase velocity of the wave exceeds that of light. Particles that travel slower than the wave would be periodically accelerating or decelerating, achieving zero net acceleration when averaged over a long time interval.

In order to make the acceleration possible, one needs to modify the waveguide to reduce the phase velocity to an appropriate value below the speed of light, so as to match the velocity of the particle.

Reduction of the phase velocity can be achieved by using iris-shaped screens installed into the waveguide with a constant step along the axes shown in Fig.5.19.

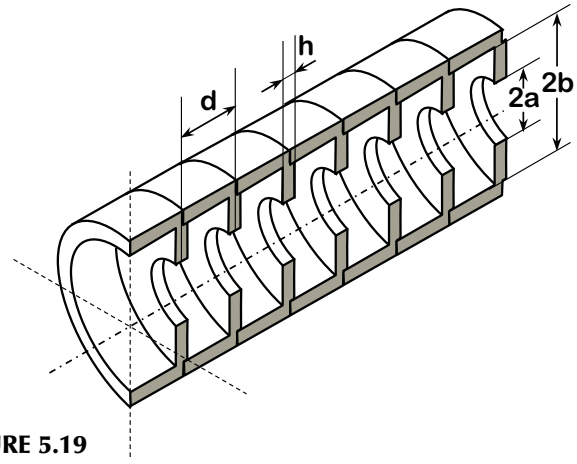



**FIGURE 5.19**
Iris-loaded accelerating structure.

The dispersion relation in a waveguide $\omega = c\sqrt{k_z^2 + (2\pi/\lambda_c)^2}$ changes due to installation of the irises. Qualitatively, behavior at low $k$ resembles that of the waveguide curve, defined by the smallest diameter of the irises. At higher $k$, with the installation of irises, the curve flattens off and crosses the boundary of $v_\varphi = c$ in the region of $k_z = \pi/d$ as illustrated in Fig.5.20.

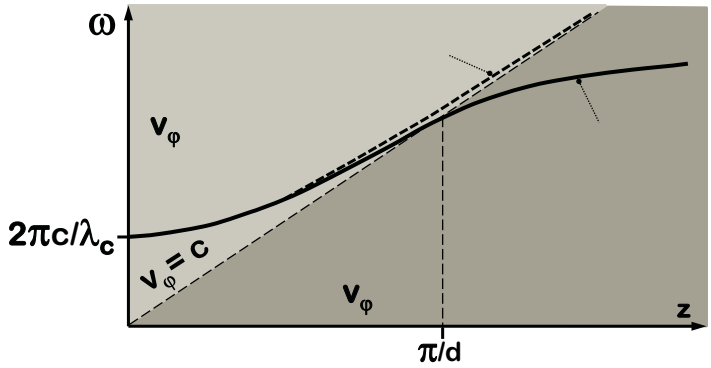



**FIGURE 5.20**
Qualitative behavior of dispersion curve in iris-loaded structures.

With a properly selected iris separation $d$ and other parameters of the irises, the phase velocity of the iris-loaded structure can be set to any value, making it suitable for particle acceleration with arbitrary $v/c$.

An extended version of the dispersion diagram curve of an iris-loaded structure can also be constructed while noting that adding a multiple of $2\pi/d$ to the wavenumber $k$ as

$$k_n = k_0 + \frac{2n\pi}{d}$$

would still satisfy the periodic boundary conditions at the

irises. Therefore, the dispersion curve repeats itself at space harmonics corresponding to different integers, $n$, as shown in Fig.5.21.

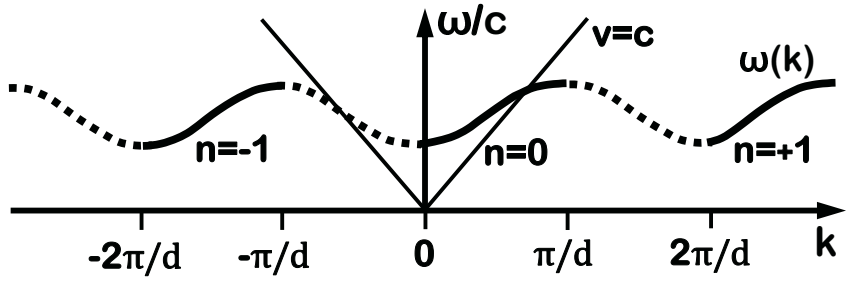



**FIGURE 5.21**
Extended dispersion diagram of an iris-loaded structure.

The first rising slope of the dispersion curve shown in Fig.5.21 is usually used for acceleration.

## 5.3 Cavities

In this section, we will consider general properties of resonant cavities, their quality factors, shunt impedance, and will introduce the definition of the resonance modes.

### 5.3.1 Waves in resonant cavities

In preparation for a discussion about the resonance modes in the cavity, we first recall a general solution of the wave equation, which can be written as

$$W(r,t) = Ae^{i(\omega t + k\cdot r)} + Be^{i(\omega t - k\cdot r)} \tag{5.14}$$

This describes the sum of two waves — one moving in one direction and the other in the opposite direction.

In the case wherein the wave is totally reflected from a conductive surface, both amplitudes need to be the same, i.e., $A = B$, and we can therefore rewrite Eq.5.14 as

$$W(r,t) = Ae^{i\omega t}(e^{ik\cdot r} + e^{-ik\cdot r}) = 2A\cos(k\cdot r)e^{i\omega t} \tag{5.15}$$

This equation describes the field configuration with a static in time amplitude $2A\cos(k\cdot r)$, therefore corresponding to a standing wave.

Consider now that the waveguide we discussed earlier has a finite length $\ell$ and its entrance and exit are closed by two conducting surfaces, forming a resonance cavity. The resonant wavelengths of this cavity can be determined noting that a stable standing wave can form in this fully enclosed cavity

if we assume that the following condition is satisfied

$$\ell = q\frac{\lambda_z}{2} \quad \text{with} \quad q = 0,\ 1,\ 2,... \tag{5.16}$$

Therefore, only certain well-defined wavelengths $\lambda$ are present in the cavity.

Near the resonant wavelength, the resonant cavity behaves like an oscillator with a high *quality factor* Q, allowing it to build up high voltages that can be used for particle acceleration. The cavities are often modeled as electrical oscillators, with their Q-value determined by losses of equivalent individual coils, capacitors and resistances of the circuit model.

### 5.3.2 Pill-box cavity

An enclosed section of a waveguide (either rectangular or cylindrical) forms the simplest RF cavity, called a *pill-box cavity*.

The conventionally accepted classification of the modes in pill-box cavities separates the cases of *transverse electric* or TE modes (zero electric field along the axis) and *transverse magnetic* or TM modes (zero magnetic field along the axis).

As Eq.5.16 suggests, many modes can exist in a cavity, as defined by the corresponding dimension of the pill-box, and the integer number of the mode. The corresponding integer indexes are used to identify a particular mode. In a rectangular pill-box, a mode can be called $TE_{klm}$ or $TM_{klm}$ where the integer indexes indicate the number of half-wavelength variations across the corresponding dimension $(x, y, z)$ of the cavity.

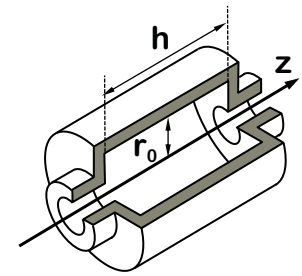



**FIGURE 5.22**
Cylindrical pill-box cavity.

Cylindrical pill-box cavities are very common in accelerators. An example of a cylindrical pill-box cavity with holes for the beam is shown in Fig.5.22. The classification of the modes in cylindrical cavities is very similar, $TE_{klm}$ or $TM_{klm}$, but in this case the indexes refer to the polar coordinates ($\varphi$, $r$ and $z$).

Examples of modes in cylindrical pill-box cavities, with longitudinal electric fields and no variation over $\varphi$ ($TM_{0lm}$), which are therefore suitable for use as an accelerating cavity, are shown in Fig.5.23.

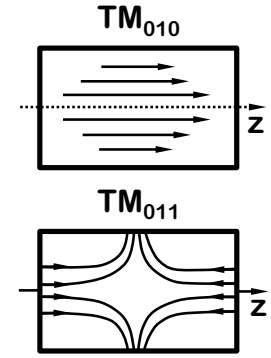



**FIGURE 5.23**
Examples of pill-box cylindrical cavity modes with electric field lines shown.

### 5.3.3 Quality factor of a resonator

The quality factor of a resonator — Q — is defined as the ratio of the energy stored in the cavity to the energy dissipated per oscillating cycle, divided by $2\pi$

$$Q = \frac{W_s}{W_d} = \omega\frac{W_s}{P_d} \tag{5.17}$$

Here, $W_s$ is the energy stored in the cavity, $W_p$ is the energy dissipated per cycle divided by $2\pi$, $P_d$ is the power dissipated in the cavity walls and $\omega$ is the frequency of the cavity.

The stored energy over the cavity volume is

$$W_s = \frac{\varepsilon_0}{2} \int |\mathbf{E}|^2 dv + \frac{\mu_0}{2} \int |\mathbf{H}|^2 dv \tag{5.18}$$

*In vacuum $H = B/\mu_0$ in SI units.*

where the first integral is the energy stored in the E-field and the second integral corresponds to the energy in the H-field. In an EM wave in space or a cavity, the energy oscillates back and forth between these two contributions.

The losses in the cavity are calculated by taking into account the finite conductivity $\sigma$ of the cavity walls. Since the linear density of the current $\mathbf{j}$ along the walls of a perfect conductor can be written as

$$\mathbf{j} = \mathbf{n} \times \mathbf{H} \tag{5.19}$$

where $\mathbf{n}$ is the vector normal to the surface, we can therefore equate the power dissipated in the cavity walls to

$$P_d = \frac{R_{surf}}{2} \int_s |\mathbf{H}|^2 ds \tag{5.20}$$

where the integral is taken over the inner surface of the conductor, and the surface resistance is given by Eq.5.8.

### 5.3.4 Shunt impedance — $R_s$

The so-called *shunt impedance* $R_s$ relates the accelerating voltage $V$ to the power that needs to be fed into the cavity to compensate for the dissipation in the walls $P_d$.

The accelerating voltage along the path followed by the beam in an electric field $E_z$ is

$$V = \int_{pass} E_z(x, y, z)\, d\ell \tag{5.21}$$

and is taken as peak-to-peak value. The shunt impedance is then defined as

$$R_s = \frac{V^2}{2P_d} \tag{5.22}$$

and is another important characteristic of an accelerating cavity.

### 5.3.5 Energy gain and transit-time factor

The energy gain of a particle as it travels a distance through the accelerating structure depends only on potential difference crossed by particle:

$$U = K\sqrt{P_{RF} l R_s} \tag{5.23}$$

where $P_{RF}$ is the RF power supplied to the cavity, $l$ is the length of the accelerating structure, $R_s$ is the shunt impedance and $K$ is a correction factor (typically $\approx 0.8$).

In accelerators such as drift tube linacs, a so-called "transit-time factor" also plays a role and needs to be taken into account in order to evaluate the average energy gain.

Consider an accelerating gap corresponding to the space between drift tubes in a linac structure (Fig.5.24). The accelerating field in this gap is uniform along the axis and depends sinusoidally on time $E_z = E_0 \cos(\omega t + \phi)$, where the phase $\phi$ refers to the particle in the middle of gap $z = 0$ at $t = 0$. The field varies as the particle traverses the gap, making the cavity less efficient and the resultant energy gain only a fraction of the peak voltage.

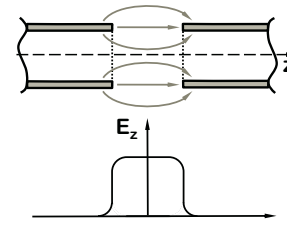



**FIGURE 5.24**
The RF gap — space between entrance and exit irises of cavity resonator in drift tube linac.

The *transit-time factor* $\Gamma$ is the ratio of the energy actually given to a particle passing the cavity center at the peak field to the energy that would be received if the field were constant with time at its peak value. Taking into account that the energy gained over the gap $G$ is:

$$V = \int_{-G/2}^{+G/2} E_0 \cos(\omega t + \phi)\, dz = E_0 G \cos\phi \frac{\sin(\omega G/2\beta c)}{\omega G/2\beta c} \qquad (5.24)$$

we write the following expression for the transit-time factor $\Gamma = \sin(\omega G/2\beta c)/(\omega G/2\beta c)$.

### 5.3.6 Kilpatrick limit

The performance of any normal conducting accelerating structure depends on its susceptibility to RF breakdown (which can occur at very high fields). Empirically derived around 1950, the *Kilpatrick limit* expresses the relation between the accelerating frequency and maximum achievable accelerating field:

$$f\,[MHz] = 1.64\, E_k^2 e^{-8.5/E_k} \qquad (5.25)$$

where $E_k$ is expressed in [MV/m] and is depicted by the lower curve in Fig.5.25.

Significant efforts and technological developments intended to improve surface quality and cleanness have resulted in a considerable increase of achievable accelerating gradients. In particular, Wang and Loew's empirical formula, devised in 1997, suggests the following behaviors:

$$E\,[MV/m] = 220 f^{1/3} \qquad (5.26)$$

where $f$ is expressed in [GHz] — shown by the upper curve in Fig.5.25.

The $E \sim f^{1/3}$ dependence in Eq.5.26 was, for a long time,

an inspiration and a driving force for developing higher gradients acceleration at higher (multi-tens of GHz) frequencies. This dependence, however, eventually was not confirmed for the practical parameters of the accelerator RF pulses, where it was observed that in these regimes the maximum gradients appear to be rather independent of the frequency.

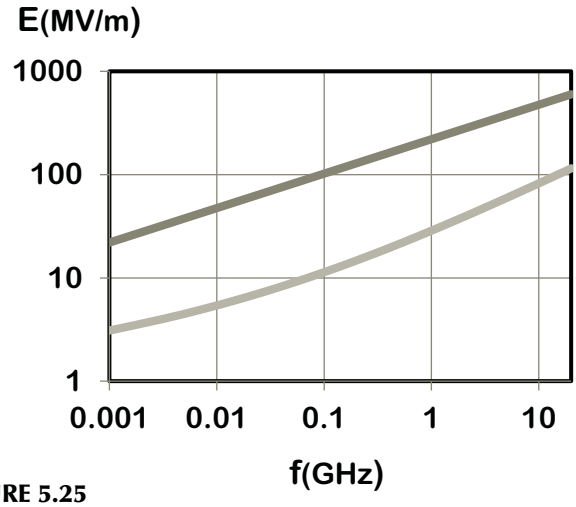



**FIGURE 5.25**
Breakdown Kilpatrick limit (lower curve) and Wang–Loew limit (upper curve).

## 5.4 Power sources

Sinusoidal power ranging from a few kW to a few MW is needed to drive the accelerating structures. This is commonly achieved by using RF power amplifiers such as triodes and tetrodes (operating from a few MHz to a few hundred MHz), inductive output tubes (suitable for CW applications, providing tens of kW at a high efficiency) and klystrons (which typically operate above a few hundred MHz) and have proven to be the most effective power generators for accelerator applications at higher frequencies.

### 5.4.1 IOT — inductive output tubes

The *inductive output tube*, or IOT, (invented by Andrew Haeff around 1939) is based on the principle that a toroidal cavity surrounding an electron beam of oscillating intensity could extract power from the beam without intercepting the beam itself.

The oscillating EM fields carried with the beam excite the modes found inside the toroidal cavity, which allows RF energy to be transferred from the beam to a waveguide or coaxial cable connected to the resonator via a coupling loop.

A schematic design of an IOT is shown in Fig.5.26.

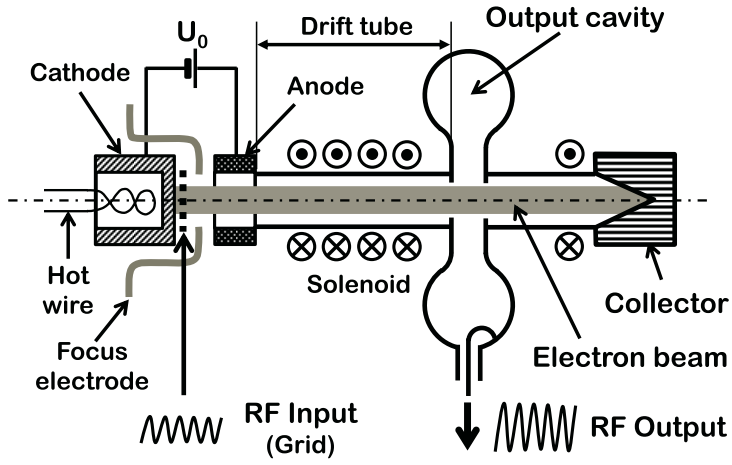



**FIGURE 5.26**
Schematic of an inductive output tube.

### 5.4.2 Klystron

Building on the success of the IOT, Sigurd and Russell Varian added a cavity resonator at the beginning of the beamline in order to provide a signal input for the inductive output tube. This input resonator acts as a pair of control grid electrodes performing *velocity modulation* in the electron beam. This results in bunching developed in the beam by the time it arrives at the output resonator. In the latter the amplitude variation will be converted to energy extracted from the beam. The Varian brothers called their invention a *klystron*.

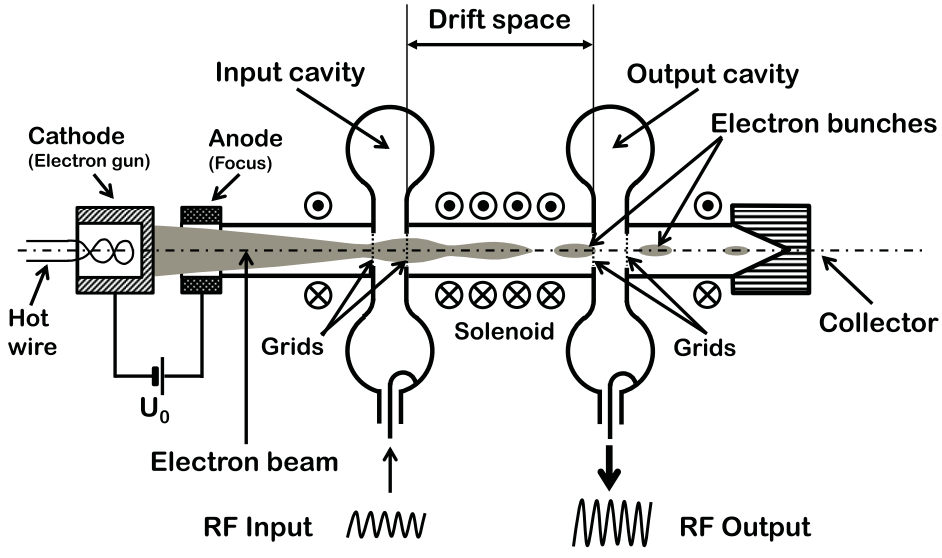



**FIGURE 5.27**
Schematic of a klystron.

A schematic design of a klystron is shown in Fig.5.27.

Klystrons are similar to small linear accelerators, and operate as follows. First, electrons are emitted from a round cathode, which has a large surface area. Electrons are then accelerated by a voltage of a few tens of kV. A round beam is formed with a current from a few amperes to tens of amperes. The beam is focused by near-cathode electrodes and is further focused by the solenoid field (which is essential to ensuring effective beam transportation). The particles leaving the cathode pass through an input cavity operating at $TM_{011}$ mode, which is fed from an external pre-amplifier.

*Permanent magnet focusing has been used in klystrons with the intention to increase their wall-plug efficiency.*

The klystron output power is given by

$$P_{klystron} = \eta U_0 I_{beam} \tag{5.27}$$

where $U_0$ is the klystron supply voltage (e.g., 350 kV), $I_{beam}$ is the electron beam current (e.g., 420 A) and $\eta$ is the klystron efficiency (e.g., 45%). The numbers given above as examples correspond to the SLAC 5054 klystron running at 2.856 GHz. The klystron efficiency (which typically ranges from 45% to 65%) is one of the most important parameters and is the subject of continuous innovations.

IOTs, klystrons and similar RF power generators — while still extremely popular — are gradually being replaced by solid-state devices, for frequencies lower than 1 GHz in particular. The solid-state RF power systems have the advantage of compactness, higher efficiency and also reliability. The modular design of solid-state RF power sources, when each module contributes only a small fraction of the total power, makes it possible to significantly increase the reliability of the entire system.

*The trend of RF power sources technology development is a good illustration of TRIZ conclusions about the evolution of technical systems.*

### 5.4.3 Magnetron

Magnetron is an RF power source popular in particular for CW applications. In a magnetron the cylindrical cathode is located in the center and magnetic field is applied along the axis of the device. When the electrons are moving from the cathode to the anode, the magnetic field turns them and makes them move on spiral trajectories, creating azimuthal variations of the electron trajectories.

At certain parameters (of the voltage and magnetic field) the azimuthal variations of the electron trajectories will match the azimuthal spacing of the cavities arranged around the perimeter on the anode body, exciting fields in the cavities. The exited fields in their turn will enhance the velocity modulation and thus the spatial density modulation of the electron beam. Increased modulations will further increase the fields generated in the cavities. The magnetron amplitude will grow until saturation (due to effects related to the electron beam space charge) and remain constant.

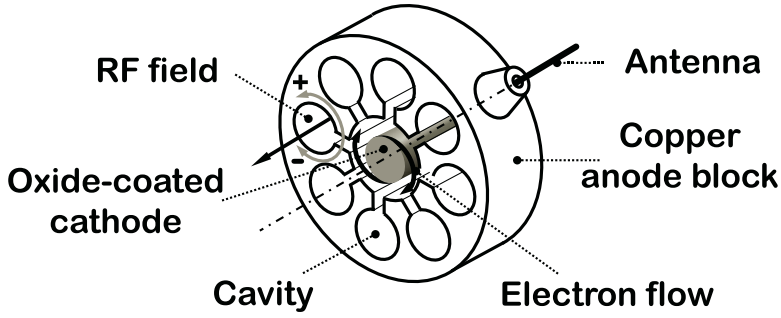



**FIGURE 5.28**
Schematic of a magnetron.

A schematic of a magnetron is shown in Fig.5.28.

A magnetron, in contrast to klystrons and IOTs, is not an amplifier, but a generator, which starts from initial noise. Therefore, the phase of the RF power generated by a magnetron is arbitrary in principle.

Correspondingly, one of the difficulties of using magnetrons in accelerators when more than one magnetron is required is the difficulty of phase-locking those devices.

### 5.4.4 Powering the accelerating structure

An accelerating structure of a linac or other accelerator is typically fueled by pulsed power RF tubes — klystrons.

RF power is transported from a klystron in a different mode than that used for acceleration. Therefore, the *RF cou-*

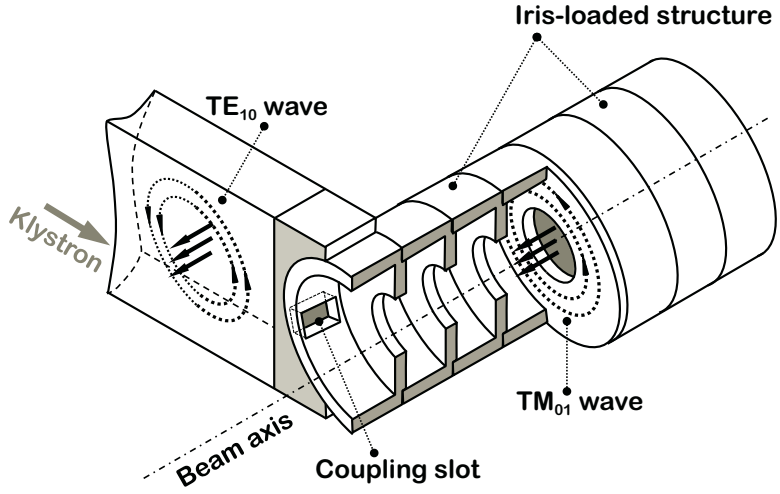



**FIGURE 5.29**
Feeding RF power into an accelerating structure. Field lines show electric and magnetic fields of the corresponding cavity modes.

*pler* of the acceleration cavity must ensure an appropriate mode conversion.

It is typical to feed the RF power into the linac structure by a $TE_{10}$ mode of EM wave in a rectangular waveguide. The mode is then converted by a coupling slot into a cylindrical $TM_{01}$ mode in the accelerating cavity, as illustrated in Fig.5.29.

## 5.5 Longitudinal dynamics

In this section, we will discuss the basics of longitudinal dynamics in travelling and standing wave linacs, as well as in synchrotrons.

### 5.5.1 Acceleration in RF structures

Particle acceleration in linacs is achieved with RF structures, using EM modes with the electric field pointing in the longitudinal direction (the direction of the charged particle's motion). The RF electric field can be provided by either *travelling wave* structures or *standing wave* structures.

The acceleration conditions demand that the phase velocity of the *travelling wave* and the particle velocity be equal, so therefore disk-loaded structures are used to slow down the phase velocity of the electric field $v_p < c$ to achieve synchronism. In an appropriately synchronized travelling wave, the bunch of charge particles experience a constant electric field

$$E_z = E_0 \cos(\phi) \tag{5.28}$$

as illustrated in Fig.5.30 (left plot).

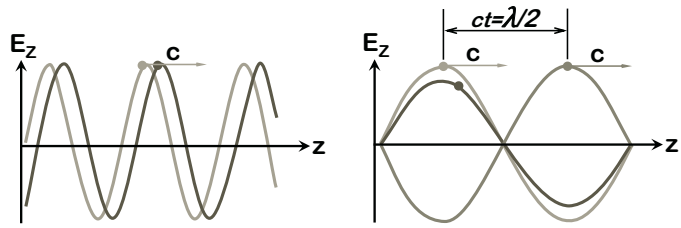



**FIGURE 5.30**
Acceleration in a travelling wave structure (left) and in a standing wave structure (right). The wave and particles' position in different moments of time are shown.

In a *standing wave* structure, the electromagnetic field is the sum of two travelling waves running in opposite directions. Only the forward-travelling wave takes part in the acceleration process.

The electric field that the particle bunch observes in a

standing wave is a varying function of time

$$E_z = E_0 \cos(\omega t + \phi) \sin(kz) \tag{5.29}$$

as illustrated in Fig.5.30 (right plot).

The standing wave, despite its seeming disadvantage in comparison with the travelling wave (in terms of the average field seen by the bunch), is actually much more suitable in certain cases, such as for superconducting cavities.

### 5.5.2 Longitudinal dynamics in a travelling wave

Consider a particle moving in the $E$ field of a travelling wave

$$E_z = E_0 \cos(\omega t - kz) \tag{5.30}$$

with a phase velocity $v_p = \omega/k$. The equations that describe the particle motion in the longitudinal plane in this field are

$$\frac{dp_z}{dt} = eE_0 \cos(\omega t - kz) \quad \text{and} \quad \frac{d\mathcal{E}}{dt} = eE_0 \dot{z} \cos(\omega t - kz) \tag{5.31}$$

We will define the synchronous particle as

$$\frac{dE_s}{dt} = eE_0 v_s \cos\varphi_s \tag{5.32}$$

and for any other particle, we will use, as coordinates, the deviations from the energy $W$ and position $u$ of the synchronous particle

$$\mathcal{E} = E_s + W \quad \text{and} \quad z = z_s + u \tag{5.33}$$

Then, after changing variables to

$$\varphi = kz - \omega t = \varphi_s - \frac{\omega}{v_s} u \tag{5.34}$$

we will obtain the system of equations for a particle motion in a travelling wave:

$$\frac{dW}{ds} = eE_0 [\cos\varphi - \cos\varphi_s] \quad , \quad \frac{d\varphi}{ds} = -\frac{\omega}{\beta_s^3 \gamma_s^3 c} \frac{W}{mc^2} \tag{5.35}$$

These describe the motion in the so-called "RF bucket" in a longitudinal phase space ($\varphi$ , W) and feature stable enclosed trajectories as well as unstable trajectories. We will discuss the phase space trajectories and motion in the "RF bucket" in detail in the following section, after deriving similar equations for the case of acceleration in a synchrotron.

### 5.5.3 Longitudinal dynamics in a synchrotron

Acceleration in a synchrotron is provided by the longitudinal electric fields generated in RF cavities placed on the orbit.

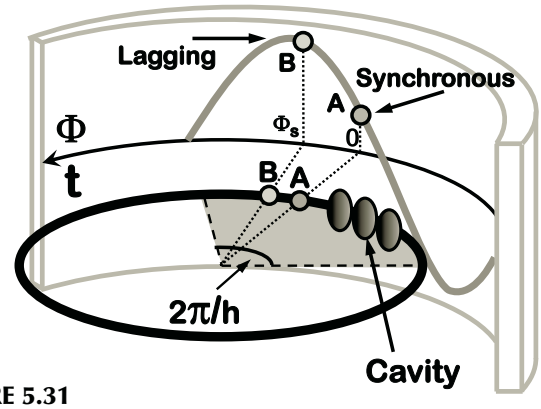



**FIGURE 5.31**
Synchronous and lagging particles in a synchrotron ring.

A particle in an RF cavity changes its energy according to the phase of the RF field in the cavity

$$\Delta E = eV(t) = eV_o \sin(\omega_{RF} t + \varphi_s) \tag{5.36}$$

The *synchronous particle* (see Fig.5.31) is the particle that arrives at the RF cavity when the voltage is such that it exactly compensates the average energy losses $U_0$

$$\Delta E = U_0 = eV_0 \sin(\varphi_s) \tag{5.37}$$

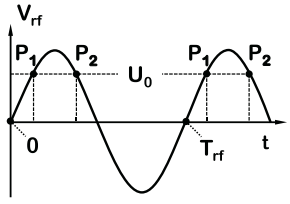



**FIGURE 5.32**
Motion in RF potential.

There are two points in time in the RF potential where the particle will get the correct energy from the RF wave: point $P_1$ and point $P_2$ (see Fig.5.32), and, as we can guess, one is stable and the other is unstable.

In Section 5.1.6 we began our discussion of the dynamics of the particle arriving earlier or later than the synchronous particle, making a simplifying assumption that the particle is ultrarelativistic.

In the case of arbitrary energy, we need to take into account the dependence of the particle's time of flight around the ring of energy, which includes dependence of the circumference and particle velocity on energy.

The synchronous particle with a nominal energy $E$ and velocity $v$ travels around the nominal circumference $C$ in time $T$ so that $T = C/v$. Taking a logarithm and differentiating this expression will yield

$$\frac{dT}{T} = \frac{dC}{C} - \frac{dv}{v} \tag{5.38}$$

The first component of the above equation, $dC/C$, is expressed via the *momentum compaction factor* $\alpha_c$, which connects the momentum deviation of the particle $dp$ with the orbit length difference $dC$:

$$\frac{dC}{C} = \alpha_c \frac{dp}{p} \tag{5.39}$$

The momentum compaction factor $\alpha_c$ depends on the design of the focusing lattice of the ring and can be either positive (which is most common) or negative (particles with higher $E$ travel over a shorter orbit — which seems counterintuitive, but is possible, although rare).

The second component of Eq.5.38 can be expressed as

$$\frac{dv}{v} = \frac{d\beta}{\beta} = \frac{1}{\gamma^2}\frac{dp}{p} \tag{5.40}$$

Taking these two components together will yield

$$\frac{dT}{T} = \left(\alpha_c - \frac{1}{\gamma^2}\right)\frac{dp}{p} \sim \left(\alpha_c - \frac{1}{\gamma^2}\right)\frac{dE}{E} \tag{5.41}$$

which shows us that the time of flight depends on the energy deviation, on the relativistic factor $\gamma$ and on the momentum compaction factor.

We now can conclude that if $\alpha_c - 1/\gamma^2 > 0$, the point $P_2$ in Fig.5.32 is stable (in contrast to phase stability as described for a linac earlier). Indeed, according to this assumption, a particle with a higher energy has a longer flight time and therefore arrives later at the RF cavity, undergoes lower RF voltage (point $P_2$ is on a negative slope), thus gaining less energy, and so tends to return to a nominal energy. Similar deliberations can show that a particle with a lower energy will, in this case, gain more energy in a manifestation of the "principle of phase stability," which enables the capture of particles in the RF potential.

In the opposite case, $\alpha_c - 1/\gamma^2 < 0$, similar logic can lead us to a conclusion of stability of the point $P_1$ (as for a linac).

We have now arrived at the need to introduce the notion of *transition energy*. In rings with positive $\alpha_c$ passing, during acceleration, the energy corresponding to the gamma factor

$$\gamma_t = \frac{1}{\alpha_c^{1/2}} \tag{5.42}$$

corresponds to the moment of stability flipping from point $P_1$ to point $P_2$ on the RF slope. Preserving the quality of the accelerated beam requires a quick switch of the phase of the RF voltage at the moment of passing the transition energy.

Now, let's derive the longitudinal beam dynamics equations for particle acceleration in a synchrotron. We start by considering a particle moving in an electric field of a travelling wave $E_z = E_0\cos(\omega t - kz)$ with a phase velocity $v_f = \omega/k$. Equations describing the motion in the longitudinal plane are

$$\frac{dp_z}{dt} = eE_0\cos(\omega t - kz) \quad \text{and} \quad \frac{d\mathcal{E}}{dt} = eE_0\dot{z}\cos(\omega t - kz) \tag{5.43}$$

We again define the synchronous particle as

$$\frac{dE_s}{dt} = eE_0 v_s \cos\phi_s \tag{5.44}$$

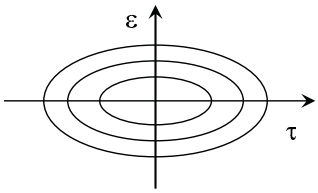



**FIGURE 5.33**
RF bucket trajectories in a linearized case are ellipses.

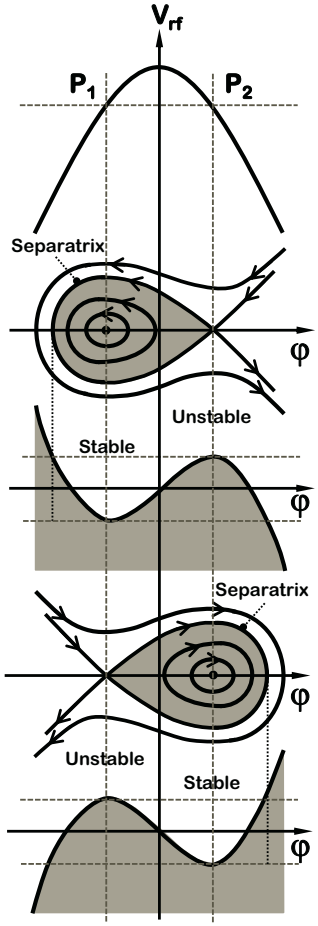



**FIGURE 5.34**
RF voltage and phase space and RF potential for cases below and above the transition energy.

and use deviations from its energy and time to describe an arbitrary particle

$$\mathcal{E} = E_s + \varepsilon \quad \text{and} \quad t = t_s + \tau \tag{5.45}$$

Using these definitions, we obtain the first equation

$$\frac{d\varepsilon}{ds} = eE_0\left[\cos(\omega\tau + \phi_s) - \cos\phi_s\right] \tag{5.46}$$

and using Eq.5.41 to define the momentum compaction factor at high energy ($\gamma \gg 1$), we obtain the second equation

$$\frac{dT}{T} \sim \left(\alpha_c - \frac{1}{\gamma^2}\right)\frac{d\mathcal{E}}{\mathcal{E}} \quad \rightarrow \quad \frac{d\tau}{dt} \sim \frac{\alpha_c}{E_s}\frac{d\varepsilon}{dt} \tag{5.47}$$

which together describe an *RF bucket* in the longitudinal phase space with coordinates $(\tau, \varepsilon)$.

Rewriting these equations for the RF bucket for $\gamma \gg 1$ in terms of derivatives of time yields

$$\varepsilon' = \frac{qV_0}{L}\left[\sin(\varphi_s + \omega\tau) - \sin\varphi_s\right] \quad \text{and} \quad \tau' = \frac{\alpha_c}{E_s}\varepsilon \tag{5.48}$$

Linearizing these equations for the motion in the RF bucket gives us

$$\varepsilon' = \frac{e}{T_0}\frac{dV}{d\tau}\tau \quad \text{and} \quad \tau' = \frac{\alpha_c}{E_s}\varepsilon \tag{5.49}$$

which corresponds to the phase space motion with elliptical trajectories as in Fig.5.33, with angular frequency defined as

$$\omega_s^2 = \frac{\alpha_c e\dot{V}}{T_0 E_0} \tag{5.50}$$

which is called *synchrotron frequency*.

In a realistic case of a practical accelerator design, we often cannot limit ourselves to a linear approximation and would need to consider the full nonlinear equations Eqs.5.48. We will explore these equations now, looking at them from a different perspective — via an analogy with classical mechanics. The equivalent equation can be rewritten as

$$\frac{d^2\varepsilon}{dt^2} = U\tau^2 \tag{5.51}$$

where $U$ is an analog of a potential energy in an oscillator.

The shape of the potential and corresponding phase space trajectories are shown in Fig.5.34 for two cases, below and above the transition energy. The potential energy analogy and its shape help to make sense of the behavior of the trajectories. They are ellipse-like in the vicinity of the stable point where the potential is parabolic, but become distorted and eventually unstable as they come closer to the *saddle point*

of the potential where $U$ flattens and passes through a local maximum, creating conditions for certain trajectories to take an arbitrarily long time for passing through the saddle point region. The trajectory that originates from the vicinity of the saddle point and separates the stable area from an unstable area is called the *separatrix*.

### 5.5.4 RF potential — nonlinearity and adiabaticity

The RF potential, as we can see in Eqs.5.51, is intrinsically nonlinear. Particles that are located close to the separatrix will have longer periods of oscillations and will therefore lag behind in their rotation in phase space with respect to particles in the center. This is illustrated in Fig.5.35, which shows the qualitative behavior of a particle bunch in an RF potential (neglecting radiation damping) with an increasing number of synchrotron periods. We can see that while the beam distortions are negligible after one synchrotron period, after ten and especially after fifty synchrotron periods the longitudinal phase space distribution is completely distorted. The nonlinearity of the RF bucket results in *filamentation* of the phase space, causing an effective increase of the longitudinal emittance.

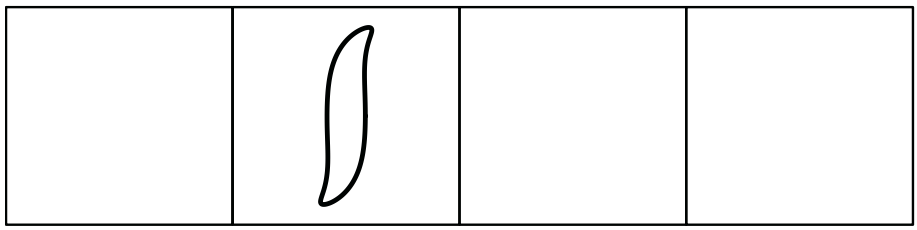


**FIGURE 5.35**
Qualitative evolution of the longitudinal phase space (energy vs phase, for vertical and horizontal axes, correspondingly) of the beam for an increasing number of synchrotron periods. Left to right: initial distribution, after one synchrotron period, after ten and after fifty periods.

Another assumption we made in the previous sections is that the acceleration is adiabatic, i.e., $d\gamma_s/ds \sim 0$. If this assumption is violated, numerical integration of the equations of motion need to be used, which will then show that the RF bucket distorts further and takes the shape of a "golf club" as illustrated qualitatively in Fig.5.36.

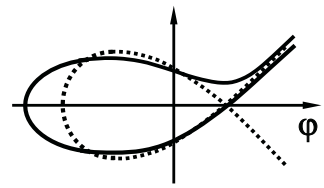



**FIGURE 5.36**
RF bucket in the case of fast acceleration.

### 5.5.5 Synchrotron tune and betatron tune

Synchrotron oscillation in circular accelerators, as we saw in the previous sections, is a multi-turn phenomenon (the synchrotron frequency may correspond to many tens or hun-

dreds of turns), i.e., such oscillations are very slow. The *synchrotron tune* is connected to the synchrotron frequency as follows

$$Q_S = \frac{2\pi T_{rev}}{\omega_S} \tag{5.52}$$

where $T_{rev}$ is the period of revolution around the orbit. Typically, $Q_S \ll 1$.

In contrast, transverse betatron oscillations, in circular accelerators with strong focusing, are necessarily fast (there are usually many tens or hundreds of oscillations per revolution period). The *betatron tune* is defined as the number of oscillations around the ring or the ratio of the betatron frequency to the revolution frequency

$$Q = \frac{\mu}{2\pi} = \frac{1}{2\pi} \quad \frac{ds'}{\beta(s')}$$

where the integral is taken along the accelerator circumference. Typically, $Q \gg 1$.

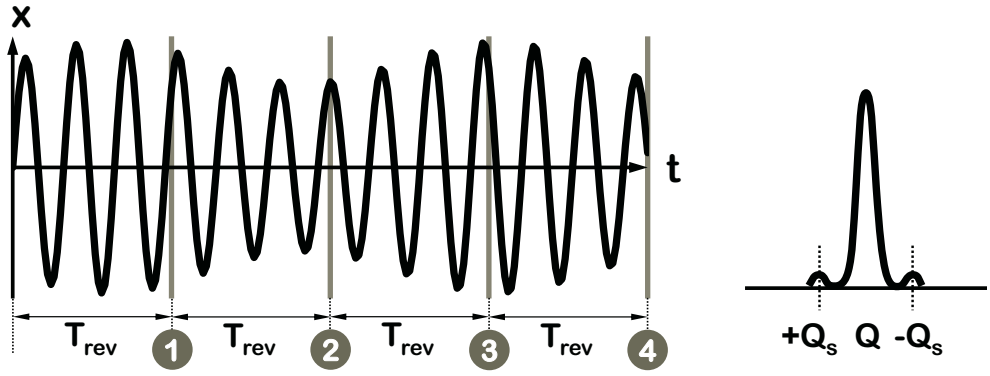



**FIGURE 5.37**
Betatron oscillations modulated by synchrotron motion (left) and a corresponding spectrum (right) with betatron tune and synchrotron sidebands.

The betatron $Q$ is momentum dependent and this can link the two motions together. When synchrotron motion is thus coupled to betatron motion, this can manifest itself in the signals of *pick-up electrodes*, which measure the beam transverse oscillations as qualitatively shown in Fig.5.37. As the main fast (betatron) signal is now modulated with a slow (synchrotron) component

$$x \propto \sin(2\pi Q\, t/T_{rev})(1 + \Delta \sin(2\pi Q_S\, t/T_{rev}))$$

the spectrum of transverse motion will now include *synchrotron sidebands* at $Q - Q_S$ and $Q + Q_S$ in addition to the main betatron frequency as is illustrated in Fig.5.37.

### 5.5.6 Accelerator technologies and applications

Table 5.1 shows typical accelerating gradients for RF cavities of various frequencies for linear accelerators.

**TABLE 5.1**
Operating frequencies and typical parameters for RF cavities

| **Warm cavities** | **Gradient** | **Repetition rate** |
|---|---|---|
| S-band (3GHz) | 15-25 MV/m | 50-300 Hz |
| C-band (5-6 GHz) | 30-40 MV/m | <100 Hz |
| X-band (12 GHz) | 100 MV/m | <100 Hz |
| **Superconducting cavities** | **Gradient** | **Repetition rate** |
| L-band (1.3 GHz) | < 35 MV/m | up to CW |

As we can see, the achievable gradient generally increases with frequency, consistent with frequency dependences of Eq.5.25 and Eq.5.26 (these relationships, however, haven't yet been demonstrated for frequencies higher than 12 GHz).

The practically achievable gradient is one of the main factors that define the size of accelerator-based facilities of various kinds. Let's consider a couple of examples.

In a *high energy physics* application, a *linear collider* (Fig.5.38) aiming at 500 GeV energy in the center of mass (CM), and built with L-band superconducting cavities, would be around 30 km long. A collider aiming at 3 TeV CM, built with an X-band normal conductive cavities, would be almost 50 km long.

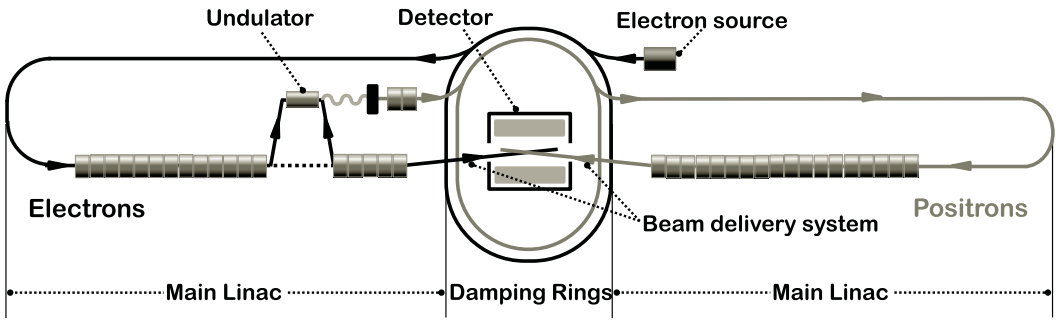



**FIGURE 5.38**
A generic linear collider.

While the linac length constitutes a major fraction of the length of a linear collider, in a *free electron laser*, the linac length is a noticeable fraction of the overall length (between around a quarter to a half). As an example, an FEL (see Fig.5.39) with an electron beam energy of around 10 GeV built with S-band or C-band technology would be about a kilometer long.

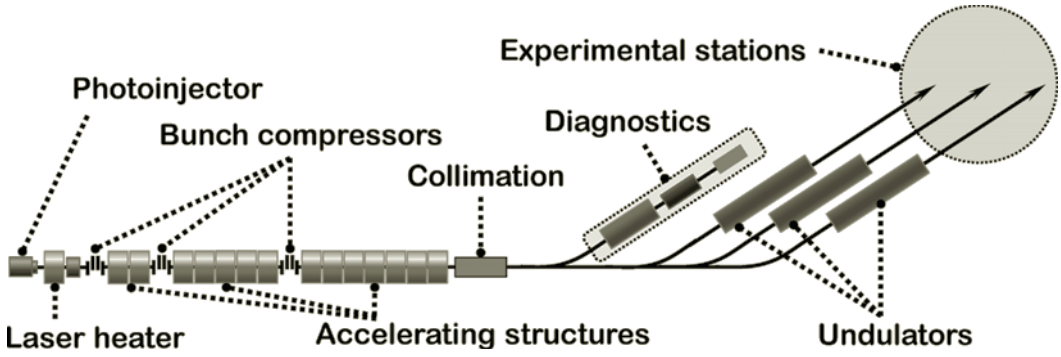



**FIGURE 5.39**
A generic free electron laser.

These examples demonstrate the motivations for developing new accelerator technologies that would make these and other accelerator-based facilities and instruments more compact. This brings us to the next chapter, where we will delve into methods of plasma acceleration.

**EXERCISES**

5.1 *Chapter materials review*.
Derive the Eq.5.35 for particle motion in the travelling wave.

5.2 *Chapter materials review*.
Discuss design approaches to the beam optics that would result in achieving a negative value of the momentum compaction factor in a synchrotron.

5.3 *Mini-project*.
Define very approximate parameters (sizes, magnetic fields, parameters of RF system) of a 200 MeV rapid-cycling proton synchrotron capable of operating at a 10 Hz repetition rate. Assume injection at 1 MeV.

5.4 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this chapter (e.g., tandem, RFQ, bunch and pulse compressors) in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

5.5 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 6 Plasma Acceleration



Plasma acceleration is an emerging and promising field, whose rapid progress is enabled by developments in laser technology — particularly by the method of chirped pulse amplification. Plasma accelerators of electrons — the primary focus of this chapter — are the backbones of future compact light sources. Proton and ion plasma acceleration, briefly discussed here and in closer detail in Chapter 9, are a potential way to improve future medical accelerators.

The aim of this chapter, after a brief introduction and discussion of the motivations for pursuit of plasma acceleration, is to develop the framework that will help us to estimate the parameters of laser plasma-based light sources, so as to be prepared for Chapter 7.

## 6.1 Motivations

The "Livingston plot," which depicts the energy of accelerated beams versus time (Fig. 1.6), illustrates the great history of accelerators and related inventions. It also shows the signs of saturation, highlighting the need for the next breakthrough in accelerator technology.

Traditionally, accelerating structures have been made from metal (normal conductive or super-conductive) and are typically limited in their accelerating gradient to $E_z <$ 100 MeV/m. This limitation is imposed by the properties of the materials — since damage to the accelerating structure's walls (deterioration of their integrity) limits the gradient.

The "accelerating structures" produced on the fly in plasma by a laser pulse are, however, made from a material that is already "damaged" (plasma), and therefore do not exhibit the same limitations due to the material's properties.

Plasma acceleration was first proposed by T. Tajima and J. Dawson in 1979, which was, in fact, too early for laser and beam technologies to be ready to realize the proposed approach. Consequent parallel developments of laser and beam technologies — specifically those aimed at creating short, powerful pulses — created the new reality making the laser plasma acceleration the area with the highest degree of synergy between the physics of plasma, lasers and accelerators.

We should also recall (see margin notes) our discussion

*Recalling the use of Accelerating Science TRIZ method:*
*1) Define the problem in terms of generic contradiction parameters (Table 1.4). To be improved: rate of E change; what gets worse: integrity.*
*2) Use the contradiction matrix (Table 1.5) to obtain the relevant inventive principle — replace material that can be damaged with other media, which either cannot be damaged (light) or is already "damaged" (e.g., plasma).*
*3) Translate the generic inventive principle into a specific solution — plasma acceleration.*

DOI: 10.1201/b18696-6

from Chapter 1 regarding the use of the AS-TRIZ approach to *post-facto* analyze the invention of plasma acceleration.

### 6.1.1 Maximum field in plasma

By using plasma as an accelerating medium, we can remove the limitation of the accelerating gradient relating to the material's damage threshold. The maximum field in plasma will still be limited, but by other factors.

Let us look again at plasma oscillation, as illustrated in Fig. 4.21, and briefly recall that this diagram allowed us to estimate the plasma frequency $\omega_p$. Assuming that a fraction of the charges is shifted by distance $x$, we selected an integral contour that enclosed the displaced charges, and we then equated, according to Maxwell's equations, the surface integral of electric field — $\mathbf{E} \cdot d\mathbf{S}$ to the volume integral of charge density — $\rho dV/\varepsilon_0$, to obtain the electric field created by the shifted charges $E = nex/\varepsilon_0$, which creates the restoring force. The equation of motion $F = md^2x/dt^2 = -eE = -ne^2x/\varepsilon_0$ then gave us the oscillation frequency $\omega_p^2 = ne^2/(\varepsilon_0 m)$, which, with use of $4\pi\varepsilon_0 r_e = e^2/(m_e c^2)$, we rewrote as: $\omega_p^2 = 4\pi n c^2 r_e$ — the angular plasma frequency.

*Plasma frequency in Hz is $f_p \sim 9000\, n^{1/2}$ where plasma density n is in $cm^{-3}$.*

Very similar calculations allow us to estimate the maximum accelerating field in plasma. Imagine that the plasma oscillation in Fig. 4.21 is excited by a charged object moving with velocity $c$. In the case where a total charge separation is achieved in plasma, the maximal field is estimated assuming

$$x \sim \lambda_p \sim \frac{c}{\omega_p} \tag{6.1}$$

which results in the following for the maximum field

$$E_{\max} \sim \frac{nec}{\varepsilon_0 \omega_p} = \frac{mc\omega_p}{e} \tag{6.2}$$

or equivalently

$$eE_{\max} \cong mc^2 \, \frac{\omega_p}{c} \tag{6.3}$$

*Maximum accelerating field in plasma* $eE_{max} \approx 1\,\mathrm{GeV/cm} \cdot \left(n/10^{18}\mathrm{cm}^{-3}\right)^{1/2}$.

We can use the practical formula $f_p \sim 9000\, n^{1/2}$ where $n$ is defined in $\mathrm{cm}^{-3}$ to obtain a formula for the maximum possible accelerating field in plasma:

$$eE_{max} \approx 1 \frac{\mathrm{eV}}{\mathrm{cm}} \cdot n^{1/2} \ \mathrm{cm}^{-3} \tag{6.4}$$

This means that 1 GeV/cm accelerating gradient can be achieved for plasma of $10^{18} cm^{-3}$ density.

Theoretical predictions, made back in 1979, of the principal feasibility of such large accelerating gradients were an essential driving force towards the development of plasma acceleration technology.

## 6.2 Early steps of plasma acceleration

We see that GeV/cm requires plasma with n=$10^{18} cm^{-3}$. The plasma wavelength

$$\lambda_p = \frac{c}{f_p} \quad \text{or} \quad \lambda_p \approx 0.1 mm \quad \overline{\frac{10^{17} cm^{-3}}{n}} \tag{6.5}$$

corresponding to a plasma density of $10^{18} cm^{-3}$ is around $\lambda_p \approx 30\,\mu m$ (or around 100 fs). Thus, short sub-100-fs pulses are needed to excite plasma towards GeV/cm accelerating gradients.

In the absence of such short laser pulses, in the late 1970s and early 1980s, other methods of plasma excitation were suggested (by J.M. Dawson, 1979) such as the plasma beat wave accelerator (PBWA) and the self-modulated laser wakefield accelerator (SMLWFA); see Fig. 6.1.

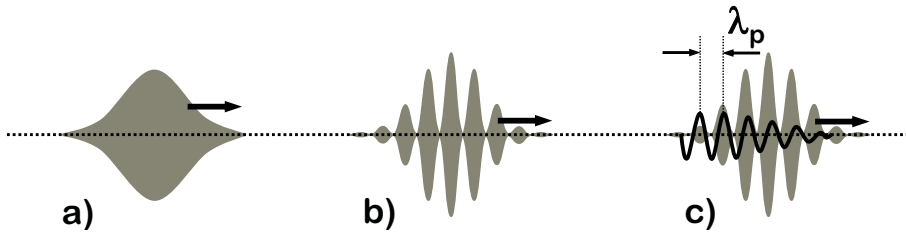



**FIGURE 6.1**
For illustration of plasma beat wave and self-modulated laser wakefield acceleration.

In the PBWA, two laser pulses with envelopes as in Fig. 6.1.a and frequencies differing by $\omega_p$ overlap to create a beating at the plasma's frequency as shown in Fig. 6.1.b. This combined laser pulse is sent into plasma where it creates plasma excitation as shown in Fig. 6.1.c.

In contrast to the previous method, in the SMLWFA, only a single laser pulse is sent into the plasma (Fig. 6.1.a), where an instability (which we will not discuss here in detail) results in a self-modulation of the long laser pulse at $\lambda_p$ (Fig. 6.1.b), which again creates plasma excitation at wavelength $\lambda_p$ (Fig. 6.1.c).

As a result of beam and laser technologies development, short sub-ps pulses of laser or beams became available and thus prompted rapid progress of plasma acceleration.

The plasma wakefield acceleration (PWFA) method uses a short, high energy particle bunch to excite the plasma (Fig. 6.2). Similarly, a short laser pulse of high intensity can be used in a laser wakefield acceleration (LWFA) method (Fig. 6.3). In both of these cases, a high amplitude plasma wave is created, which can then be used for acceleration. We will discuss these methods in detail in the following sections.

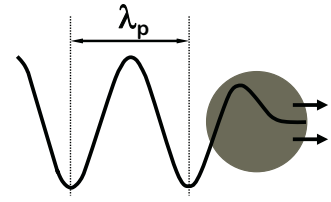



**FIGURE 6.2**
Plasma wakefield acceleration — PWFA.

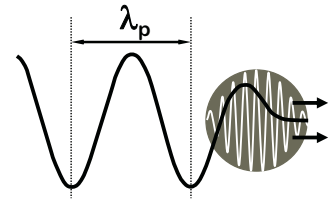



**FIGURE 6.3**
Laser wakefield acceleration — LWFA.

## 6.3 Laser intensity and ionization

Laser acceleration requires laser pulses of short duration and high intensity. In order to prepare for a quantitative discussion of the intensities required for plasma acceleration, let us introduce the basic concepts related to this subject.

### 6.3.1 Laser pulse intensity

Laser intensity (in a vacuum) is defined (in SI and Gaussian units respectively) as

$$I = \frac{1}{2}\varepsilon_0 E_{\max}^2 c \quad \text{(SI)} \tag{6.6}$$

*Recall*
$\varepsilon_0 \approx 8.8 \cdot 10^{-12} A^2 s^4/(kg\, m^3)$

$$I = \frac{1}{8\pi} E_{\max}^2 c \quad \text{(Gaussian)} \tag{6.7}$$

The intensity $I$ is usually measured in Watts per $cm^2$. The corresponding relation between electric field and intensity in practical units is:

*It is useful to remember that in an EM wave the field amplitudes of 300 V/cm and ≈1 Gauss are equivalent.*

$$E_{\max}\left[\frac{V}{cm}\right] \quad 2.75 \times 10^9 \quad \frac{I}{10^{16} W/cm^2}^{1/2} \tag{6.8}$$

Similarly, for the magnetic field:

$$B_{\max}[Gauss] \quad 9.2 \times 10^6 \quad \frac{I}{10^{16} W/cm^2}^{1/2} \tag{6.9}$$

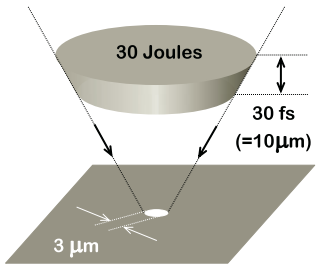



**FIGURE 6.4**
Laser focused to a tight spot.

For example: a laser with 30 Joules energy in a 30-fs (10 $\mu$m)-long pulse corresponds to (assume rectangular distribution in space and time) a peak power of $10^{15}$ watts and, if focused to a spot with a diameter of 3 $\mu$m as illustrated in Fig. 6.4, produces the intensity in the focus of $10^{22}$ W/cm$^2$. According to the equations given above, the fields in the focus of such a laser will approach 10,000 Mega Gauss.

### 6.3.2 Atomic intensity

In order to develop a quantitative understanding of laser intensity values, it is best to compare the field of an intense laser with *atomic fields* — particularly with the field in a hydrogen atom.

The Bohr radius is given by:

$$a_B = \frac{\hbar^2}{me^2} = 5.3 \times 10^{-9} cm \tag{6.10}$$

The corresponding field is then defined as:

$$E_a = \frac{e}{a_B^2} \quad \text{(Gaussian units)} \tag{6.11}$$

$$= \frac{e}{4\pi\varepsilon_0 a_B^2} \quad \approx 5.1 \times 10^{11} \frac{V}{m} \quad \text{(SI)}$$

The corresponding atomic intensity is thus equal to

$$I_a = \frac{\varepsilon_0 c E_a^2}{2} \quad 3.51 \times 10^{16} \frac{W}{cm^2} \tag{6.12}$$

A laser with intensity higher than the above will ionize gas immediately. However, as we will show in the next sections, ionization can occur well below this threshold due to multi-photon effects or tunneling ionization.

### 6.3.3 Progress in laser peak intensity

Lasers able to produce peak intensity of atomic levels given by Eq. 6.12 were not available until the mid-1980s. This is illustrated in Fig. 6.5, which shows a qualitative overview of the progress in laser peak intensity throughout history.

The invention of the chirped pulse techniques — CPA and OPCPA — was a breakthrough in laser peak power, allowing reaching and exceeding atomic intensities (indicated by the line **b** in Fig. 6.5).

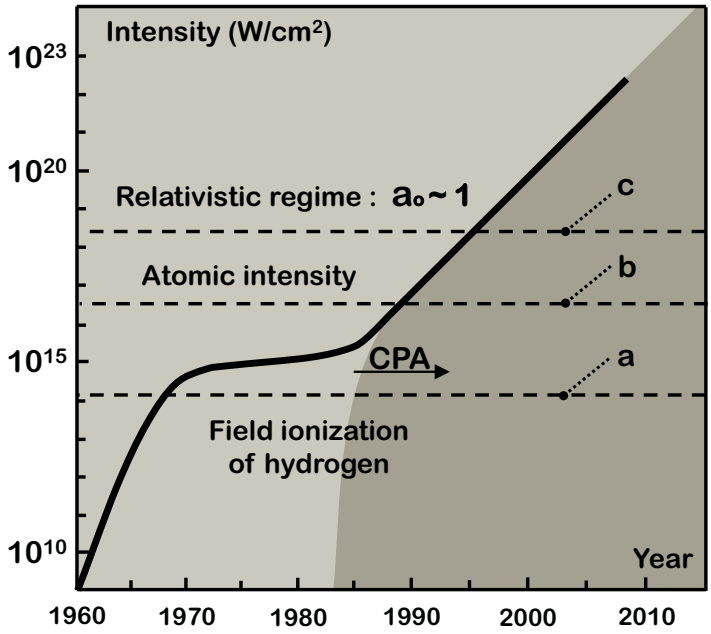



**FIGURE 6.5**
Qualitative overview of the progress in laser peak intensity.

Fig. 6.5 also shows two other important intensity levels: the first one corresponds to the field ionization of hydrogen (line **a** in this figure), discussed in the next section, and the second one corresponds to the relativistic optics case (line **c**) when electrons become relativistic in the laser field (discussed in Section 6.3.6).

One more vital intensity limit relevant to Fig. 6.5 — but not shown as it would be significantly off-scale (around $2 \cdot 10^{29}$ W/cm$^2$) — is the *Schwinger intensity limit*, which corresponds to the case when the laser field can produce $e^+e^-$ pairs from a vacuum (this will be discussed further in Section 6.3.8).

### 6.3.4 Types of ionization

There are several types of ionization of interest, some of which are shown in Fig. 6.6, depicting the potential well of an electron in an atom.

In *direct ionization* (Fig. 6.6.a), the photon transmits enough energy to an electron to overcome the potential barrier in one interaction. In a *multi-photon ionization* (Fig. 6.6.b), the electron obtains the energy needed to overcome the potential barrier via the process of multi-photon absorption. *Tunneling ionization* (Fig. 6.6.c) can occur when an electron quantum tunnels through the potential barrier.

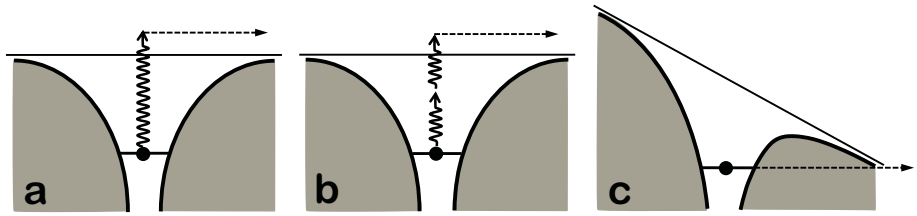



**FIGURE 6.6**
Types of ionization: (a) direct, (b) multi-photon, (c) tunneling.

The tunneling ionization, as presented in Fig. 6.6.c, and with somewhat larger laser field intensity, will turn into another ionization mechanism called barrier suppression ionization.

### 6.3.5 Barrier suppression ionization

*Barrier suppression ionization* (BSI) occurs when the laser field distorts the potential of an atom in such a way that the electron can freely escape the potential well.

The Coulomb potential of a hydrogen atom distorted by a homogeneous field $E$ can be written as (in Gaussian units):

$$V(x) = -\frac{e^2}{x} - eEx \tag{6.13}$$

The distorted potential is shown in Fig. 6.7. The position of the maximum of the potential on the right side of the plot is:

$$x_{\max} = (e/E)^{1/2} \tag{6.14}$$

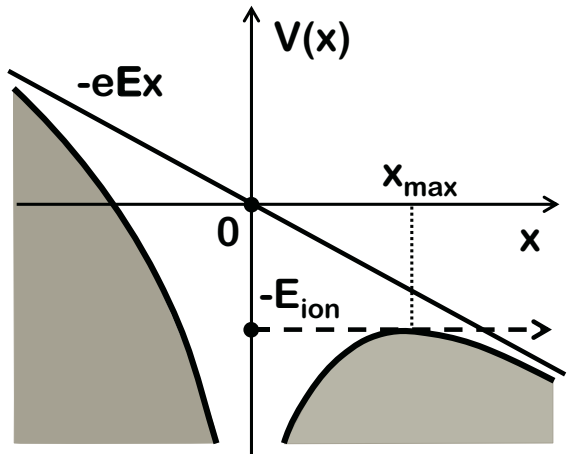



**FIGURE 6.7**
Barrier suppression ionization.

and the value of the potential at the maximum is

$$V(x_{\max}) = 2\ e^3 E^{\ 1/2} \tag{6.15}$$

Equating the potential value at the maximum $V(x_{\max})$ to the hydrogen atom ionization potential $E_{ion}$

$$E_{ion} = \frac{e^2}{2a_B} \approx 13.6eV \tag{6.16}$$

gives us the critical field for the hydrogen atom

$$\varepsilon_c = \frac{e}{16a_B^2} = \frac{E_a}{16} \tag{6.17}$$

As we can see, it is a small fraction of the atomic field $E_a$ given by Eq. 6.11. The laser intensity corresponding to the BSI mechanism is then

$$I_c = \frac{I_a}{256} \approx 1.4 \cdot 10^{14} W/cm^2 \tag{6.18}$$

which is more than two orders of magnitude lower than the atomic intensity $I_a$ given by Eq. 6.12. The intensity $I_c$ is indicated by line **a** in Fig. 6.5.

### 6.3.6 Normalized vector potential

The laser field can be written in terms of the vector potential of the laser field **A** as

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{c\partial t}\ , \qquad \mathbf{B} = \nabla \times \mathbf{A} \tag{6.19}$$

For a linearly polarized field:

$$\mathbf{A} = A_0 \cos(kz - \omega t)\,\mathbf{e}_\perp \tag{6.20}$$

where $\mathbf{e}_\perp$ is transverse unit vector. We can see that the field and vector potential amplitudes are connected via

$$E_0 = \frac{A_0 \omega}{c} \tag{6.21}$$

Comparing momentum gained by an electron $e^-$ in one cycle of laser field

$$e\,E\,\Delta t \sim \frac{eE}{\omega} \tag{6.22}$$

with its rest mass $m_e c$, we can see that it is better to define the *normalized vector potential* as

$$\mathbf{a} = \frac{e\mathbf{A}}{m_e c^2} \tag{6.23}$$

with its amplitude given by

$$a_0 = \frac{eE_0}{m_e \omega c} \tag{6.24}$$

The amplitude $a_0$ will indicate if the electron motion in the laser field is relativistic: $a_0 \gg 1$, or nonrelativistic: $a_0 \ll 1$.

The normalized vector potential amplitude in practical units can be written as

$$a_0 \approx \left( \frac{I\left[W/cm^2\right]}{1.37 \cdot 10^{18}} \right)^{\frac{1}{2}} \cdot \lambda\,[\mu m] \tag{6.25}$$

where $\lambda = 2\pi c/\omega$ is the wavelength of the laser. For example, for a red laser with $\lambda = 0.65$ $\mu$m, the value $a_0 = 1$ reached at intensity of $I \approx 3 \cdot 10^{18}$ $W/cm^2$ (as indicated by line **c** in Fig. 6.5).

### 6.3.7 Laser contrast ratio

As we see in Fig. 6.5, different phenomena related to laser-matter interaction and plasma acceleration occur at significantly different intensities. This brings us to a dialogue regarding the temporal *contrast ratio* of a laser pulse.

The spatial contrast — the ratio of intensity at the laser focus to the intensity outside of the focus — is a standard concept intuitively known to everyone from everyday life.

For CPA-compressed pulses, which involve manipulations and exchanges between energy and longitudinal phase space coordinates, it is appropriate to introduce the notion of the temporal *contrast ratio* — a function of time given by the ratio of the peak laser intensity to the intensity in the front or back of the pulse.

A qualitative spatial profile of a CPA-compressed laser pulse is shown in Fig. 6.8 in terms of the contrast ratio, in logarithmic scale. It is typical that a short sub-ps pulse is accompanied by many tens of ps low-intensity pulses, as well

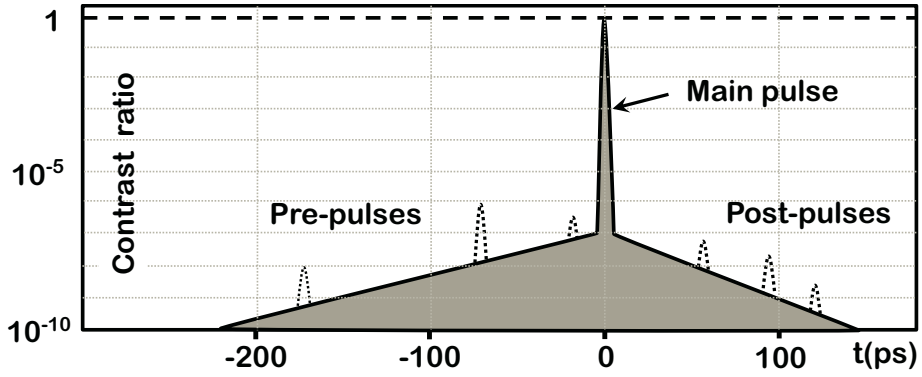



**FIGURE 6.8**
Qualitative temporal profile of a CPA-compressed laser pulse.

as short pre-pulses or post-pulses, which are typically caused by nonlinear properties of the elements of the CPA system and non-ideal properties of the initial laser pulse.

The high contrast ratio is often the key parameter in plasma acceleration, as even a relatively low intensity can either ionize or destroy the target long before the arrival of the main high-intensity short pulse.

### 6.3.8 Schwinger intensity limit

A laser of high enough intensity can generate $e^+e^-$ pairs from a vacuum. According to the time-energy uncertainty principle

$$\Delta E\, \Delta t \geq \frac{\hbar}{2} \tag{6.26}$$

a virtual $e^+e^-$ pair (thus $\Delta E \sim m_e c^2$) can appear for a short duration of time $\Delta t \sim \hbar/(m_e c^2)$. If an electric field $E_S$ acting on the pair during $\Delta t$ increases the momentum of $e^+$ or $e^-$ by about $mc$ (i.e. $\Delta t \cdot e E_S \sim mc$), the particles then become real and thus materialize from the vacuum thanks to the laser field. The corresponding laser field (ignoring factors of two in the estimations) is hence given by

$$E_S = \frac{m_e^2 c^3}{e\hbar} \approx 1.3 \cdot 10^{18} \ \ \text{V/m} \tag{6.27}$$

and is called a *Schwinger limit* field — the scale above which the linear electrodynamics become invalid. The corresponding laser intensity is

$$I_S \approx 2 \cdot 10^{29} \ \ \text{W/cm}^2 \tag{6.28}$$

Reaching such laser intensity in practice would undoubtedly create a new scientific and technological breakthrough.

## 6.4 The concept of laser acceleration

We are now ready to discuss the concept of laser plasma acceleration — see Fig. 6.9 — wherein a powerful laser pulse enters gas (which can either be pre-ionized or not). We first note that the contrast ratio of the laser is not infinite, and so the ionization front starts in the gas at the front tail of the laser pulse, much in advance of the arrival of the main laser pulse. We then note that the main laser pulse needs to be of a length similar to or shorter than the plasma wavelength in order to excite the plasma efficiently.

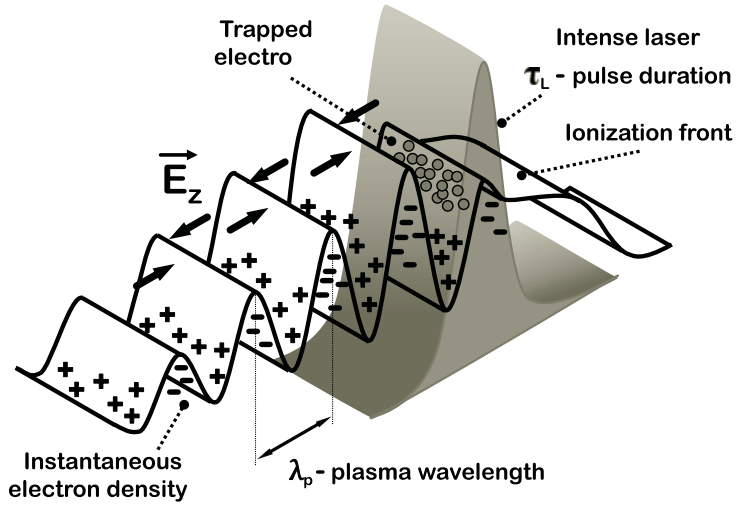



**FIGURE 6.9**
Laser acceleration — conceptually. Linear regime.

The electrons of the plasma can be trapped in the wave and then accelerated. Maximum acceleration can occur when the laser pulse causes total separation of the electrons and ion charges of the plasma; this regime is nonlinear, and the cavity that is formed in the plasma and can trap and accelerate electrons is called a *bubble*. Usually, electrons are trapped and accelerated in the first bubble. The mechanism of bubble formation in a strongly nonlinear approximation will be discussed in detail in the following section.

### 6.4.1 Ponderomotive force

The formation of a bubble is the result of *ponderomotive force* that a laser pulse confined in space exerts on the plasma electrons.

We start from the assumption that the laser field $E$ is homogeneous:

$$E = E_0 \cos(\omega t) \tag{6.29}$$

The corresponding transverse motion of electrons is:

$$\ddot{y} = \frac{F}{m} = \frac{eE}{m} \quad \Rightarrow \quad y = -\frac{eE_0}{m\omega^2}\cos(\omega t) \tag{6.30}$$

We then assume that the field $E$ has a gradient in transverse direction $y$ and thus can be expressed as:

$$E = E_0(y)\cos(\omega t) \approx E_0\cos(\omega t) + y\frac{\partial E_0}{\partial y}\cos(\omega t) \tag{6.31}$$

We then find the force acting on an electron $e^-$ averaged over time is:

$$\langle F\rangle_t = \left\langle -\frac{eE_0}{m\omega^2}\cos(\omega t)\cdot\frac{\partial E_0}{\partial y}\cos(\omega t)\right\rangle_t \tag{6.32}$$

Replacing $\langle\cos^2\rangle$ with 1/2, we rewrite it as

$$\langle F\rangle_t = -\frac{e^2}{2m\omega^2}E_0\frac{\partial E_0}{\partial y} = -\frac{e^2}{4m\omega^2}\frac{\partial E_0^2}{\partial y} \tag{6.33}$$

We see from this equation that, since the intensity $I \propto E^2$, the ponderomotive force is proportional to the gradient of the laser intensity

$$\langle F\rangle_t \propto -e^2\frac{\partial I}{\partial y} \tag{6.34}$$

and that the direction of the force is such that the ponderomotive force pushes electrons out from the high intensity region.

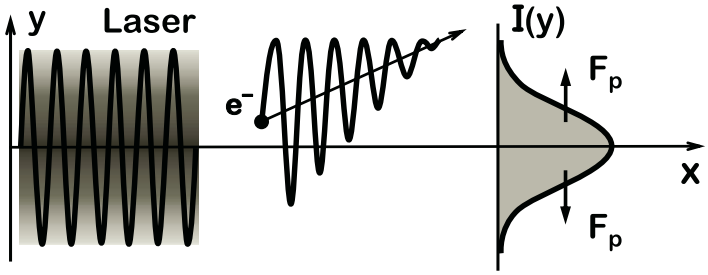



**FIGURE 6.10**
For illustration of the mechanism of the ponderomotive force. Laser with transverse intensity gradient (left), qualitative picture of the electron motion (middle) and the intensity profile with direction of the force shown (right).

We also note that the ponderomotive force is independent on the sign of the charge — positrons will be pushed out of the laser pulse just as well.

An alternative explanation of the ponderomotive force can also be suggested (see Fig. 6.10): the electrons oscillating in the time-varying laser field are pushed away more forcefully when they are in a higher intensity region — which on average repels the charged particles from the laser pulse's high intensity area.

### 6.4.2 Laser plasma acceleration in nonlinear regime

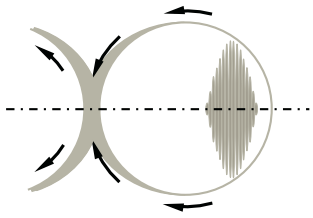


**FIGURE 6.11**
Bubble formation.

Ponderomotive force plays a key role in the formation of the accelerating *bubble* in the nonlinear (also called *blow-out*) regime of laser plasma acceleration.

The ponderomotive force of a short (typically ~50 fs) and intense (typically $\sim 10^{18}$ W/cm$^2$) laser pulse expels plasma electrons while heavier ions stay at rest. The expelled electrons are immediately attracted back to the ions, forming the first bubble, as shown in Fig. 6.11 and Fig. 6.12, thus creating a plasma wave that trails behind the laser pulse. The gradient of the density of electrons creates spatial oscillation of the electric field within plasma (reaching ~100 GV/m) which can accelerate particles.

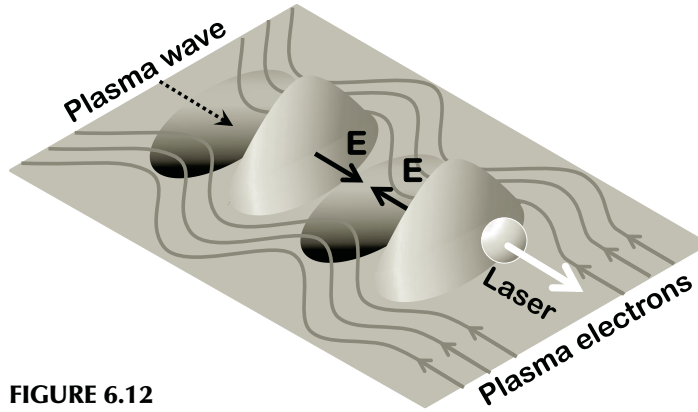



**FIGURE 6.12**
Laser plasma acceleration in nonlinear regime — conceptually.

Having formed the first bubble, the electrons continue their oscillations around the ions, but their motions quickly become incoherent and so the second and subsequent bubbles gradually become smaller. In a sense, only the first bubble (and sometimes the second) is useful for acceleration.

### 6.4.3 Wave breaking

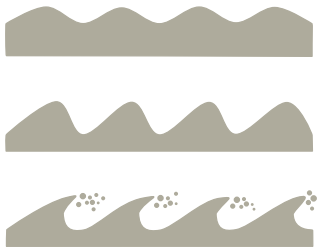


**FIGURE 6.13**
Wave breaking concept — the wave nonlinearity gradually rises from top to bottom.

The high accelerating gradient in the plasma is useful only if a particle beam can be injected into the bubble. Luckily, self-injection of background plasma electrons into the plasma bubble can occur through the *wave breaking* phenomenon.

Wave breaking transpires when, within the nonlinear regime, certain particles outrun the wave, as is represented in Fig. 6.13 in the analogy with ocean waves.

Other methods of getting particles into the bubble include injection of an external electron beam (challenging if the bunches are short) and various other methods that create an electron bunch inside of the bubble in the right place and at the right time. These typically involve using multiple laser pulses and mixes of gases with different ionization potentials.

### 6.4.4 Importance of laser guidance

As a laser pulse travels through the gas or plasma, several competing effects are taking place.

The most notable is *diffraction*; indeed, a laser beam focused to a size of several tens of $\mu$m will diffract very fast. Other effects include *dephasing* — the gradual separation of the accelerating beam (which quickly become relativistic) from the laser (which propagates in gas or plasma slower than the speed of light) — and also *depletion* — the gradual decrease of the laser intensity.

Additional peculiar effects include *longitudinal compression* of the laser pulse by plasma waves; *self-focusing*, in particular due to the relativistic effect (the electrons of plasma at the axis become relativistic and have higher masses, affecting the plasma refraction coefficient); and *ionization-caused diffraction* (gas on the axis where intensity is higher will be ionized first, affecting diffraction).

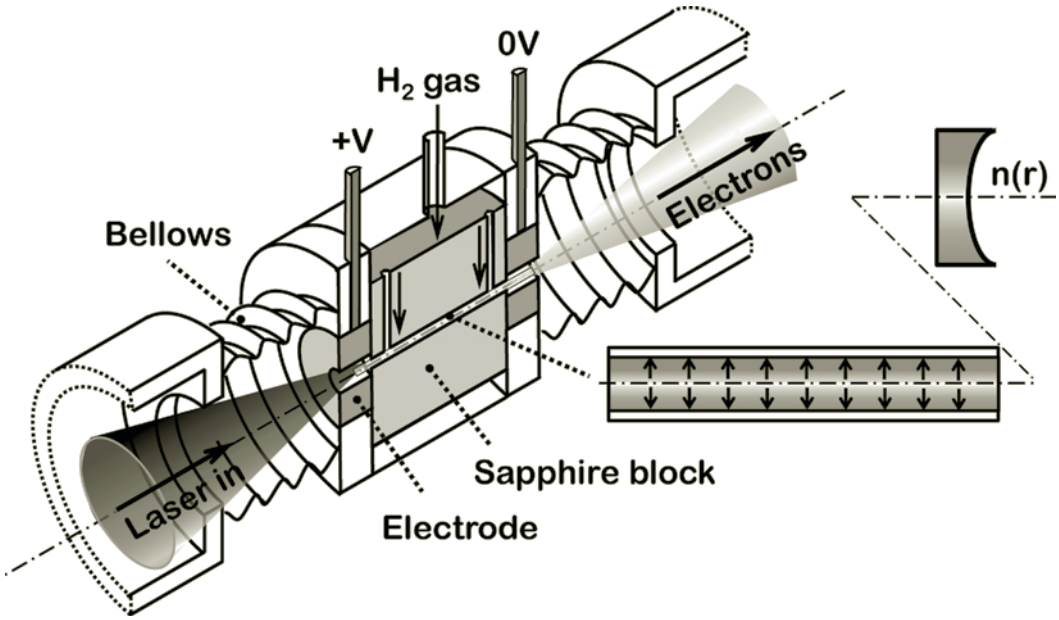



**FIGURE 6.14**
Capillary channel technique of laser plasma acceleration.

A possible solution that could solve some of the issues listed above involves creating a channel within the plasma where a special density profile $n(r)$ will be formed to assist guiding the laser pulse for a significant distance.

This solution was realized in practice in the form of a capillary discharge channel developed at Oxford University by S. Hooker (ca. 2006). A schematic of the capillary channel is shown in Fig. 6.14. In this example, a sub-mm hole is created in a sapphire block, hydrogen gas is delivered to the capil-

lary via side holes, and discharge electrodes act to pre-form plasma with a density profile featuring minimum on the axis (as gas near cold walls of the capillary has higher density according to $P = nkT = conts$) — such a density profile is suitable for refraction-assisted laser guidance.

The capillary channel technique was essential in exceeding the GeV barrier in laser plasma acceleration for the first time ever (W. Leemans et al., 2006), creating a monoenergetic 1 GeV beam after accelerating in just 3 cm of plasma.

## 6.5 Betatron radiation sources

Beams accelerated in a laser-formed plasma bubble can oscillate, which generates synchrotron (betatron) radiation. Strong radial electric fields within plasma bubbles are responsible for the electrons experiencing transverse oscillations. This can generate bright betatron radiation in an extensive range of photon energy (around 1 - 100 keV). In this section we will estimate the expected parameters of radiation produced by a laser plasma source.

### 6.5.1 Transverse fields in the bubble

Transverse oscillations of the accelerating electron beam in the plasma bubble are caused by a transverse focusing force that is produced by ions. We can assume that the ions are heavy and are stationary within the bubble. The ions produce a focusing force that can be determined using

$$\mathbf{E} \cdot d\mathbf{S} = 4\pi \quad \rho dV \quad \text{(Gaussian units)} \tag{6.35}$$

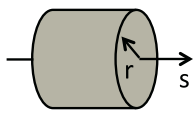



**FIGURE 6.15**
Cylindrical symmetry in the plasma bubble.

and by assuming cylindrical symmetry as in Fig. 6.15. The focusing force is therefore

$$eE = 2\pi ne^2 r \tag{6.36}$$

An electron with relativistic factor $\gamma$ will oscillate in this field as

$$\frac{d^2r}{ds^2} = \frac{2\pi ne^2 r}{\gamma mc^2} = \frac{\omega_p^2}{2\gamma c^2} r \tag{6.37}$$

The period of oscillation is thus given by

$$\lambda = \sqrt{2\gamma}\ \lambda_p \tag{6.38}$$

(and is changing during electron acceleration).

### 6.5.2 Estimations of betatron radiation parameters

In Chapter 3 we estimated the characteristics of synchrotron radiation, in particular the energy loss per unit length:

$$\frac{dW}{ds} = \frac{2}{3}\frac{e^2\gamma^4}{R^2} \tag{6.39}$$

the characteristic frequency of photons:

$$\omega_c = \frac{3}{2}\frac{c\,\gamma^3}{R} \tag{6.40}$$

and the number of photons emitted per unit length:

$$\frac{dN}{ds} = \frac{\alpha\,\gamma}{R} \tag{6.41}$$

Fig. 6.16 shows a qualitative representation of the evolution of the plasma bubble and oscillation of the accelerating beam. We assume that the beams are self-injected into the bubble due to the wave breaking phenomena when two beamlets overshoot and enter the bubble symmetrically from the top and bottom (as shown in Fig. 6.16.a). Their initial transverse velocity forces the beamlets to oscillate in the focusing field of the ions. The beamlets then continue to simultaneously accelerate while exhibiting transverse oscillations.

Let's assume that the amplitude of beam oscillations in the plasma bubble is equal to $r_b$ and the period of oscillations is $\lambda$. The radius of the curvature of the beam trajectory in this case is equal to:

$$R = \frac{\lambda^2}{4\pi^2 r_b} \tag{6.42}$$

Substituting the period of oscillation given by Eq. 6.38, we obtain for the radius of the curvature

$$R = \frac{\gamma\,\lambda_p^2}{2\pi^2 r_b} \tag{6.43}$$

Substituting this into Eq. 6.40, we get an estimation of the radiation wavelength for the laser plasma betatron source:

$$\lambda_c = \frac{1}{3\pi}\frac{\lambda_p^2}{r_\beta}\frac{1}{\gamma^2} \tag{6.44}$$

Using Eq. 6.41 together with Eq. 6.43 we can also estimate the number of photons $N_\gamma$ emitted per $\lambda$:

$$N_\gamma \approx \sqrt{2\gamma}\;\; 2\pi^2\,\alpha\,\frac{r_b}{\lambda_p} \tag{6.45}$$

Let us consider a practical example of a beam accelerated in the bubble characterized by $\lambda_p = 0.03$ mm, to up to 1 GeV

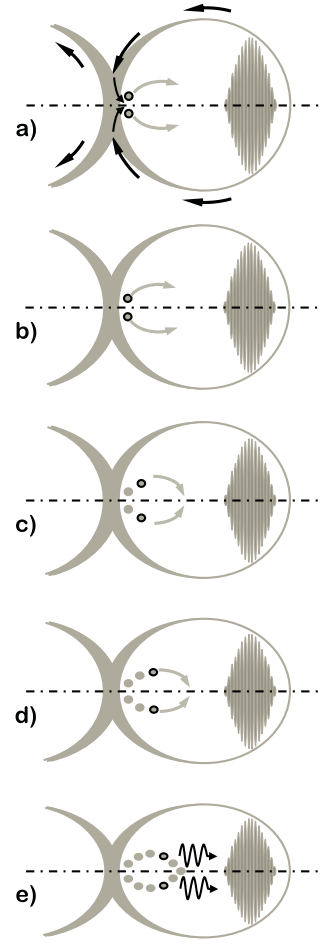



**FIGURE 6.16**
Laser plasma betatron source — conceptually. Wave breaking and self-injection — (a). Oscillation of accelerating electron beams in the plasma bubble — (b)-(d), sequential time moments. Betatron radiation produced by oscillating beams — (e).

($\gamma = 2 \cdot 10^3$). The synchrotron radiation is most notable at higher energies, thus we will ignore lower energy radiation occurring during acceleration. The oscillation amplitude $r_b$ can only be guessed very approximately and usually needs to be obtained via careful simulations. In a very rough approximation, $r_b$ is around 1%–10% of the bubble size (which is $\sim \lambda_p$). We therefore assume that $r_b$=0.001mm. Substituting this into the above equations, we obtain $\lambda_c = 0.025$ nm (or $\sim$ 50 keV) and $N_\gamma$ per $\lambda$ is $\sim$ 0.3 per each accelerated electron. Considering that the accelerating bunch can carry tens of pC to nC charge, we can conclude that such a light source can generate many hard X-ray photons.

## 6.6 Glimpse into the future

Plasma acceleration is a technique that opens new opportunities for creating scientific, technological and medical instruments. In this section we will give a brief review of the progress made in laser acceleration to date, and, after comparing the evolution of plasma accelerators to the evolution of computers in the latter half of the 20th century, we will take a glimpse into the future.

### 6.6.1 Laser plasma acceleration — rapid progress

The last decade has yielded rapid progress in the field of laser plasma acceleration. The pace of research and development in this area received a significant boost in 2004 when the first quasi-monoenergetic beam was generated.[1]

In 2006, Oxford and Berkeley teams[2] broke the GeV barrier in laser plasma acceleration and demonstrated quasi-monoenergetic properties of this accelerated beam. Demonstration of these promising properties further increased the research momentum, and applications of the accelerated beam started to be developed, based on generation of radiation in conventional as well as plasma wigglers (betatron radiation).

The first use of laser plasma-produced betatron radiation for biological imaging was reported[3] in 2011. Multi-GeV laser plasma acceleration[4] was mastered to produce around 4 GeV beams and further progress is expected.

[1] S. Mangles, Nature, 2004.
[2] W. Leemans et al., Nature Physics 2006.
[3] S. Kneip et al., Applied Physics Letters, 2011.
[4] W. Leemans et al., Phys. Rev. Letters, 2014.

### 6.6.2 Compact radiation sources

Laser-driven plasma accelerators can already generate electron beams with several GeV of energy, ~10 fs bunch duration and ~10-100 pC of charge per bunch.

These parameters make laser plasma technology potentially suitable for creating compact radiation sources (see Fig. 6.17). As the beams in these sources are created by a laser in the first place, such sources would also have the advantage of automatic synchronization of accelerated electron beams and generated X-rays with the initial laser pulses. Such synchronization is a powerful asset for *time-resolved studies*.

The far-from-desirable parameters of laser plasma light sources so far include the repetition rate, wall-plug efficiency and beam qualities (energy spread and emittance).

The repetition rate and efficiency are presently limited by laser technology and are the subjects of active research by many groups. Various promising ideas that are beyond the scope of this book have been suggested and are being developed.

The beam quality, which is a focus of significant attention, is gradually improving, and eventually the laser plasma accelerated beam might be suitable for the generation of coherent radiation in a free electron laser application.

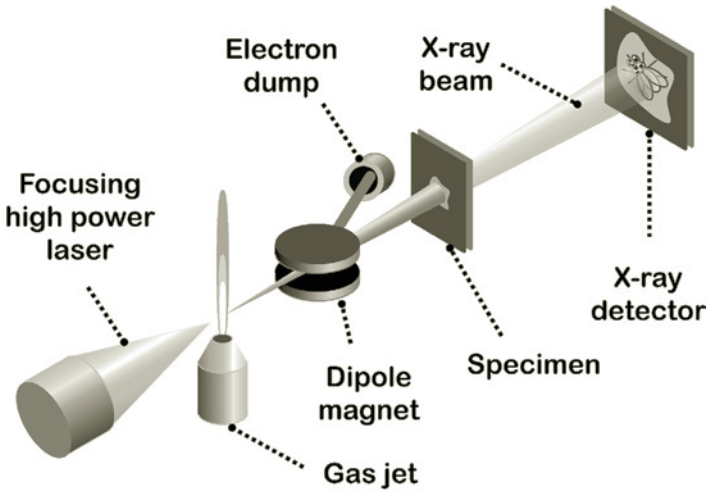



**FIGURE 6.17**
Laser plasma betatron radiation light source — conceptually.

Modern synchrotron-based light sources are large machines with perimeters of several hundred meters. The linac-based free electron lasers can be around a kilometer or more in length. Both of these types of light sources operate with electron beams of a few to about 10 GeV.

Despite the fact that similar electron energies can be reached in a much more compact laser plasma accelerator, it is unlikely that laser plasma-based light sources would en-

tirely replace conventional light sources in the foreseeable future. It is more probable that the types of light sources will evolve in a similar manner to that of computers.

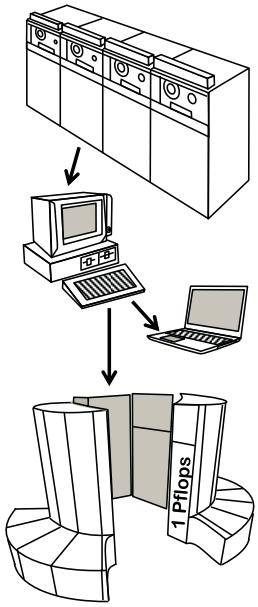



**FIGURE 6.18**
Computers' evolution.

### 6.6.3 Evolution of computers and light sources

Early computers were large, bulky and slow. Development of compact personal computers started in the early 1980s, but was not accepted immediately: "IBM bringing out a personal computer would be like teaching an elephant to tap dance," as newspapers mocked around 1981. Still, as we know today, a plethora of large computers and super-computers co-exist with a variety of personal and compact computers, from laptops to mobile phones and smart watches (Fig. 6.18).

The expected future evolution of light sources may follow a similar pattern. The synchrotron-based light sources and FELs — (a) and (b) in Fig. 6.19 — will eventually be joined by compact plasma-based light sources — (c) in Fig. 6.19 — as a result of intense research, commercialization, and work with users, the industry and economists — efforts of all of which will result in a change of the paradigm.

All types of light sources will then continue to co-exist and national-scale facilities will be complemented by a variety of compact plasma acceleration based light sources.

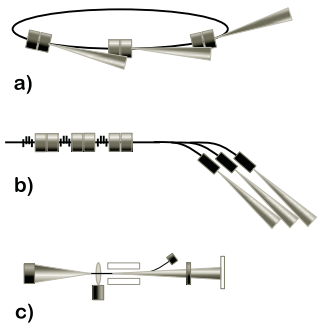



**FIGURE 6.19**
Light sources' evolution.

## 6.7 Plasma acceleration aiming at TeV

While application of plasma acceleration to compact light sources is practically within reach, the application of plasma acceleration technology to high energy physics discovery machines is significantly further away. In this section, we will briefly review some of the primary challenges to plasma acceleration on the way to TeV energy.

### 6.7.1 Multi-stage laser plasma acceleration

In laser plasma acceleration, the laser pulse propagating through a medium (plasma) has $v < c$ and the accelerating electrons that quickly become relativistic will soon dephase from the plasma wave. Consequently, if we are aiming at multi-tens of GeV or TeV acceleration of electrons, many stages of acceleration will be necessary.

The length of a single stage can be estimated by taking into account that the group velocity of a laser pulse is given by

$$v_g = \sqrt{1 - \omega_p^2/\omega^2} \tag{6.46}$$

and that the dephasing occurs when an electron outruns the wave by a half of a period. For a relativistic electron the de-

phasing time $t_d$ is thus given by

$$(c-v_g)\, t_d = \lambda_p/2 \tag{6.47}$$

We then substitute the expression for the group velocity and get the estimate for the dephasing length:

$$L_d \approx \lambda_p\, \omega^2/\omega_p^2 \tag{6.48}$$

For example, for a laser with a wavelength of 1 $\mu$m and $\lambda_p = 30$ $\mu$m, the dephasing length is $L_d \approx 30$ mm. With an accelerating gradient (at the corresponding plasma density) of 1 GeV/cm, a single stage could yield around 3 GeV. Acceleration to a TeV would thus require several hundred stages. Preservation of beam qualities during multi-stage acceleration is the research area that promises significant advances in the near future.

### 6.7.2 Beam-driven plasma acceleration

Plasma can be excited not with a laser pulse, but with a short intense bunch of charged particles (e.g., electrons). In this case, the bubble will be formed due to the bunch's field, and will have a very similar shape and properties to the laser example.

This *beam-driven acceleration* approach (Fig. 6.20) has an advantage in that the driver beam has $v = c$, and thus dephasing of the witness beam from the driver is no longer an issue. Another advantage of this method is that the driver beam can carry much more energy than a laser pulse.

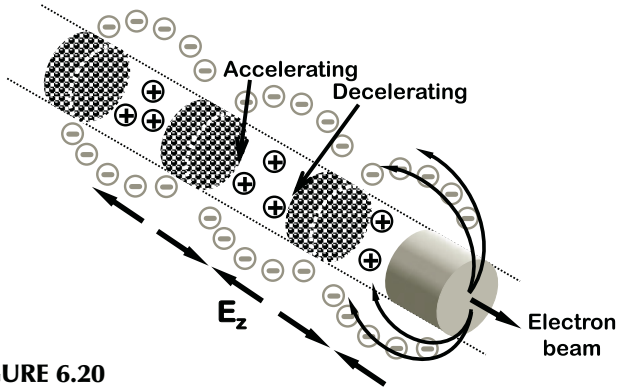



**FIGURE 6.20**
Beam-driven plasma acceleration — conceptually.

These advantages manifested themselves via a much higher beam energy achieved in beam-driven acceleration — the maximum final energy obtained so far is around 80 GeV with an initial beam energy of 42 GeV (which acted in this SLAC linac energy doubling experiment[5] both as a witness as well as a driver).

[5] I. Blumenfeld et al., Nature, 2007.

## 6.8 Laser-plasma and protons

So far, we have only discussed acceleration of electrons. However, plasma acceleration of protons and ions is possible and is being actively developed. We will touch here on this subject only briefly, and leave the details for Chapter 9.

The major reason why laser plasma acceleration of protons should be developed is its potential value to proton therapy. Conventional proton therapy systems — which require around 250 MeV protons and include beam sources (cyclotron or synchrotron), beamlines and especially the beam delivery gantries — are large and expensive. A desire to create compact laser plasma acceleration proton therapy systems is one of the main motivations for development of plasma acceleration of protons.

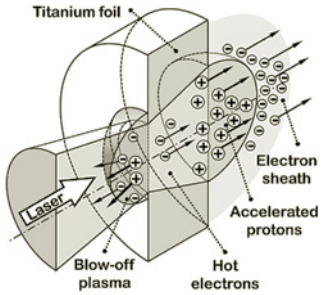



**FIGURE 6.21**
Sheath laser plasma acceleration of protons or ions.

Proton and ion laser plasma acceleration is a rapidly developing area, but it is not quite yet ready for the textbooks. Different models, with different assumptions and simplifications, are created to explain and predict beam properties in different regimes of acceleration. One of the models, illustrated in Fig. 6.21, is sheath laser acceleration of protons. In this case, the laser heats and ionizes a foil, creating a sheath of hot electrons moving away from the foil, which in turn pull and accelerate the ions from the plasma. This particular mechanism, unfortunately, creates a very wide energy spectrum for the accelerated beam. Moreover, its scaling with the laser power is not favorable. We will discuss other more suitable mechanisms in Chapter 9.

## EXERCISES

6.1 *Chapter materials review*.
What laser intensity (in W/cm$^2$) would correspond to a normalized vector potential of $a_0 = 10$, and what are the maximum values of the electric and magnetic fields in the laser wave, for a ruby laser or for a $CO_2$ laser?

6.2 *Chapter materials review*.
Assume that we get 1 GeV electron bunch from a laser plasma accelerator, and would like to create an undulator from plasma using the focusing force of the plasma's ions. Suggest the plasma parameters and the amplitude of the beam oscillation that would correspond to the radiation at the boundary of the undulator regime, i.e., with the undulator parameter equaling $K = 1$.

6.3 *Mini-project*.
Assume that the LHC proton beam of $E = 7$ TeV is going to be used as a driver for the plasma acceleration of electrons. Select the parameters for the plasma and estimate to what bunch length you would need to compress the LHC beam so that it could be used for plasma acceleration. Define, roughly, the parameters of the corresponding bunch compressor.

6.4 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this chapter in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

6.5 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 7 Light Sources



In this chapter, we will focus on light sources — first synchrotron radiation light sources and then Compton light sources. For the former, we will base our observations on results derived in Chapter 3, and will then introduce the necessary formalism for the latter.

## 7.1 SR properties and history

We know from courses on electrodynamics that electromagnetic radiation is emitted by charged particles when they are accelerated. In the particular case when the relativistic particles move on a curved trajectory — i.e., when they are accelerated radially (when the velocity is perpendicular to the acceleration vector) — the emitted electromagnetic radiation is called *synchrotron radiation*.

This SR was at first considered a nuisance, as it causes energy losses in the accelerated particles. Its unique properties, however, were eventually shown to have the potential to pave the path to a new and important type of scientific instrument: the SR sources.

### 7.1.1 Electromagnetic spectrum

The usefulness of SR and Compton sources is due to the fact that they can cover a large range of the electromagnetic spectrum (Fig. 7.1), from infrared (IR) and ultraviolet (UV) to vacuum ultraviolet (VUV: the wavelength range strongly absorbed in air — thus the name), and on to hard X-rays and $\gamma$-rays.

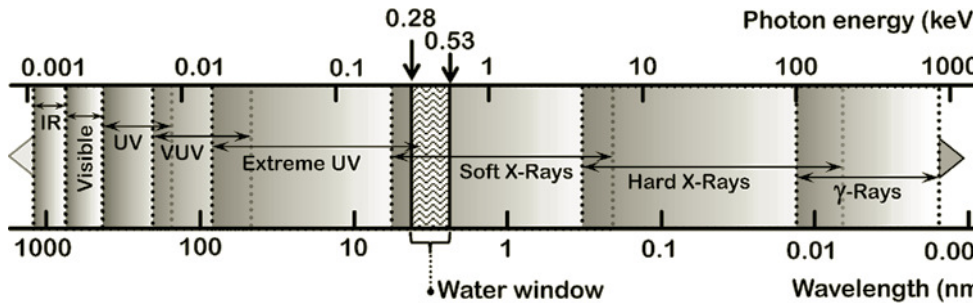



**FIGURE 7.1**
Electromagnetic spectrum covered by SR and Compton sources.

DOI: 10.1201/b18696-7

The EM spectrum shown in Fig. 7.1 also indicates a particular range of wavelengths — the so-called *water window* is defined as the range between the K-absorption edges of oxygen (0.53 keV) and carbon (0.28 keV) — see Fig. 7.2. In this range of soft X-ray energies, water is relatively transparent, which simplifies SR experiments when biological samples need to be used as water solutions.

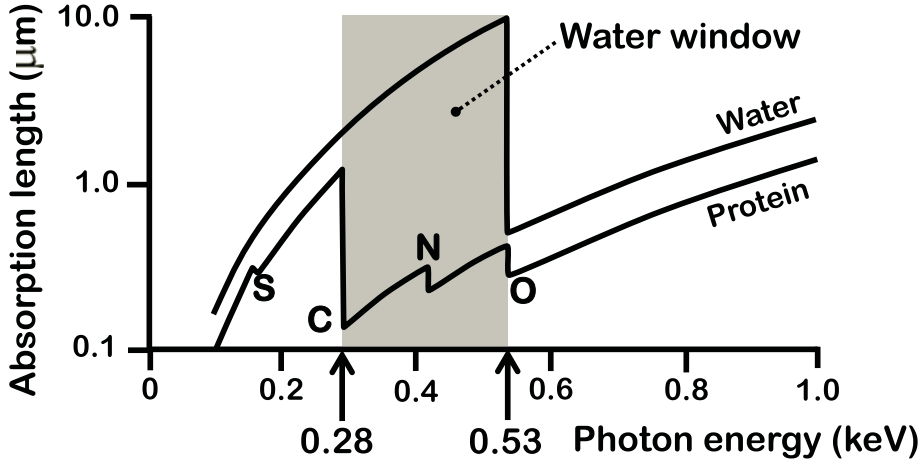



**FIGURE 7.2**
Photon attenuation in water in comparison with a typical protein.

The success and widespread use of SR and Compton sources is thanks to their flexible spectral parameters. This allows one to select the photon energy required for an experiment by adjusting the energy of the beam or laser, or the field strength of wigglers or undulators (useful, for example, when trying to fit within the range of the water window).

### 7.1.2 Brief history of synchrotron radiation

In 1944, D. Ivanenko and I. Pomeranchuk predicted that the maximum energy of electrons in a betatron is limited due to energy losses caused by radiation of relativistic electrons.

This radiation was first observed around 1947, by accident, in a General Electric 70 MeV synchrotron (this gave the name *synchrotron* to the observed radiation), in the *visible* spectrum. It is interesting to note that earlier deliberate attempts to find this radiation in a betatron had failed, as researchers looked for the radiation in a microwave range where the betatron walls were opaque.

The first physics experiments with SR were conducted in 1956 at Cornell, on a 320 MeV synchrotron. In this run, D. Tomboulian and P. Hartman studied the spectral and angular properties of the radiation and also made the first soft X-ray spectroscopy experiments, investigating the trans-

parency of beryllium and aluminum foils near the K and L edges.

The National Bureau of Standards (now National Institute of Standards and Technology) was the next to use SR properties to their advantage, modifying a section of a vacuum chamber of a 180 MeV electron synchrotron to enable access to SR. Soon, it was apparent that the era of SR light sources had begun.

## 7.2 Evolution and parameters of SR sources

A large demand for new scientific instruments stimulated enormous advances in SR light source technology. In just a few decades, several generations of SR sources technology have been developed, each exhibiting an improvement with every evolutionary step.

### 7.2.1 Generations of synchrotron radiation sources

It is now a tradition to distinguish the many generations of SR sources according to the following classifications. The first-generation light sources are the accelerators built for high energy physics, nuclear physics or other purposes, which used for synchrotron radiation experiments parasitically.

A large demand for SR experiments resulted in the construction of dedicated accelerators, creating the second-generation of purpose-built synchrotron light sources. The SRS at Daresbury, England, was the first dedicated machine (operated between 1981 and 2008).

The second-generation light sources employed SR emitted from bending magnets. Advances in accelerator science and technology, inspired by a demand from SR users, quickly created an opportunity for the next technological breakthrough: the third-generation light sources — accelerators optimized for high brilliance due to low electron beam emittance and the use of *insertion devices* (wigglers and undulators). Examples of such SR sources include the European Synchrotron Radiation Facility (ESRF) in France, the Diamond light source in the UK and many others.

The third-generation of SR sources is presently the most widespread. There are several tens of such machines around the world and the number is growing, following demand in the field of science. The brightness of the third-generation machines is several orders of magnitude higher than that of the previous generation (see Fig. 7.3) and exceeds — by about ten orders of magnitude — the brightness of the sources available in the beginning of the 20th century.

The fourth-generation light source was brought to fruition via the *free electron laser* idea, which was developed

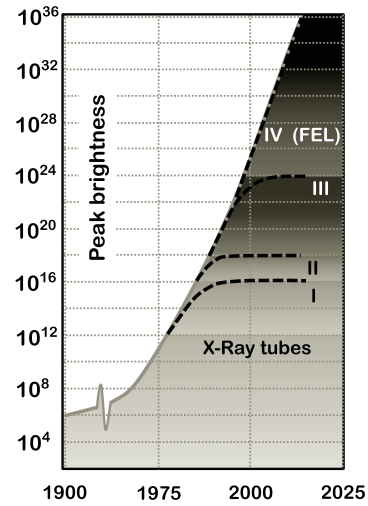



**FIGURE 7.3**
Generations of SR sources. Brightness is expressed in the units of the number of photons per $s \cdot mm^2 \cdot mrad^2 \cdot 0.1\%BW$.

in the 1970s by John Madey. An essential part of the FEL, the undulator (invented in 1947 by Ilya Ginzburg) provided a significant cornerstone for the foundation of FEL technology development.

The fourth-generation light sources enjoy all the latest developments of accelerator science: they are linac-based with low-emittance photo-injectors, assisted with several bunch compressors that help to obtain the ultra-short bunches. Examples of such FELs include: FLASH (Germany), LCLS (USA), and SACLA (Japan).

The next generation of SR light sources will inevitably arrive, and very soon. The exact design and underlying technology are not yet determined, and in fact several ideas are in competition, one being the *ultimate storage rings* discussed in Chapter 3 and another being related to plasma acceleration light sources discussed in Chapter 6, among others.

### 7.2.2 Basic SR properties and parameters of SR sources

Without repeating the formulae, which can be found in Chapter 3, let us briefly recall our approach for deriving important parameters of SR sources and of the synchrotron radiation itself.

Recall that a simple back-of-the-envelope treatment of SR was possible when the amount of radiation *left behind* was determined from simple geometrical consideration and the volume integral over the field squared gave us the energy lost per unit of length. This led us immediately to the estimation of the cooling time and, after considering the quantum (statistical) character of radiation, we came up with an estimate for the equilibrium horizontal emittance of the beam (with vertical equilibrium emittance determined by the coupling of the ring).

Knowledge of equilibrium emittances, as well as knowledge of the single photon emittance that we estimated on the way, firstly led us to an estimation of the SR flux and then of the brightness of SR sources.

High-intensity photon flux (defined as the number of photons per second and per spectral bandwidth) allows for either rapid experiments or the use of weakly scattering objects or crystals. The brightness of a light source (also called, interchangeably, brilliance or spectral brightness) is determined as the number of photons emitted per second per surface area and per solid angle and per fraction of a spectral bandwidth. The high brilliance of SR sources is enabled by low emittance (and thus a low emitting area) and a low divergence of radiation emitted by an ultra-relativistic beam.

$$Brilliance = Photons/s \cdot mm^2 \cdot mrad^2 \cdot BW$$

SR covers a broad light spectrum, from microwaves to hard X-rays, hence allowing a variety of experiments. Experiments that require precise photon energies can benefit either from the use of a monochromator or from the ability to adjust

the emission wavelength of the insertion devices. Polarization of SR can be either linear or circular, as required by the experiment, and depends on the design of insertion devices.

The coherence properties of SR sources vary — from partial coherence in third-generation sources to full temporal coherence in early FELs and full temporal and spatial coherence in the latest designs of pre-seeded FELs. We will discuss some of the corresponding methods in Chapter 10.

The temporal structure of SR radiation in third- and fourth-generations span many orders of magnitude, starting from tens of ps SR flashes in ring-based sources to ultra-short tens of fs flashes in FELs. The sources are designed to have high spatial and temporal stability of the emitted radiation, to submicron levels in space, and tens of fs levels in time.

Discussing the dependences of SR properties on parameters in Chapter 3 led us to the concept of the ultimate brightness of a *diffraction-limited* SR source — the *ultimate storage ring* where the equilibrium horizontal emittance is at least as small as the emittance of radiated photons. Reaching the ultimate parameters, as Eq. 3.30 suggests, may require increasing the perimeter of the ring and decreasing the average of $\eta^2/\beta_x$ i.e., employing tighter focusing, which can be achieved by shortening the distances between focusing elements.

*The tendencies to increase the ring perimeter and tighten the focusing (increased segmentation of the ring) are demonstrated by the developing plans of upgrades of many third-generation SR sources as well as by the emerging designs of ultimate storage rings.*

## 7.3 SR source layouts and experiments

The modern third-generation light sources are state-of-the-art facilities that provide multiple X-ray beamlines for a variety of experiments (Fig. 7.4).

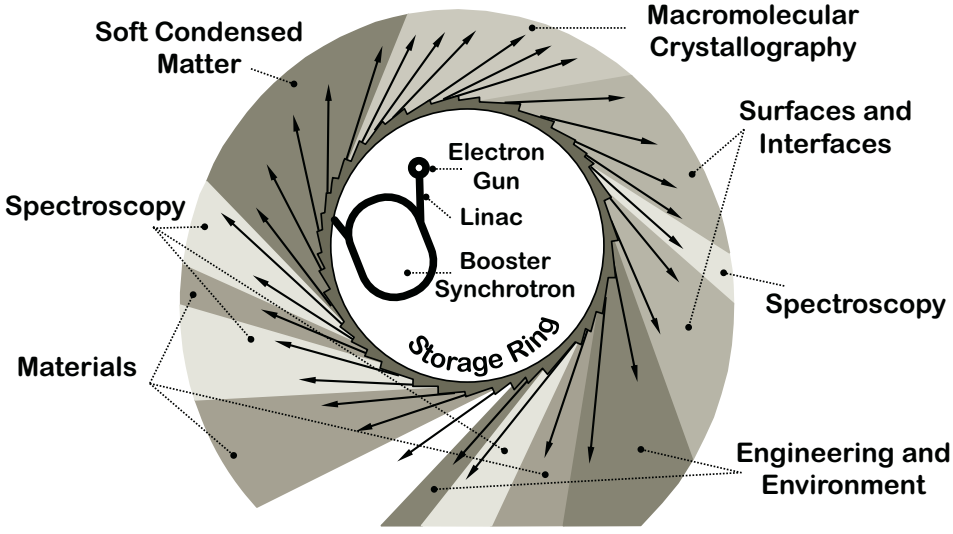



**FIGURE 7.4**
Generic SR light source with multiple X-ray beamlines and showing typical allocation of beamlines to experiments.

In this section, we will look into a generic layout of such sources, and will briefly touch on their experimental capabilities.

### 7.3.1 Layout of a synchrotron radiation source

A schematic of a generic third-generation synchrotron radiation source is shown in Fig. 7.5. Electrons are typically generated in an RF gun and accelerated in a *linac* (usually to a few hundred MeV), further accelerated to the required energy (of a few GeV) in a *booster*, and then injected into the storage ring where the circulating electrons emit an intense beam of synchrotron radiation.

The optics of the storage ring are arranged in such a way so they have many empty drift sections where *insertion devices* (ID) — wigglers and undulators — can be installed. Each of these IDs will direct light into a corresponding X-ray beamline, which can then be tailored to a particular type of experiment (life science, materials, etc.). The typical amount of X-ray beamlines is a couple of dozen, as illustrated in Fig. 7.4.

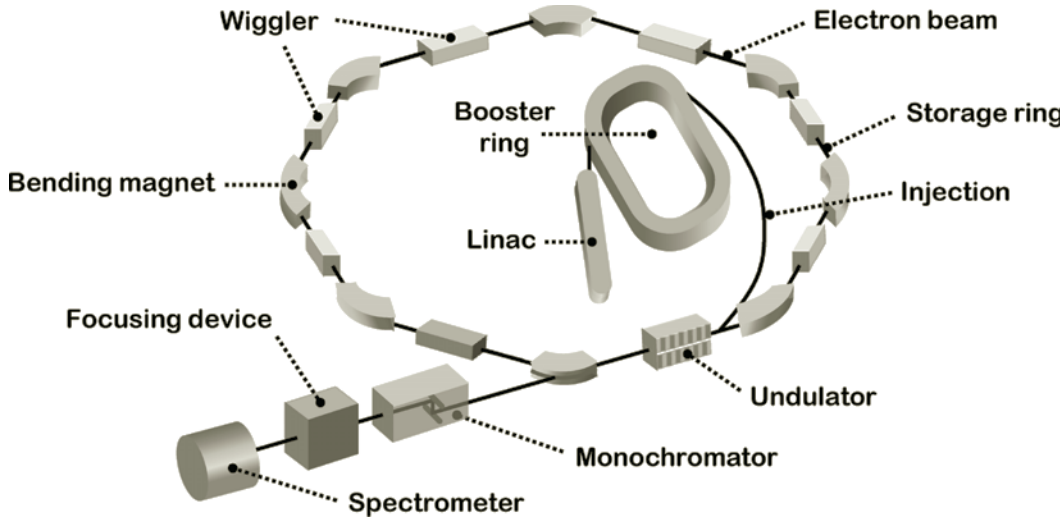



**FIGURE 7.5**
Schematics of a generic third-generation SR light source.

*For example, for $B = 1.4$ T and $E = 3$ GeV $\varepsilon_c = 8.4$ keV.*

The second-generation sources use radiation emitted in bending magnets, which emit a continuous spectrum characterized by critical energy $\varepsilon_c$, which can be estimated as $\varepsilon_c(\text{keV}) = 0.665\ \text{B(T)}E^2$ GeV.

The third-generation employs insertion devices (undulators and wigglers). Either of these devices is a periodic array of magnetic poles that provide a sinusoidal magnetic field **B** on axis: $\mathbf{B} = (0, B_0 \sin(k_u z), 0)$ where $k_u = 2\pi/\lambda_u$. The maximal radius of the curvature of the orbit in the sinusoidal

field $R$ defines two distinct regimes (see Section 3.3.5). If $2R/\gamma \ll \lambda_u/2$ (parameter $K \sim \gamma\, \lambda_u/R \gg 1$), then the radiation emitted at each period of sin-like field is independent; this corresponds to a wiggler regime (similar to radiation from a sequence of bends). The insertion devices working in a wiggler regime are usually used to achieve high photon energies and flux.

In contrast, the undulator regime corresponds to the case $2R/\gamma \gg \lambda_u/2$ (or $K \ll 1$), which means that the entire wiggling trajectory will contribute to radiation. Undulators are typically used to generate high brilliance radiation in a quasi-monochromatic spectrum — the bandwidth of undulator radiation is inversely proportional to the number of undulator periods $N_u$ and can be estimated as $\Delta f/f \sim 1/N_u$. The precise definition and meaning of parameter $K$, as well as the wavelength of undulator radiation, will be discussed in the next chapter, in relation to FELs.

To a large extent, the scientific performance of third-generation SR sources depends on their stability. The current in the storage ring, decaying between injection cycles (due to the *Touschek effect*: intrabeam scattering resulting in a change of particle momentum and its consequent loss on the energy acceptance aperture) forces the power of emitted SR to change, affecting the temperature regime and stability of the ring and of the X-rays' beamlines. The *top-off injection* (also called *top-up*) is the operation regime (see Fig. 7.6) that keeps the beam current in the ring almost constant, improving the stability significantly. In this regime, a small amount of current is injected into the ring much more frequently than in the standard regime.

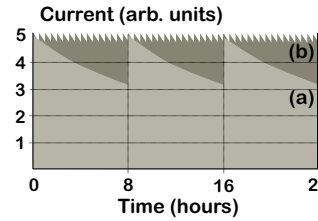



**FIGURE 7.6**
Current in SR light source without (a) and with (b) top-up injection mode.

### 7.3.2 Experiments using SR

Synchrotron radiation allows for a wide array of experiments, ranging from utilization of phenomena in X-ray scattering, X-ray absorption or X-ray fluorescence, to various advanced methods that enhance the resolution of obtained images (e.g., relying on X-ray absorption near the atomic spectral edges of particular elements contained in the studied samples). Already, a variety of imaging methods allow for the use of SR in biological, chemical, medical and material studies, and in many other areas of science and technology.

SR experiments often require an X-ray beam with a well-defined wavelength. *Monochromatization* is typically performed by the crystal monochromators. A variety of configurations of crystal monochromators are possible — Fig. 7.7 shows two particular arrangements. In both of these cases, the geometry is selected in such a way that the desired X-ray wavelength $\lambda$ corresponds to *Bragg conditions*

$$n\,\lambda = 2\,d\,\sin\theta \tag{7.1}$$

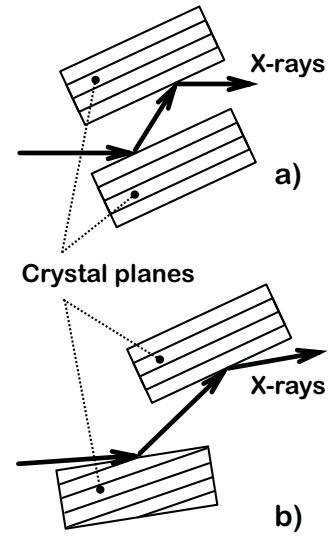



**FIGURE 7.7**
Crystal monochromator of X-rays. Symmetric case (a) and asymmetric case (b).

which correspond to the maximal reflection of X-rays from the crystal. In the above equation, $n$ is an integer, $d$ is the distance between the crystal planes and $\theta$ is the angle between X-rays and scattering crystal planes. The monochromator plates are usually made from crystals of Si or Ge. The symmetrical configuration shown in Fig. 7.7.a is standard, while the asymmetrical one (Fig. 7.7.b) allows for the increase in angular resolution — thus narrowing the resulting energy spread of the X-ray beam.

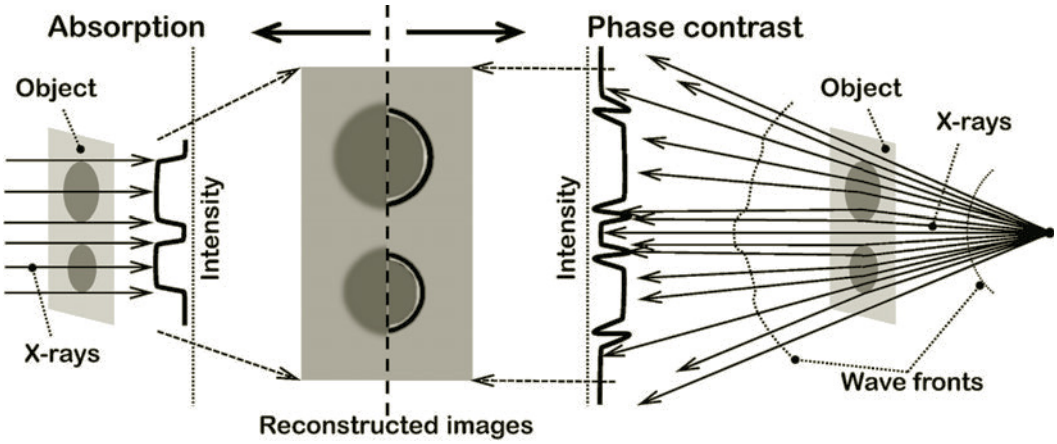



**FIGURE 7.8**
Absorption (left) and phase contrast (right) X-ray imaging and comparison of reconstructed image (middle).

Due to the small size of the area that emits X-rays, the SR light sources can utilize an advanced technique called *phase contrast imaging* — shown in Fig. 7.8 in comparison with standard absorption imaging.

Phase contrast imaging is particularly appropriate for studies of biological objects where the density difference, and thus absorption difference between different tissues, is minimal, which complicates the goal of achieving high-resolution images relying on absorption (left part of Fig. 7.8). However, benefiting from the point-like nature of the emitting source, one can increase the distance between the object and the detector plane, and rely instead on refraction of X-rays caused by the density variations in the object. The consequent interference pattern on the detector plane will have much sharper features, thus reconstructing images with better resolution and contrast (right side of Fig. 7.8).

The phase contrast imaging technique is especially beneficial for laser plasma betatron light sources (see Chapter 6), as the emitting areas can have sizes below a micrometer. The relatively low average brightness of such sources would then be compensated by higher spatial resolutions, which are ad-

ditionally enhanced by extremely short temporal durations of X-ray flash.

To conclude this section, let us recapitulate the capabilities of the modern SR sources by referring to a well-known example. In 1952, DNA structure was studied in R. Franklin and R. Gosling's experiments using an X-ray tube that had a brilliance of around $10^8$ ph/sec/mm$^2$/mrad$^2$/0.1 BW . At that time, the duration of exposure needed to acquire the necessary statistics was typically as long as one day (around $10^5$ sec). The modern third-generation light sources with a brilliance of the order of $10^{20}$ can provide the same exposure in just 100 ns.

The much higher brightness is not the only advantage of modern SR sources. Engaging lasers in combination with SR sources creates a completely new type of experiment — a *pump-probe* configuration, in which a laser pulse synchronized with the beam revolution in an SR source excites the object just before the main X-ray pulse arrives from the SR source (as illustrated in Fig. 7.9).

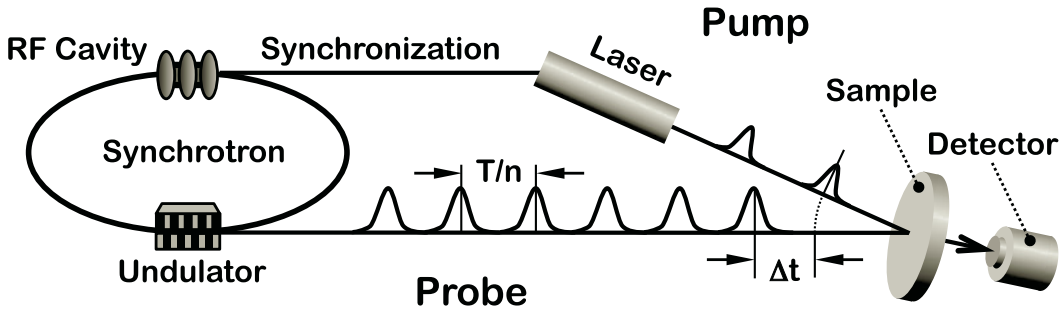



**FIGURE 7.9**
Pump-probe experiment arrangement. Here $T$ and $n$ are revolution period and number of bunches in the SR ring, $\Delta t$ is time delay between the pump laser pulse and SR probe pulse.

The specific feature of this kind of experiment is the ability to vary the time delay $\Delta t$ between the pump and probe pulses. The pump-probe experiments are therefore in particular useful for studies of ultra-fast phenomena such as spin dynamics in metals, structural molecular dynamics of proteins, photosynthesis, ultrafast photo-switching and many others.

## 7.4 Compton and Thomson scattering of photons

Thomson and Compton processes describe the scattering of an EM wave or photon on a charged particle. The Compton scattering, in particular, can be very useful for creating com-

pact X-ray sources — enabled by the development of electron accelerators and laser technologies.

In this section — after reviewing the basic formalism of the Thomson and Compton processes — we will discuss the typical design and characteristics of Compton X-ray sources.

### 7.4.1 Thomson scattering

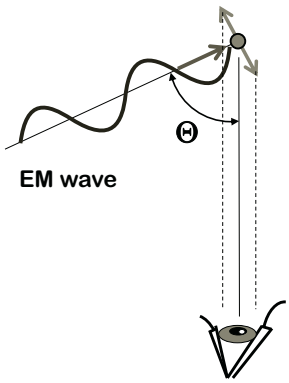



**FIGURE 7.10**
Thomson scattering.

The elastic scattering of an electromagnetic plane wave by an electron at rest (or low energy $E$) with mass $m_e$ and charge $q$, is a process known as *Thomson scattering*.

The total cross section of a classical Thomson scattering is given by the following equation:

$$\sigma_{Th} = \frac{8\pi}{3} r_e^2 \approx 0.665 \cdot 10^{-28} \, [m^2] \tag{7.2}$$

And the differential cross section, illustrated in Fig. 7.10, is equal to

$$\frac{d\sigma}{d\Omega} = \frac{1}{2} r_e^2 \;\; 1 + \cos^2\theta \tag{7.3}$$

Thomson scattering is an approximation of an elastic process — the energies of the particle and photon are the same before and after the scattering (i.e., the recoil of the electron can be neglected, in contrast to the Compton scattering).

### 7.4.2 Compton scattering

Compton scattering describes the inelastic process where we can no longer neglect the transfer of energy between the particle and the photon.

We are, in particular, interested in the instance when a collision between a high-energy electron and a low-energy photon results in a substantial fraction of the electron energy being transferred to the photon. In the laboratory reference frame, this manifests as backscattering of the photon with a significant energy boost; this process is known as *Compton backscattering* (or inverse Compton scattering), as illustrated in Fig. 7.11.

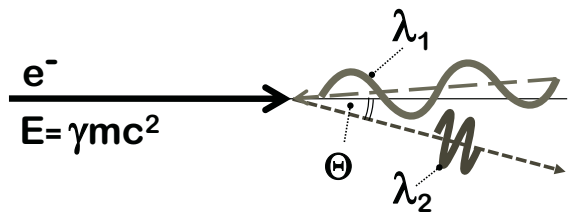



**FIGURE 7.11**
Compton backscattering. Initial photon with wavelength $\lambda_1$ and after scattering with $\lambda_2$.

In the approximation of small angles and relativistic electrons (see Fig. 7.11), the wavelength of the photon after Compton backscattering is described by the following equation:

$$\lambda_2 = \lambda_1 \ \ 1 + \theta^2\gamma^2 \ / \ 4\gamma^2 \tag{7.4}$$

As we can see from Eq. 7.4, in the case of relativistic electrons, the photon gains considerable energy after interaction: its wavelength is shortened by the factor of $4\gamma^2$. Let's consider two examples in the case of green light with $\lambda_1$= 532 nm (corresponding photon energy is 2.33 eV). If the total electron energy is 5.11 MeV ($\gamma$ = 10) then $\lambda_2$ = 1.33 nm (equivalent to 0.93 keV energy of the photons). For a slightly larger energy of electrons of 18.6 MeV ($\gamma$ = 36.5) the scattered wavelength would reach an angstrom: $\lambda_2$ = 0.1 nm (or 12.4 keV).

*Green laser (532 nm) scattered from an 18.6 MeV electron beam turns into X-rays with 0.1 nm wavelength.*

A derivation of the Compton process kinematics can be obtained through considering the relativistic invariants of electrons and photons before and after collision, as illustrated in Fig. 7.12. Below, we will reproduce only the final results.

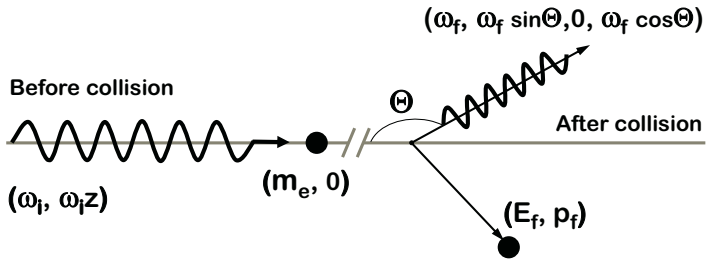



**FIGURE 7.12**
Compton scattering in the rest frame of an electron and relativistic invariants.

The resulting expression for the total Compton scattering cross section in the center of mass reference frame is as follows (we assumed $c$ = 1 in this and the next section):

$$\sigma_{tot} = \frac{2\pi\ m_e^2\ r_e^2}{E_{cm}^2} \ln\left[\frac{E_{cm}^2}{m_e^2}\right] \tag{7.5}$$

where

$$E_{cm} = m_e^2 + 2p_f\omega_f\,(1-\cos\theta) \tag{7.6}$$

In the laboratory reference frame the total cross section is:

$$\sigma_{tot} = \frac{2\pi r_e^2}{x_1}\left(1 - \frac{4}{x_1} - \frac{8}{x_1^2}\right)\ln(1+x_1) + \frac{1}{2} + \frac{8}{x_1} - \frac{0.5}{(1+x_1)^2} \tag{7.7}$$

The differential cross section in the laboratory frame:

$$\frac{d\sigma}{d\Omega} = 2r_e^2\left(\frac{\omega_f}{m_e x_1}\right)^2\left(4y(1+y) - \frac{x_1}{x_2} - \frac{x_2}{x_1}\right) \tag{7.8}$$

where the parameters are defined as follows:

$$x_1 = 2\gamma \frac{\omega_i}{m_e}(1-\beta\cos\phi_1); \quad x_2 = -2\gamma \frac{\omega_f}{m_e}(1-\beta\cos\phi_2); \quad y = \frac{1}{x_1}+\frac{1}{x_2}$$

Here $\beta$ is the relativistic factor and angles $\varphi_1$ and $\varphi_1$ are defined as shown in Fig. 7.13 in the next section.

### 7.4.3 Compton scattering approximation

We will now consider the approximation of the Compton cross section given in the previous section in case $x_1 \ll 1$ or $\gamma\ \omega_1 \ll m_e$. In this instance, the expression in the curved parenthesis in Eq. 7.7 will simplify to

$$\left\{\left(1-\frac{4}{x_1}-\frac{8}{x_1^2}\right)\ln(1+x_1)+\frac{1}{2}+\frac{8}{x_1}-\frac{1}{2(1+x_1)^2} \quad \rightarrow \quad \frac{4x_1}{3}\right.$$

Therefore the total cross section in this approximation is given by the following equation:

$$\sigma_{tot} = \frac{8\pi\, r_e^2}{3} \tag{7.9}$$

which shows that the total Compton scattering cross section is very close to the Thomson one.

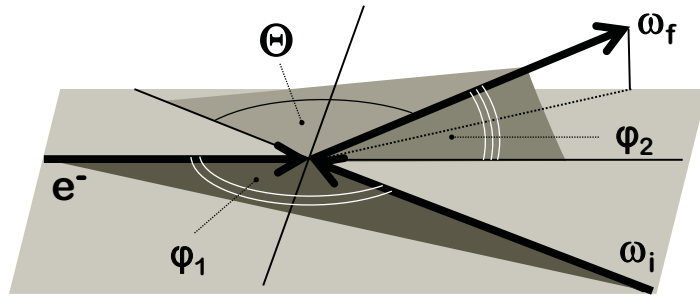



**FIGURE 7.13**
Compton scattering — definition of frequencies and angles.

### 7.4.4 Compton scattering characteristics

In the approximation described above, we can evaluate the rate of emitted X-rays as the product of the Thomson cross section and the *luminosity* $\mathcal{L}$ characterizing interaction between the electron and laser beams. Assuming head-on collision of the beams, we write for the luminosity:

$$\mathcal{L} = \frac{N_e N_\gamma f}{2\pi\,\sigma_x\sigma_y} \tag{7.10}$$

In the above, $N_e$ and $N_\gamma$ are the numbers of electrons and photons in the colliding bunches and $f$ is the repetition frequency of collisions. The beam sizes in Eq. 7.10 are the convolution of electron and photon beam sizes:

$$\sigma = \sqrt{\sigma_e^2 + \sigma_\gamma^2} \tag{7.11}$$

The above equations help us to estimate the rate of X-ray production as

$$\frac{d\mathcal{N}_\gamma}{dt} = \sigma_{tot}\,\mathcal{L} \tag{7.12}$$

In relativistic approximation $\gamma \gg 1$, the expression for the final frequency of Compton scattered photons is given by the following equation

$$\omega_f \approx \frac{2\gamma^2\omega_i(1-\cos\phi_1)}{1+(\gamma\phi_2)^2+2\gamma\frac{\omega_i}{m_e}(1-\cos\phi_1)} \tag{7.13}$$

which identifies the following characteristics of Compton scattering. There is a clear dependence between scattered photon energy and its angle — this can be useful for the selection of monochromatic beam with help of collimation. The majority of the X-ray flux is emitted into a cone with an opening angle of $4/\gamma$. The photons of maximum energy $\omega_c = 4\omega_i\gamma^2$ come from a head-on collision ($\varphi_1 = \pi$), while the photons of half-maximum energy come from a $\varphi_1 = \pi/2$ collision.

The equations and dependencies defined above aid us in estimating basic parameters of Compton-based light sources. Let us now consider the design of such light sources and their typical characteristics.

## 7.5 Compton light sources

A generic Compton light source based on a linear accelerator is shown in Fig. 7.14.

In this design, a train of electron bunches produced by

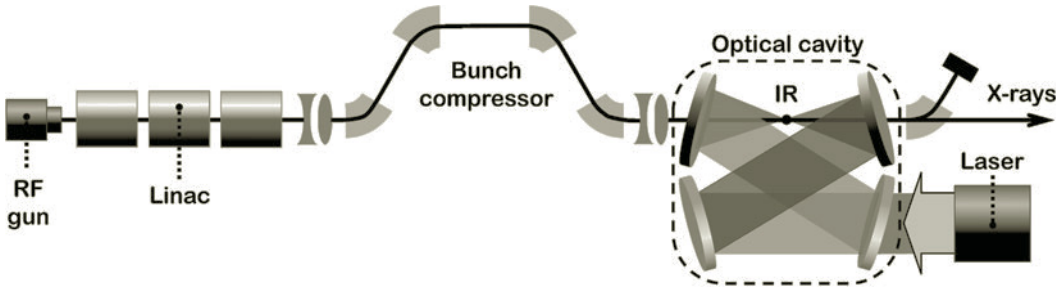



**FIGURE 7.14**
Generic Compton source of linac type.

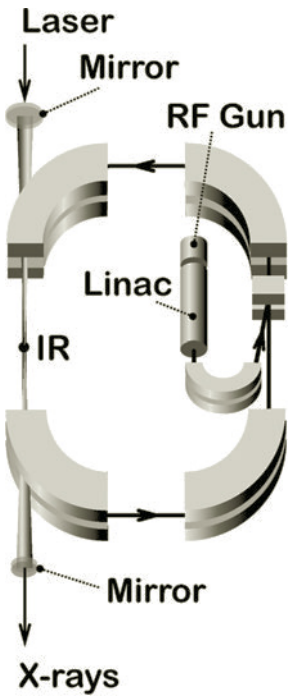



**FIGURE 7.15**
Generic Compton light source based on electron storage ring.

a photo RF gun is accelerated in a linac up to a few tens of MeV (or several hundred MeV, depending on the application). It is then compressed longitudinally and sent for collision with laser bunches accumulated in the laser cavity. The laser pumps the cavity through a semi-transparent mirror and the laser intensity buildup in the cavity can exceed the intensity in the single laser pulse by more than a hundredfold, increasing the brightness of the linac-based Compton source.

Further enhancement of the brightness of the linac-based Compton source can be achieved by employing the method of *energy recovery*. The Compton cross section of electron-photon interaction is rather low and therefore, in typical configurations, the majority of electrons will pass through the laser pulses without interacting. The major fraction of the electron beam can therefore be decelerated, after Compton interaction, and its energy recovered, reducing the required RF power and increasing the current of the electron beam. The energy recovery based Compton sources are mostly suitable for superconducting linac technology.

Another type of Compton source is one based on an electron storage ring, as illustrated in Fig. 7.15. In this case, a short linac injects an electron beam into a compact ring, where — due to SR — the electron beam emittances are cooled, thus helping to achieve a higher luminosity of electron-laser interaction. The laser cavity is typically located around or inside one of the straight sections of the ring with the interaction region (IR) in the center.

Parameter ranges of existing or planned Compton sources allow for their application in a variety of areas. In particular, 10-30 MeV accelerator produces (with a typical laser) X-rays tuneable from a few keV to around 50 keV, which can be applied to high resolution clinical imaging systems or various types of biomedical research.

Either the linac- or ring-based Compton source can be rather compact, fitting in a room of a few meters by a few meters, which enables the use of such sources in areas where it would not be possible before. For example, a THOMX[1] Compton source is being considered for use in cultural heritage applications, and might be installed in a museum for nondestructive studies of precious paintings without the need to transport them.

Compton sources aimed at larger energies of X-ray photons, toward the 1–5 MeV range, have their particular niche. The phenomenon of nuclear resonance fluorescence helps to create imaging instruments with excellent isotopic sensitivity; therefore such Compton sources can also assist in nuclear-waste management.

[1] A. Variola et al., THOMX Conceptual Design Report. LAL RT 09/28, SOLEIL/SOU-RA-2678, 2010.

The Compton sources based on a linac, mentioned above, require superconducting RF (SRF) technology to achieve maximal brightness. Compton sources are compact and therefore may be suitable for small organizations, labs, hospitals or universities. However, most of the modern SRF cavities with frequencies around 1 GHz require operation at 2 K temperature. There is, nonetheless, a contradiction here as the cryogenics system, in particular for 2 K, can be bulky and expensive.

A possible solution for this issue is to revert to lower frequency SRF cavities, which operate at 4 K temperatures, resulting in a significant reduction in complexity, size and cost of the cryogenic system. The practical frequency range of SRF cavities that work at 4 K is around 200-500 MHz. The larger transverse size and lower accelerating gradient of the lower frequency cavities can sometimes be compensated by modifying their design, e.g., use of spoke cavities.

## EXERCISES

7.1 *Chapter materials review*.
Describe the methods of the creation of a monochromatic X-ray beam in Compton sources.

7.2 *Chapter materials review*.
In laser plasma acceleration, the final energy of an accelerated electron beam is 1 GeV. The wavelength of the laser used for laser plasma acceleration is 800 nm. Part of the same laser pulse is redirected with mirrors to collide head-on with the accelerated electron beam. Estimate the energy of photons created in such a Compton source and the angular spread of the photons.

7.3 *Mini-project*.
Select a desired photon energy of a laser plasma acceleration betatron X-ray source (e.g., from 1 to 100 keV) and devise a consistent set of basic parameters describing the source. Discuss the justifications for selecting particular values of certain parameters (for plasma or laser, etc.). Estimate the brightness of the source.

7.4 *Mini-project*.
Select a desired photon energy of a Compton X-ray source (e.g., from 1 keV to 10 MeV) and devise a consistent set of basic parameters describing the source. Discuss the reasons for selecting particular values of certain parameters (for electron beam, laser, etc.). Estimate the brightness of the source.

7.5 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this chapter in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

7.6 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

# 8 Free Electron Lasers



In this chapter we will continue to build upon the results of the previous chapters — particularly Chapter 3 (Synchrotron Radiation) and Chapter 7 (Light Sources) — and will discuss the present reigning champion among the X-ray light sources: the free electron laser.

We will begin with a brief, historical introduction. Then, we will recall the properties of radiation from a sequence of bends, wigglers and undulators, and then discuss how their radiation spectra compare. Next, we will follow up on undulator resonance conditions and microbunching. Finally, we will discuss the precise physical meaning and exact definition of the undulator parameter $K$, which was introduced in earlier chapters as a qualitative factor.

Following this introduction of basic FEL concepts, we will discuss FEL designs and parameters, as well as possible future advances in the evolution of FEL technology.

## 8.1 FEL history

The FEL concept, as well as the term itself, was suggested by John Madey in the early 1970s during his work at Stanford University. His research benefited from the earlier work by Hanz Motz who, in 1953, built an undulator (which was proposed in Vitaly Ginzburg's 1947 theoretical paper, wherein the undulator was described as a device generating electromagnetic radiation via relativistic electrons).

In 1971, John Madey wrote his first FEL-related published journal entry on the subject of stimulated emissions of radiation in a periodic magnetic field. This was shortly followed by a patent on FEL filed in 1972 — but that's a whole 'nother story. The FEL created by Madey used a 43 MeV electron beam to create radiation with wavelengths of 3.4 $\mu$m in a 5 m helical undulator, with a 3.2 cm period and a field of 0.24 T.

Following this pioneering work, many FELs have been created all around the world — their wavelengths gradually shortening as the technology matured. The most modern FELs have recently reached the Angstrom range, creating unsurpassed possibilities for discovery science, bio-medical studies and technology-aimed research.

DOI: 10.1201/b18696-8

## 8.2 SR from bends, wigglers and undulators

We started our discussion of FELs in Chapter 3, where we approached radiation from wigglers and undulators and compared them to radiation from bending magnets. We recall that, for relativistic electrons with $\gamma \gg 1$, the emitted photons go into $1/\gamma$ cone and, if the radius of the curvature of the trajectory in the magnetic field is $R$, then the external observer will see the photons emitted during the particle travel along the arc $2R/\gamma$. This allowed us to estimate the characteristic frequency of SR as $\omega_c = 1.5c\,\gamma^3/R$.

*Recall that $\gamma^3$ dependence of $\omega_c$ is due to the length of the emitting arc ($\propto 1/\gamma$) and photon and particle velocity difference ($v - c \propto 1/\gamma^2$).*

We also recall that the extra factor of $\gamma^2$ appears in this formula for $\omega_c$ due to the difference between the speed of photons $c$ and the speed of particles $v$, estimated as $(1 - v/c) = 1/(2\,\gamma^2)$. Knowledge of the characteristic frequency allows us to determine the spectral characteristic of the SR emitted from the bending magnets.

### 8.2.1 Radiation from sequence of bends

Assume that a set of bending magnets are arranged in a sequence with +–+– polarity with period $\lambda_u$, so that the particle trajectory through this sequence of magnets wiggles as illustrated in Fig. 8.1.

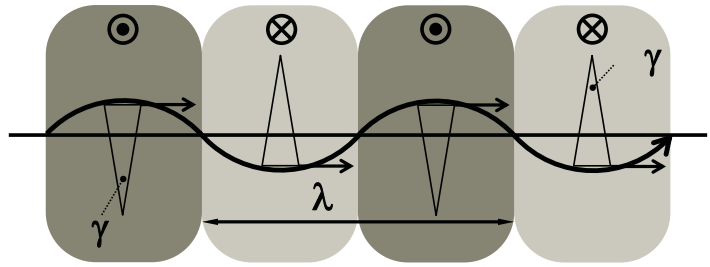



**FIGURE 8.1**
Trajectory and radiation in a sequence of bending magnets.

*Precise definition of K follows in just a couple of pages.*

If the length of the emitting region (that a remote observer can see) is much less than the length of an individual bend, i.e., $2R/\gamma \ll \lambda_u/2$, then the radiation emitted in each bend is independent. Such an arrangement of bends is called a wiggler (and corresponds to $K \gg 1$ where $K \sim \gamma\,\lambda_u/R$, and where $\lambda_u/R$ can be noted as being approximately equal to the maximal angle of the trajectory).

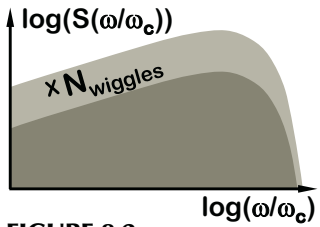



**FIGURE 8.2**
Wiggler (top) and bending magnet (bottom) SR spectra.

In a wiggler configuration, the spectrum of emitted SR is thus expected to resemble to the spectrum from the bends — the spectrum shape will be similar to the spectrum from a bend while the amplitude will be multiplied by the number of wiggles, as shown qualitatively in Fig. 8.2.

The opposite regime $2R/\gamma \gg \lambda_u/2$ is different — the entire wiggling trajectory contributes to radiation. It is logical

to assume that there is a wavelength in this $K \ll 1$ undulator regime that would be resonant with the emitting particle. Indeed, it will emit radiation at this wavelength at each wiggle, allowing coherent buildup of the amplitude. We will look at the spectrum of undulators and also this resonant condition in detail in the following sections.

### 8.2.2 SR spectra from wiggler and undulator

Despite the similarity of the overall shape of the SR spectra of bend magnets and wigglers, the external observer will note an important difference in the details of the spectra.

The fields emitted from wigglers and detected by the external observer manifest themselves as periodic signals — short flashes repeating with a period corresponding to the time of flight between wiggler periods, as illustrated in Fig. 8.3 and Fig. 8.4.

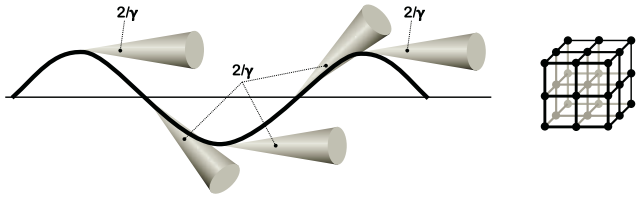



**FIGURE 8.3**
Radiation from wiggler, regime of $K \gg 1$.

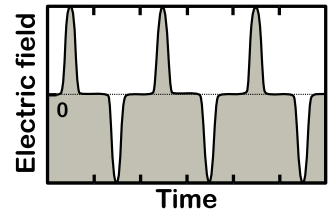



**FIGURE 8.4**
Time profile of radiation observed from wiggler.

Given the periodic nature of radiation emitted by wigglers, the spectrum of SR wiggler radiation should consist of harmonics defined by the wiggler period corrected by the factor $(1 - v/c) = 1/(2\gamma^2)$, which takes into account the relative velocity of particles and radiation (illustrated in Fig. 8.5 at left). The relative width of each peak in the wiggler spectrum corresponds to the number of wiggles $N_w$, i.e., $\Delta\lambda/\lambda \approx 1/N_w$.

As the the entire trajectory contributes to radiation emitted from the undulator, the time structure of the observed radiation is periodic and continuous, as shown in Fig. 8.6 and Fig. 8.7.

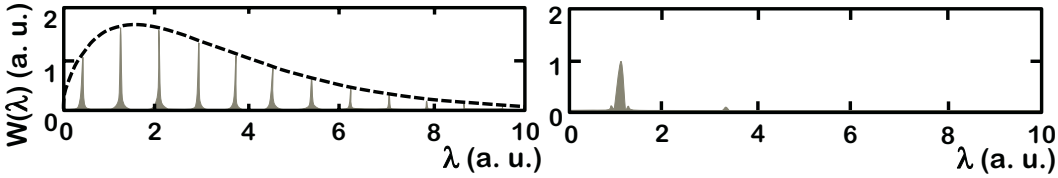



**FIGURE 8.5**
Spectrum from wiggler (left) and undulator (right), qualitative comparison. Dashed line on the left spectrum corresponds to the spectrum from bends of the same strength. Horizontal axis is in units of $\lambda_u/(2\gamma^2)$.

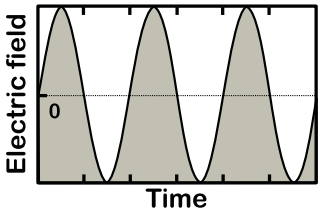



**FIGURE 8.7**
Time profile of radiation observed from undulator.


**FIGURE 8.6**
Radiation from undulator, with $K \ll 1$.

The spectrum corresponding to the observed undulator radiation will thus contain just one harmonic at a wavelength close to $\lambda_u/(2\gamma^2)$. We will clarify this statement in the next section.

### 8.2.3 Motion and radiation in sine-like field

In the example that we considered in Fig. 8.1, we assumed that the segmented field of the wiggler is uniform and sharply changes its sign during the transitions between segments. This may correspond to zero-aperture magnets but, in practice, is not possible.

A much better way to estimate the field of a wiggler or undulator is to assume that the field is sine-like, i.e.,

$$B_y(z) = B_0 \sin(k_u z) \tag{8.1}$$

where $k_u = 2\pi/\lambda_u$.

Let us consider a trajectory through such a sine-like field and let us parameterize it in such a way that the maximum angle of the trajectory is equal to $K/\gamma$, as shown in Fig. 8.8.

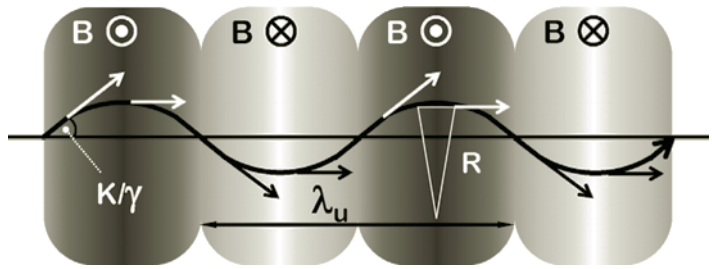



**FIGURE 8.8**
Trajectory and radiation in sine-like field.

The trajectory parametrization can therefore be written as

$$x = \frac{K}{\gamma}\frac{\lambda_u}{2\pi}\sin\left(\frac{2\pi z}{\lambda_u}\right) \quad \text{and} \quad x' = \frac{K}{\gamma}\cos\left(\frac{2\pi z}{\lambda_u}\right) \tag{8.2}$$

This shows us that if $K < 1$, then the trajectory angle is always less than $1/\gamma$ and the external observer will be able to see the emitted fields without interruptions; ultimately, the entire trajectory contributes to radiation.

Let us now connect the maximum field in the undulator

$B_0$ with parameter $K$. The bending radius is connected to the curvature of trajectory via

$$\frac{d^2x}{dz^2} = \frac{1}{R}$$

which gives us

$$K = \frac{\lambda_u \gamma}{2\pi R} \tag{8.3}$$

and by substituting the expression for the bending radius in the magnetic field $R = p\,c/(e\,B_0)$ we obtain

$$K = \frac{\lambda_u e B_0}{2\pi mc^2} \tag{8.4}$$

This finally give us the precise definition of the undulator parameter.

We are now ready to discuss the basics of FEL operation.

## 8.3 Basics of FEL operation

In a third-generation light source, the phase relationship between the radiation emitted by each electron is random and therefore the spatial and temporal coherence of the radiation is poor. Even if an undulator were inserted into the ring, the electrons would emit radiation *incoherently*.

In contrast to the third-generation light sources, operation of a free electron laser relies on *microbunching* of the beam caused by interaction of the radiation with the beam, i.e., the beam interacts with itself via the radiation it emits.

Microbunching in FEL happens primarily at the resonant wavelength determined by the undulator parameter.

Once the electron beam is microbunched, each microbunch emits radiation as a single particle of a large charge, in phase with each other — i.e., *coherently*. Correspondingly, the radiation power and brightness of FEL will scale as $N_e^2$ and not as $N_e$ as in third-generation sources, thus giving an enormous boost in performance.

We will now consider the relevant phenomena — resonance condition, energy exchange and microbunching — step by step.

### 8.3.1 Average longitudinal velocity in an undulator

The average longitudinal velocity in an undulator is an important parameter that determines the resonant wavelength. If the particle moves in a free space, its longitudinal velocity is approximated as

$$v_{z0} = \beta\, c \approx c\left(1 - \frac{1}{2\gamma^2}\right) \tag{8.5}$$

For the sine-like trajectory that we parametrized above, the transverse velocity is given by

$$v_x = \beta c \frac{K}{\gamma} \sin\left(\frac{2\pi z}{\lambda_u}\right) \tag{8.6}$$

In the second-order approximation, the longitudinal velocity can be written as follows

$$v_z \approx \beta c \left(1 - \frac{1}{2}\frac{v_x^2}{\beta^2 c^2}\right) \tag{8.7}$$

and therefore the average longitudinal velocity can be expressed as

$$\langle v_z \rangle \approx c\left(1 - \frac{1}{2\gamma^2}\left(1 + \frac{K^2}{2}\right)\right) \tag{8.8}$$

We can see that, in comparison with free space, there is an additional longitudinal *retardation*, which is due to the transverse velocity in an undulator. This retardation is equal to

$$c\,\frac{K^2}{4\gamma^2} \tag{8.9}$$

and is determined by the undulator parameter.

### 8.3.2 Particle and field energy exchange

Energy exchange between an EM wave and electron depends on the electric field and velocity:

$$\frac{dW}{dt} = e\,\mathbf{E}\cdot\mathbf{v} \tag{8.10}$$

If the electrons have only the longitudinal velocity (depicted in the left image of Fig. 8.9) no energy can be transferred between the electrons and the EM wave, as in this case $e\mathbf{E}\cdot\mathbf{v} = 0$.

On the other hand, an electron beam with sine-like trajectory as in an undulator overlaid with an EM wave (right plot in Fig. 8.9) can exhibit an energy exchange between the EM wave and electrons as in this case

$$e\,\mathbf{E}\cdot\mathbf{v} \neq 0 \tag{8.11}$$

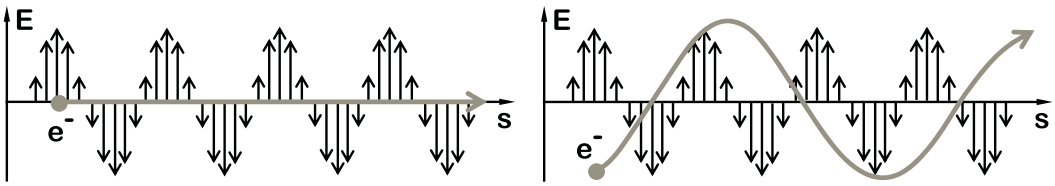



**FIGURE 8.9**
EM wave and particle trajectory — straight (left) and wiggling (right) in an undulator.

because $\mathbf{v}_\perp \neq 0$. Therefore, if electrons have a transverse velocity, energy can be transferred between electrons and EM wave — which is the principle FEL relies upon.

### 8.3.3 Resonance condition

For certain $\lambda$ of an EM wave, a resonant energy transfer between electrons and the EM wave can occur — as illustrated in Fig. 8.10.

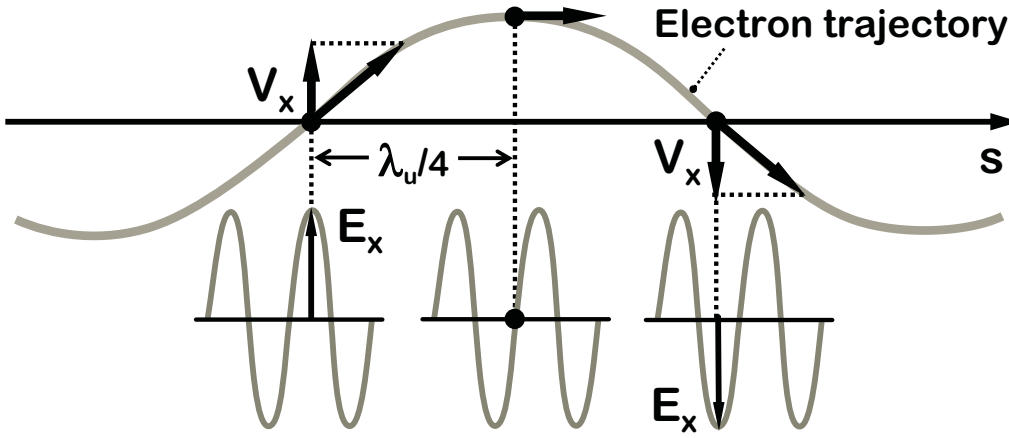



**FIGURE 8.10**
EM wave-particle resonance condition of energy transfer.

The necessary condition for the resonant energy transfer is that the EM wave slips forward with respect to an electron by a $\lambda/2$ per half period of electron trajectory, i.e.,

$$\lambda = \lambda_u (1 - \langle v_z \rangle / c) \tag{8.12}$$

Taking into account the average velocity in an undulator defined by Eq. 8.8, we therefore obtain, for the resonant EM wavelength:

$$\lambda = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2}\right) \tag{8.13}$$

We note that, in the undulator case when $K \ll 1$, the resonance wavelength is very close to the relativistically transformed undulator period $\lambda_u/(2\gamma^2)$.

We should also note that slippage by $3(\lambda/2)$, $5(\lambda/2)$, $7(\lambda/2)$ and so on would also be in resonance, which may result in generation of odd, higher harmonics.

### 8.3.4 Microbunching conceptually

The interaction of particles with the resonant EM wave that we defined in the previous section can create energy modulation in the particle beam.

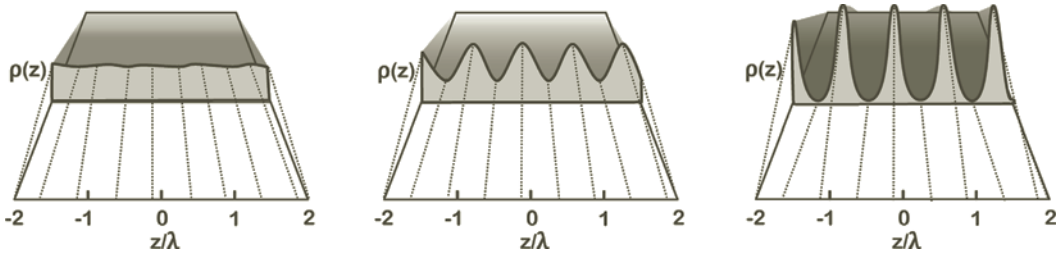



**FIGURE 8.11**
Microbunching. Density of the beam along the longitudinal coordinate for the initial noise (left), intermediate regime of microbunching (middle) and saturated microbunching (right).

As particles move along the curved sine-like trajectory, this energy modulation can result in different routes taken over different trajectories, depending on the particles' energy. Different path lengths can in turn create density modulations along the beam.

An initial EM wave of resonant wavelength can be external (*seeding*) or can emerge from the noise that is always present in the beam. The latter corresponds to the *self amplified spontaneous emission* (*SASE*) process, which is illustrated in Fig. 8.11. Here the corresponding harmonics from the noise of the initial distribution evolve through the linear regime into saturation regime where complete modulation of density eventually occurs. The resulting microbunches emit coherently at wavelength $\lambda$ with radiated power $P \sim N^2$.

## 8.4 FEL types

The two major kinds of FELs are single pass and multi pass. While multi-pass FELs were built primarily in the earlier days of FEL technology, single-pass devices are dominating the arena today.

### 8.4.1 Multi-pass FEL

Multi-pass FEL is similar to a standard laser: mirrors help to build up the optical amplitude on the resonance harmonic while the electron beam trajectory is arranged to pass through the undulator while avoiding mirrors, as shown in Fig. 8.12.

Similar to a normal laser, a single-pass FEL stores the radiation in a cavity. Such systems usually have low gain and therefore many reflections of radiation off of mirrors are needed to create sufficient amplitude. The multi-pass design is suitable in particular for the light sources used in visible or near-visible ranges of radiation.

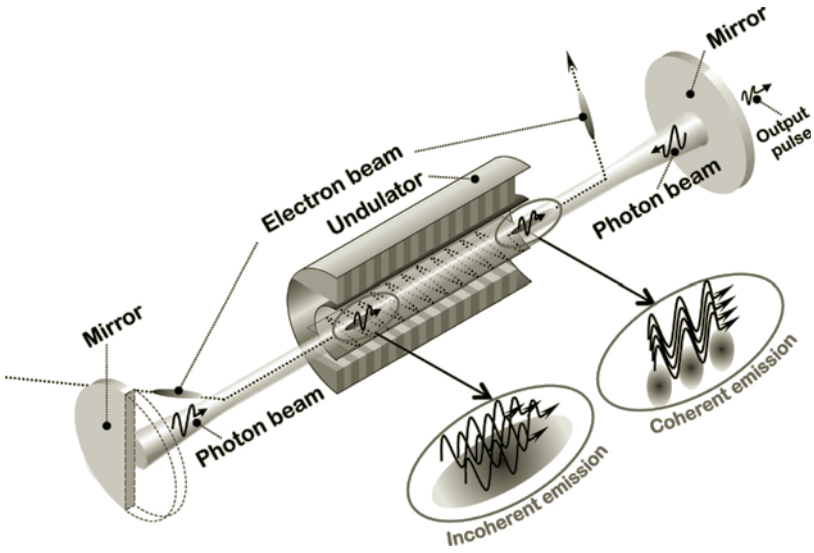



**FIGURE 8.12**
Multi-pass FEL.

### 8.4.2 Single-pass FEL

In a single-pass FEL, the radiation has to grow within a single passage of the beam through the undulator (Fig. 8.13). Such FELs can be either seeded or the SASE type.

The need to use single-pass systems is actually a necessity dictated by an absence of good mirrors in the X-ray spectral region. The single-pass system has to be a high-gain system, which puts extreme constraints on the quality of the electron beam, as well as on the accuracy of the undulator. A typical permanent magnet undulator is shown in Fig. 8.14.

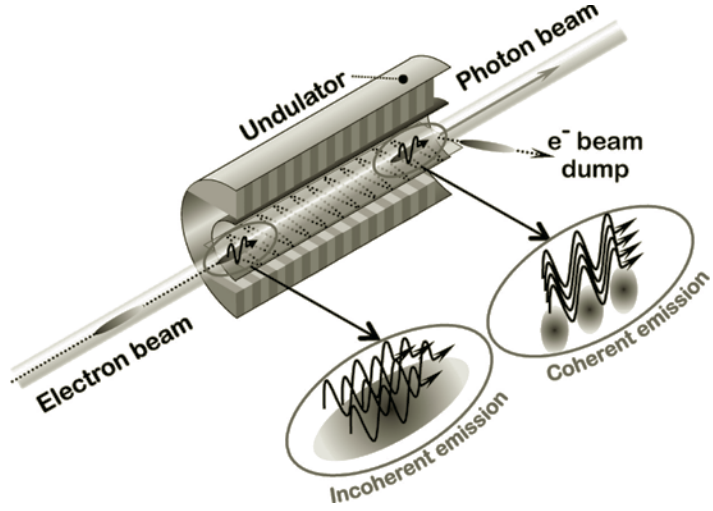



**FIGURE 8.13**
Single-pass FEL.

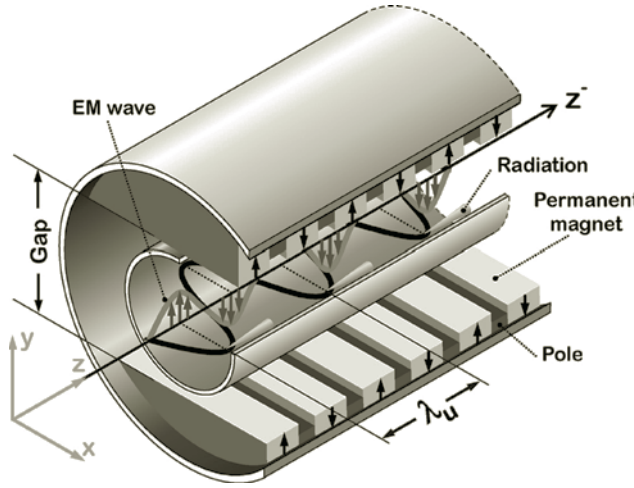



**FIGURE 8.14**
Radiation in an FEL undulator composed of permanent magnets.

## 8.5 Microbunching and gain

The term FEL contains the word *laser*; however, FEL can be explained entirely using the approach of classical electrodynamics. In this section, we will consider microbunching in detail and then discuss the FEL gain factor.

### 8.5.1 Details of microbunching

Interaction of the radiation emitted in an undulator with the electron bunch itself can, in certain conditions, be sufficiently strong to generate a significant modulation of the electrons' energy in the beam.

In this case, the energy change of the particles will occur due to the coupling between the transverse (typically horizontal) oscillation of the electron in the undulator and the transverse (thus also horizontal) component of the electric field of the emitted EM plane wave. The energy change in this case can be written as

$$\frac{d\mathcal{E}}{dt} = e\mathbf{E}\cdot\mathbf{v} = eE_x v_x \tag{8.14}$$

One can also contrast this with the case of acceleration in the RF cavities, where the energy change occurs because of the coupling between the longitudinal velocity of the electron in the RF cavity and the longitudinal component of the electric field in the RF cavity:

$$\frac{d\mathcal{E}}{dt} = e\mathbf{E}\cdot\mathbf{v} = eE_z v_z \tag{8.15}$$

We will now consider the process that leads to mi-

crobunching in detail. The relevant equations of motion are

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E} + \frac{e}{c}\mathbf{v}\times\mathbf{B} \quad \text{and} \quad \frac{d\mathcal{E}}{dt} = e\mathbf{E}\cdot\mathbf{v} \tag{8.16}$$

where momentum and energy are

$$\mathbf{p} = m_e\gamma\mathbf{v} \quad \text{and} \quad \mathcal{E} = m_e c^2\gamma \tag{8.17}$$

and where **B** and **E** are the magnetic field of the undulator and of the undulator radiation.

We will define the undulator magnetic field as

$$\mathbf{B} = B_0 \;\; 0 \;\; , \quad \cos(k_u z), \quad 0 \tag{8.18}$$

where $B_0$ is the amplitude of the undulator fields. And we will define the undulator harmonic radiation as

$$\mathbf{E} = E_0 \;\; \cos\alpha \;\; , \quad 0, \quad 0$$

$$\mathbf{B} = E_0 \;\; 0 \;\; , \quad \cos\alpha, \quad 0 \tag{8.19}$$

where $E_0$ is the amplitude of the radiation harmonic and

$$\alpha = kz - \omega t + \varphi \quad \text{and} \quad \omega = kc \tag{8.20}$$

As you can see, we have simplified the undulator radiation by representing it as a plane wave.

The equations defined above can be integrated. Let's first change the independent variable from $t$ to $z$ and then integrate the equation. We can find that the transverse velocity will be expressed as

$$\beta_x = -\frac{K}{\gamma}\sin k_u z - \frac{eE_0}{m_e\omega c\gamma}\sin\alpha \tag{8.21}$$

where the second term corresponds to modulation of the velocity due to the plane EM wave, which represents the radiation.

The energy change expressed in terms of relativistic factors can be expressed as

$$\dot{\gamma} = -\frac{eE_0}{m_e c\gamma}\cos\alpha\cdot\left[\frac{eE_0}{m_e\omega c}\sin\alpha + K\sin k_u z\right] \tag{8.22}$$

These two equations, Eq. 8.21 and Eq. 8.22, constitute a system of first-order differential equations for $(z,\gamma)$. We will solve these equations in the following approximations. We will assume that the radiation amplitude is small and will therefore keep only the terms with the first order in $E_0$. We will also assume that the gain is small, i.e., the energy change of the electrons is low ($\Delta\gamma \ll \gamma$). And finally, we will assume that the radiation wavelength is very close to the fundamental undulator radiation wavelength and that we will average

all quantities over one undulator period in order to remove fast oscillations.

We begin by introducing the following variable:

$$\zeta = k_u z + \alpha = (k + k_u)z - \omega t + \varphi \tag{8.23}$$

The above-defined system of the first-order differential equations can be transformed into a single differential equation of second order:

$$\ddot{\zeta} = -\frac{eE_0(k_u + k)[J_0(\xi) - J_1(\xi)](1 + K^2/2)K}{2m_e\gamma^4}\sin\zeta \tag{8.24}$$

where $\xi = K^2/(4 + 2K^2)$ and $J_0$ and $J_1$ are Bessel functions.

Eq. 8.24 (above) can be expressed as

$$\ddot{\zeta} + \Omega^2 \sin\zeta = 0 \tag{8.25}$$

which is the so-called FEL-pendulum equation, which describes the interaction of particles with radiation in an FEL.

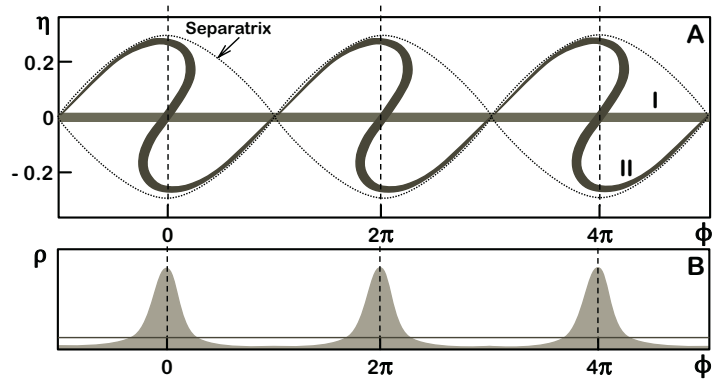



**FIGURE 8.15**
Illustrating solutions of FEL-pendulum equation and microbunching for different initial conditions. The initial beam (I) is on-energy and when bunched (II) demonstrates symmetrical profile of beam density (B).

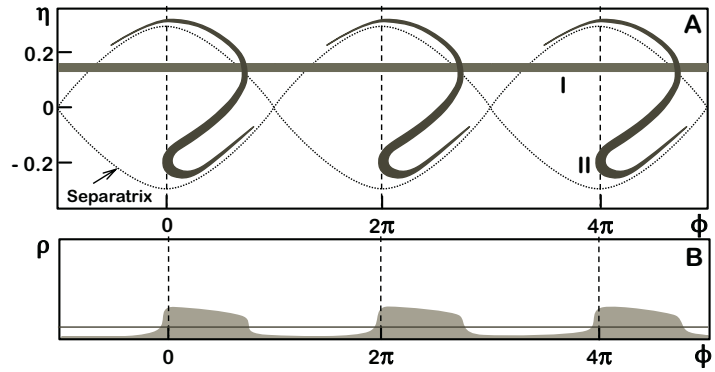



**FIGURE 8.16**
Microbunching in a case when the initial beam is slightly off energy.

Take note that the equations Eq. 8.24 or Eq. 8.25 describe the familiar motion in a separatrix where the period of oscillations depends on initial conditions, particularly whether the particle is infinitely close to the separatrix, in which case the period becomes infinitely long. Examples of solutions to this equation for different initial conditions are shown in Fig. 8.15 and Fig. 8.16.

In the EM wave-particle interaction process described by the above equations, each electron will gain or lose energy depending on the relative phase $\zeta(0)$ between the transverse oscillation in the undulator and the phase of the radiation plane wave. The energy change (ignoring the small terms of the order of $O(\Omega^2)$) can be written as

$$\Delta\gamma = -\frac{eE_0K\,[J_0(\xi)-J_1(\xi)]\,L}{2m_ec^2\beta_{z0}\gamma_0}\frac{\sin(\nu/2)}{\nu/2}\sin(\zeta(0)+\nu/2) \quad (8.26)$$

where $L$ is the length of the undulator and

$$\nu = \left(k + k_u - \frac{\omega}{c\beta_0}\right)L \quad (8.27)$$

By averaging the above equation over the initial phases $\zeta(0)$ of the electrons, we obtain the average energy variation:

$$\langle\Delta\gamma\rangle_\varphi = \frac{eE_0K\,[J_0(\xi)-J_1(\xi)]\,\Omega^2}{8m_ec\gamma_0}\left(\frac{L}{c\beta_{z0}}\right)^3\frac{d}{d\nu}\left(\frac{\sin\nu/2}{\nu/2}\right)^2 \quad (8.28)$$

As there is a balance and conservation of energy, the variation of the electron energy is equivalent to a variation of the generated EM wave's energy in the FEL. This allows us to define the FEL gain.

### 8.5.2 FEL low-gain curve

The FEL gain $G$ can be defined as a relative change of the wave's energy, equaling to the change of the energy of all electrons involved in the interaction:

$$G = \frac{\Delta E_{tot}}{W_0^L} = -m_ec^2\frac{N}{W_0^L} < \Delta\gamma >_\varphi \quad (8.29)$$

where $W_0^L$ is the initial energy of the wave over the entire length of the undulator.

For a bunch with the peak current $I$ and transverse area $\Sigma_b = F\Sigma_L$ the gain can be written as

$$G = -\frac{\pi K^2[J_0(\xi)-J_1(\xi)]^2k_uL^3(1+\beta_{z0})}{2\gamma^3\beta_{z0}^3}\frac{F}{\Sigma_L}\frac{I}{I_0}\frac{d}{d\nu}\left(\frac{\sin\nu/2}{\nu/2}\right)^2 \quad (8.30)$$

The corresponding low-signal and low-gain FEL curve is shown in Fig. 8.17.

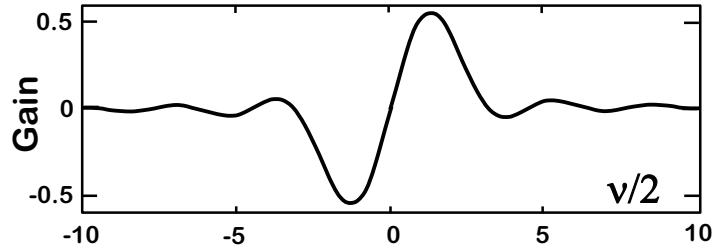



**FIGURE 8.17**
FEL low-gain curve.

The positive gain in Fig. 8.17 corresponds to an amplification of the EM wave (a standard FEL case) while the negative gain corresponds to an acceleration of the beam and decrease of the EM wave's amplitude. The latter case relates to the so-called *inverse FEL*.

### 8.5.3 High-gain FELs

Most modern FELs, especially those aimed at hard X-rays, are high-gain systems. In these FELs, the gain is so large that the EM wave amplitude changes within a single pass in the undulator, and therefore our estimations determined in the previous section need to be revised.

A detailed analysis — which takes into account the wave equation with driving terms determined by the oscillating current density of the beam — predicts an exponential growth of the radiation power of

$$P(s) \propto e^{s/L_g} \tag{8.31}$$

with the exponent given by the following

$$L_g = \frac{1}{\sqrt{3}}\left(\frac{4\gamma^3 m_e}{\mu_0 K^2 e^2 k_u n_e}\right)^{1/3} \tag{8.32}$$

The exponential growth continues until saturation is reached, at which point the emitted power starts to oscillate as illustrated in Fig. 8.18.

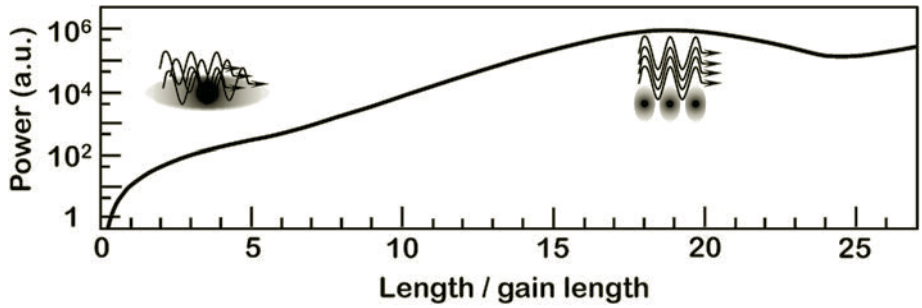



**FIGURE 8.18**
High-gain FELs, typical behavior of the emitted power — exponential growth eventually turned into saturation.

Analysis of a high-gain curve such as the one shown in Fig. 8.18 helps us to determine an optimal length of the undulator for a particular FEL design.

## 8.6 FEL designs and properties

We will now review typical accelerator parameters, requirements and radiation characteristics of modern FELs.

### 8.6.1 FEL beam emittance requirements

As we have determined in Chapter 3, the emittance of a synchrotron radiation photon beam is given by Eq. 3.33 and is equal to

$$\varepsilon_{ph} = \frac{\lambda}{4\pi}$$

It is intuitive to assume that efficient generation of FEL radiation (the term, *lasing*, is often used) requires a fair amount of overlap between the electron and the photon beam.

The geometrical emittance of an electron beam needed for efficient lasing, therefore, needs to be smaller than the one for the photon beam. Thus,

$$\varepsilon \leq \frac{\lambda}{4\pi} \tag{8.33}$$

or in terms of the normalized emittance:

$$\varepsilon_N \leq \gamma \frac{\lambda}{4\pi} \tag{8.34}$$

As an example, for $\lambda = 0.2$ nm and $\gamma = 3 \cdot 10^4$ ($\approx 15$ GeV), the required normalized emittance is $\leq 0.5$ mm·mrad, which results in a necessity to use a very bright electron source. We can also note that the requirement for geometrical emittance can be eased for higher energy electron beams since, during acceleration, the geometrical emittance decreases in inverse proportion to the beam energy.

An important clarification of the above emittance requirement relates to the concept of *slice emittance*. As we discussed above, radiation slips with respect to the electron beam by $\lambda$ for every $\lambda_u$. Even if the undulator is 100 m long, for any reasonable undulator period (e.g., $\lambda_u \approx 1$ cm) and wavelength (assume $\lambda \approx 0.1$ nm), the total slippage will be around 1 $\mu$m. A typical electron bunch is usually much longer. Therefore, only a small longitudinal fraction of the bunch contributes to a particular spatial portion of generated radiation. The requirement for the emittance defined above in Eq. 8.33 is thus applicable to the slice that is generating this portion of the radiation.

### 8.6.2 FEL and laser comparison

It can be useful to compare the properties and features of FELs and lasers side by side — see Table 8.1. Amongst many similarities, the most notable difference is in their theoretical foundations: while the laser is a purely quantum device, an FEL is entirely classical.

**TABLE 8.1** FEL and laser comparison

| | **LASER** | **FEL** |
|---|---|---|
| Characteristics | Source of narrow, monochromatic and coherent light beams | Source of narrow, monochromatic and coherent light beams |
| Configurations | Oscillator or amplifier | Oscillator or amplifier |
| First demonstration | 1960 | 1977 |
| Laser media | Solids, liquids, gases | Vacuum with electron beam in periodic magnetic field |
| Energy storage | Potential energy of electrons | Kinetic energy of electrons |
| Energy pump | Light or applied electric current | Electron accelerator |
| Theoretical basis | Quantum mechanics | Relativistic mechanics and electrodynamics |
| Wavelength definition | Energy levels of laser medium | Electron energy, magnetic field strength and period |

### 8.6.3 FEL radiation properties

Due to the coherent nature of radiation, FELs can provide peak brilliance of around 8–10 orders of magnitudes larger than that of storage ring light sources (see Fig. 8.19).

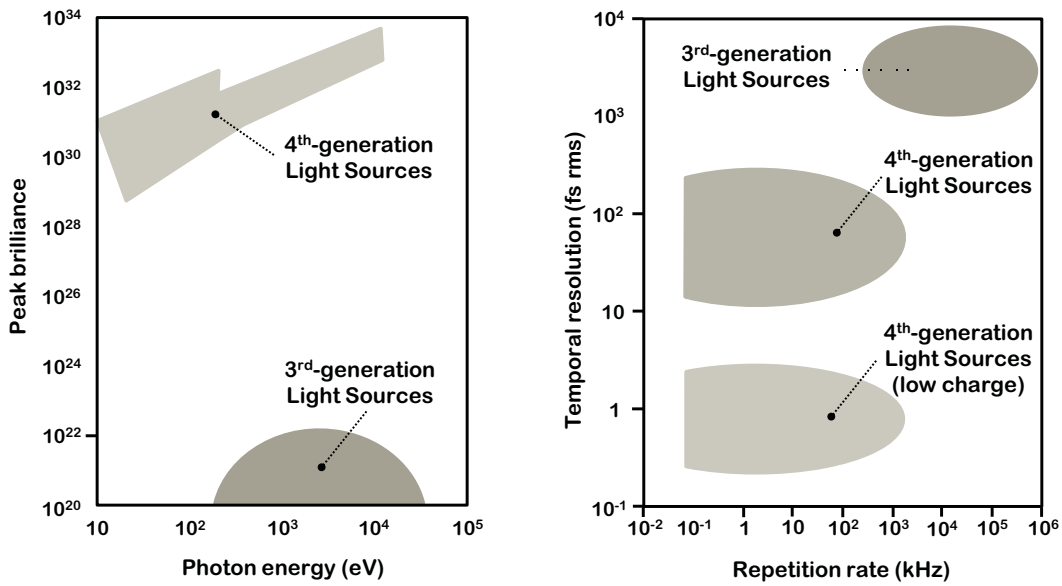



**FIGURE 8.19**
Peak brilliance (left) and temporal resolution (right) of typical FEL in comparison with third-generation SR sources.

Since the repetition rates of the typical FELs is still below that of storage rings, the average brilliance of FELs exceeds that of third-generation light sources typically by about 2–4 orders of magnitude.

The FELs are also unsurpassed in temporal resolution, exceeding the third-generation by 2–4 orders of magnitude, able to reach femtosecond resolutions in special configurations of low-charge sliced beams (the methods of ultra-short bunch generation are discussed in more detail in Chapter 10).

### 8.6.4 Typical FEL design and accelerator challenges

Modern high-brightness FELs aimed at hard X-ray ranges are sophisticated machines. A typical layout of such an FEL is shown in Fig. 5.39. A high-brightness laser-driven photo RF gun, operating from a kHz to potentially MHz range, can power a linear accelerator, a normal conducting (NC) or a super conducting (SC). Typically, two bunch compressors, at intermediate and final energies, are used to achieve the shortest possible beam length. The accelerated and compressed electron beam is sent into an undulator, or switched between different undulators.

Practical constraints (size, cost) split the FEL designs into two families — NC and SC FELs, which operate under two distinct sets of parameters. Normal conducting FELs can have the highest electron beam energy and therefore the shortest X-ray wavelength; however, they are pulsed and usually have rather low (sub-kHz) repetition rates.

Superconducting FELs, on the other hand, can be either pulsed (but with long pulses and many bunches in a single pulse) or are continuously operating (CW) with repetition rates of bunches up to MHz. They, however, are generally limited in electron energy and cannot reach the shortest wavelength (see Table 8.2).

**TABLE 8.2**
NC and SC FEL, typical parameters

| Characteristics | NC-FEL | SC-FEL |
|---|---|---|
| Linac energy | 6–15 GeV | 2–4 GeV |
| Linac frequency | S, C, X band | L band |
| Repetition frequency | 50–200 Hz | kHz–MHz |
| Operation mode | pulsed | long pulse to CW |
| Minimum wavelength | 0.1 nm | 0.3–1 nm |

Accelerator physics challenges in FELs start from the source — a very low emittance gun is needed, seeing as the normalized emittance cannot be improved in the linac. Acceleration and compression through the linac needs to be performed to keep the emittance low while managing the collective effects of high-brightness beams (which is not trivial,

particularly at a low energy). Compression of the bunch to a tens of fs range is challenged by *coherent synchrotron radiation* (CSR), which will be discussed in closer detail in Chapter 10.

Another group of challenges related to FELs is simultaneously connected to both the physics and technologies of accelerators and lasers. Firstly, this refers to the synchronization of conventional lasers and FEL electron sources for pump-probe experiments. And secondly, conventional lasers are essential to applications in FELs for the seeding of radiation in order to ensure improved temporal coherence of generated X-rays, as well as for *high harmonic generation* (HHG) (discussed further in Chapter 10).

## 8.7 Beyond the fourth-generation light sources

Let us summarise the properties of third- and fourth-generation light sources and outline prospects for future generations.

The third generation (storage rings) and FEL have complementary properties. The storage ring synchrotron light sources are extremely stable, can serve several tens of user beamlines, and their radiation can approach full transverse coherence for ultimate diffraction-limited storage rings.

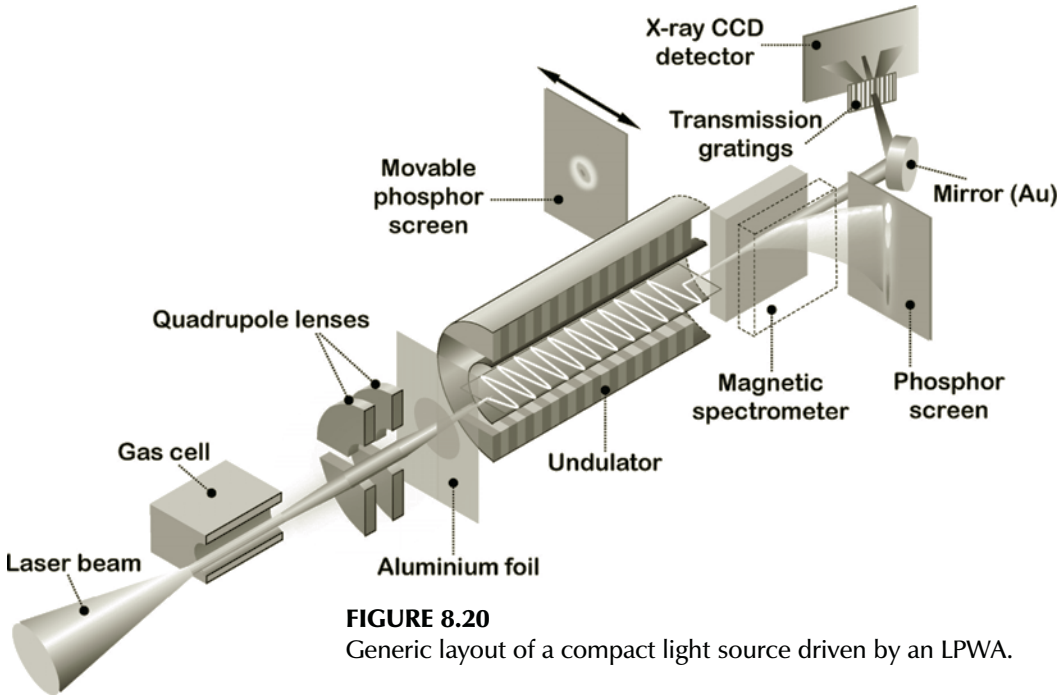



**FIGURE 8.20**
Generic layout of a compact light source driven by an LPWA.

The fourth-generation (FELs) sources have high brightness, short pulses and full transverse coherence. However, FELs can serve only a few beamlines at a time. Hard X-ray

high brightness FELs are machines that require considerable resources for their construction and can usually be realized only as national-scale facilities.

New technological solutions are required if we are to build more economical and compact radiation sources.

Progress with laser plasma accelerators in the last several years has opened up new possibilities to work towards compact betatron radiation sources and has also created an aspiration in the community to use this technology to drive compact FELs.

The first generation of undulator radiation in the soft X-ray range driven by an LWFA-produced beam has already been obtained. In this first experiment, a laser plasma wakefield accelerator produced 55–75 MeV electron bunches, which were then sent to an undulator to generate visible to IR-range synchrotron radiation.[1]

A generic layout of a compact light source based on laser plasma acceleration and an undulator is shown in Fig. 8.20. Such compact sources are often referred to as *table-top* although at present time a more appropriate term would be "room-sized."

Laser plasma wakefield accelerators demonstrated the possibility of generating a GeV beam with promising electron beam qualities, including a normalized emittance of the order of 1 mm mrad and an energy spread of close to 1% for the entire bunch.

Recall that lasing in FEL requires the beam slice to have appropriately small emittances; these characteristics are already within reach for laser plasma acceleration and will likely be achieved with relatively modest improvements on what has presently been obtained. The beam energy spread will, however, require more noticeable improvements from the presently achieved values of a few percent for the entire beam to around a few hundredth of a percent for a radiation-generating slice.

Furthermore, an FEL based on laser plasma acceleration will require significant improvement to the stability of LPWA beams; at present, the repeatability of its beam parameters is poor and important characteristics of beams can often exhibit nearly 100% pulse-to-pulse fluctuations.

We should also note that an election beam generated in a plasma bubble typically has inconvenient ratios between its size and its angular spread. While the beam size may be of the order of a micron, the angular spread can reach several milliradians. Even if the calculated emittance of such a beam is very small, its usefulness strongly depends on the ability to quickly capture this beam into an appropriately designed focusing channel, in order to avoid effective emittance growth due to filamentation — nonlinear wrapping-around in the

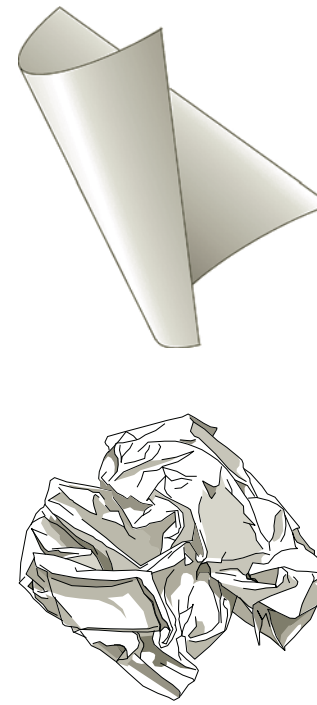


**FIGURE 8.21**
For illustration of filamentation. An intact paper sheet (top) may have very low volume; however, when crumpled (bottom) it will have its effective volume increased by orders of magnitude.

[1] Schlenvoigt et al., Nature Phys. 4, 130 (2008).

beam phase space as illustrated (in a *synectics* approach of using analogies) by a paper sheet that increases its effective volume by orders of magnitude after being crumpled — see Fig. 8.21.

Preservation of the emittance of an LPWA beam during its transport from the plasma bubble into the undulator would require properly catching the accelerated beam into a focusing system. The focusing distances of the first lenses of the catching optics need to be comparable to the focusing in the plasma bubble, and the focusing distances of further lenses should gradually increase to match the focusing strength of the undulator beamline. An alternative approach would be to create an undulator inside of the plasma — a concept which, if proven feasible, may create light sources of ultimate compactness.

**EXERCISES**

8.1 *Chapter materials review*.
Evaluate the slice emittance requirements for an FEL based on 1 GeV electron beam. Discuss the factors affecting the requirements for the slice energy spread of this beam in this FEL.

8.2 *Chapter materials review*.
Discuss the phenomenon of filamentation in plasma acceleration in connection to the TRIZ inventive principle of preliminary action.

8.3 *Mini-project*.
Define, very approximately, the main parameters (energy, length, undulator field and step size) of a linac-based FEL aimed at 15 KeV X-ray energy.

8.4 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this chapter in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

8.5 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 9 Proton and Ion Laser Plasma Acceleration



In Chapter 6, we primarily discussed the plasma acceleration of electrons. There are strong motivations, however, for development of plasma acceleration of protons and ions, as further advances in this area could improve current ways of treating tumors.

The advantages of using protons and ions (in comparison to using electrons and X-rays) are associated with the phenomenon of *Bragg peak* — we will start our discussion by looking at this phenomenon in detail.

Regarding the application of accelerators for therapy, it is important to know how either the electrons or protons can affect the living cells. We will therefore briefly overview, in this chapter, how the mechanisms of DNA respond to radiation.

The conventional accelerator systems used for particle beam therapy for tumors typically require protons with around 250 MeV of energy. The accelerator is typically a synchrotron or a cyclotron (so far the most popular). A beam therapy facility usually includes beam delivery gantries (which send the beams in a selected target volume), as well as collimators and degraders. We will review the design and functionality of conventional beam-therapy facilities in a later section.

A conventional accelerator system for particle beam therapy, with a beam source and several gantries, may require several thousands of square meters of space, making these respective facilities large and expensive to maintain. On the other hand, plasma acceleration of protons and ions is one of the most promising means by which we can make beam therapy more affordable and more accessible to patients. After we discuss the motivations to apply plasma acceleration to this area, we will briefly overview the present understanding of the different regimes and mechanisms of proton/ion plasma acceleration.

We would like to once again note that proton/ion laser plasma acceleration is a rapidly developing area and, moreover, it involves rather complicated physics. There are various regimes of acceleration identified and several mechanisms that explain the experimental results. The different mechanisms of acceleration compete, and often more than one mechanism acts at the same time. The mechanisms of acceleration are often just approximate models that represent a gradual improvement of our understanding of proton/ion

DOI: 10.1201/b18696-9

plasma acceleration. The state of this rapidly developing area makes it hardly suitable for textbooks just yet. We will therefore grant only a cursory look at an overview of the presently identified mechanisms and their scaling rules. We will conclude this chapter with a glimpse into the future.

## 9.1 Bragg peak

The treatment of tumors with the help of X-rays or proton/ion beams is based on the delivery of energy to malignant cells of the target volume, intended to prevent and eradicate the growth of these unwanted cells. We will review the biological effects of radiation in the next section, and will now compare the effects of the photons with the effects of charged particles.

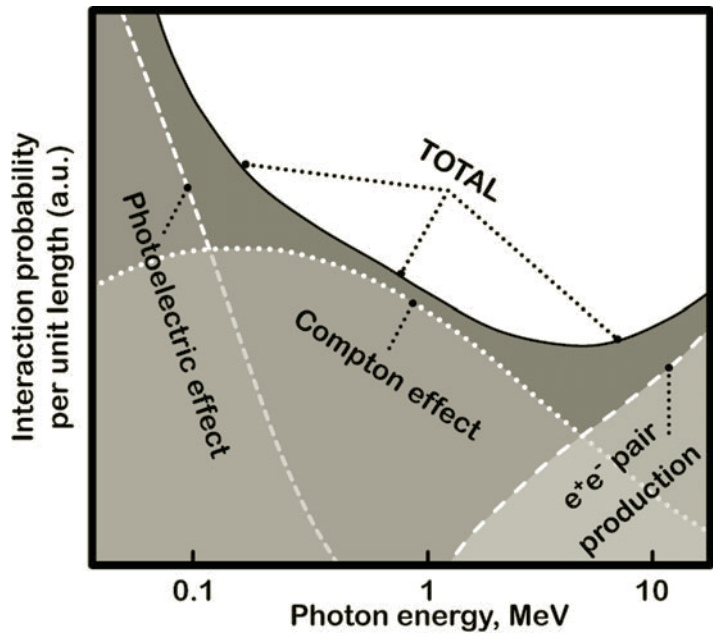



**FIGURE 9.1**
Photon matter interaction, qualitatively.

Photons penetrating through a medium will lose their energy due to several factors. An incident photon can interact with an atom and get absorbed, causing a *photoelectron* to be ejected from the atom. The effect is dominant at lower energies (see Fig. 9.1). An incident photon can also lose part of its energy via the Compton effect — the inelastic collision of photons with the electrons of the atoms. Finally, if the energy of the photon is sufficient, it can create an $e^+e^-$ pair. This effect has a threshold character and manifests itself for photons with energies greater than about an MeV.

The photons that are used in radiation therapy are usually produced from electron beams of several MeV, converted on a solid target to photons via the process of *bremsstrahlung*. As interaction between the photon beam and medium results in a gradual decrease of intensity with the depth of the absorbing medium, the so-called *stopping power* $S(E)$ is gradually decreasing too, as shown in Fig. 9.2. The stopping power is defined as $-dE/dx$ — the energy loss per unit length taken with the minus sign.

In contrast to photons, charged particles lose their energy in matter primarily through a Coulomb interaction with the outer-shell electrons of the absorber's atoms. Excitation and ionization of atoms result in a gradual slowdown of the particle. A slower moving particle will interact with an atom for longer time, resulting in a larger energy transfer. Therefore, the charged particles will have an increased energy loss per unit length at the end of their passage through the medium. Qualitatively, this is the origin of the *Bragg peak* — the peak in energy loss that occurs just before the particles come to a complete stop.

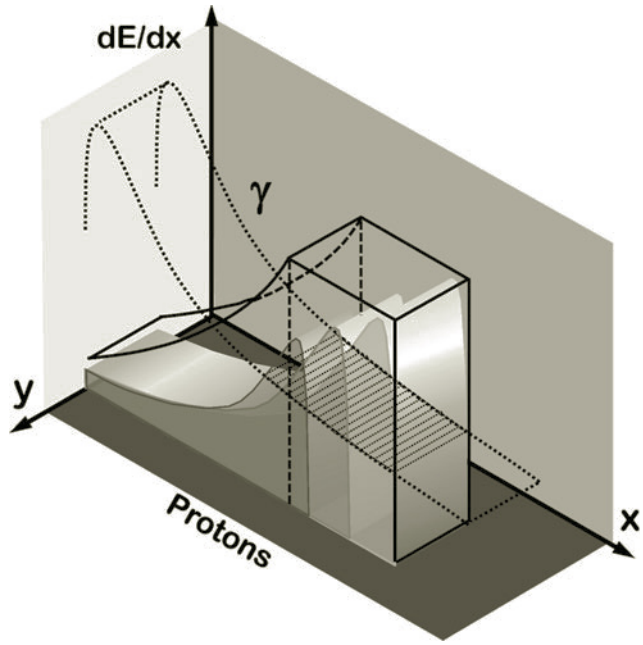



**FIGURE 9.2**
Absorption of photons (dotted lines) in comparison with absorption of protons in media. Overlaying multiple Bragg peaks creates a near uniform dose distribution in a certain target volume.

The above-mentioned phenomenon is named after William Bragg, who discovered it in 1903. The Bragg peak is the reason why protons are better for treatment of tumors than X-rays, in some cases, as protons localize the deposited dose in

the destined target volume (see Fig. 9.2), this minimizes the impact on healthy tissues — especially in cases where the target volume is located close to critical organs.

Quantitatively, the Bragg peak can be explained by the following formula for the mean energy loss of moderately relativistic heavy particles (*Bethe equation*):

$$\langle -\frac{dE}{dx} \rangle \approx K z^2 \frac{Z}{A} \frac{1}{\beta^2} \left( \frac{1}{2} \ln \frac{2 m_e c^2 \beta^2 \gamma^2 W_{max}}{I^2} - \beta^2 \right) \qquad (9.1)$$

Here, $z$ is the charge number of an incident particle, $Z$ and $A$ are the charge number and atomic mass of the absorber, respectively, and $\beta$ and $\gamma$ are the relativistic factors of the incident particle. The parameters under the logarithm are: $I$ — the mean excitation energy of the atom's electron, and $W_{max}$ — the maximum energy transfer in a single collision. For a particle with mass $M$, the latter is defined as $W_{max} = 2 m_e c^2 \beta^2 \gamma^2 / (1 + 2\gamma m_e / M + (m_e/M)^2)$. The coefficient $K$ is defined as follows: $K = 4\pi N_A r_e^2 m_e c^2$ where $N_A$ is Avogadro's number. The coefficient $K$ approximately equals $0.3\ MeV \cdot cm^2/mol$.

*Penetration ranges for protons in water:*
*250 MeV — 38 cm;*
*200 MeV — 26 cm;*
*150 MeV — 15.6 cm;*
*100 MeV — 7.6 cm;*
*50 MeV — 2.2 cm.*

The usefulness of the Bragg peak for treating tumors was first realized by Robert R. Wilson in 1946. Overlaying several Bragg's peaks described by Eq. 9.1 creates a uniform dose distribution in a given volume (as illustrated in Fig. 9.2) — this is often called a *spreadout Bragg peak*. Such overlaying requires an adjustment to the energy and intensity of each individual proton beam.

The ideally sharp Bragg peak is, in practice, somewhat spread — firstly due to the statistical character of interaction, and secondly due to nuclear interactions between the protons and absorber, which happen with some probability.

The sharpness of the Bragg peak is an enabling feature of proton therapy, but simultaneously it is a factor that increases the sensitivity of the method to errors, especially to the errors in the predicted depth range. In a particular case when the target volume is located near a critical organ, ideally, one can completely eliminate irradiation of the critical organ while filling the entire target volume uniformly. In practice, however, one cannot obtain a sharp irradiation boundary of the irradiated volume, due to the necessity to allow for some uncertainties of the depth range. Possible motion of the critical organs during irradiation — as well as shrinkage of the tumor (and possibly corresponding shift of the critical organs) as the treatment progresses — are also important factors, which need to be taken into account in proton therapy planning.

The above-mentioned sensitivity to errors places a particularly strong requirement on the energy of protons in the cases when plasma acceleration is used. The beam needs to have a well-defined energy. This can be ensured either via predictable plasma acceleration or by an appropriate energy-selection system.

## 9.2 DNA response to radiation

The underlying mechanism of X-ray and proton cancer therapy is the ability to prevent the replication of malignant cancer cells. This control is achieved by damaging the *DNA* of cancerous cells. Let's now consider the basic facts about DNA's response to radiation.

The effect of ionizing radiation on DNA can produce many outcomes, from neutral to negative (stimulating growth of cancer cells), or positive (eliminating cancer cells) depending on the initial state of the irradiated object, the radiation dose and the use of *concomitant agents* — pharmaceutical substances used prior or together with irradiation.

In some cases, radiation simply does not cause DNA damage, but passes through the cell without any side effects. With an increased dose, the radiation starts to physically affect the DNA. The effects are distinguished as direct and indirect.

Direct radiation effects take place when X-rays create ions and corresponding electrons (e-) that physically break the nucleotide pairs of the DNA (see Fig. 9.3).

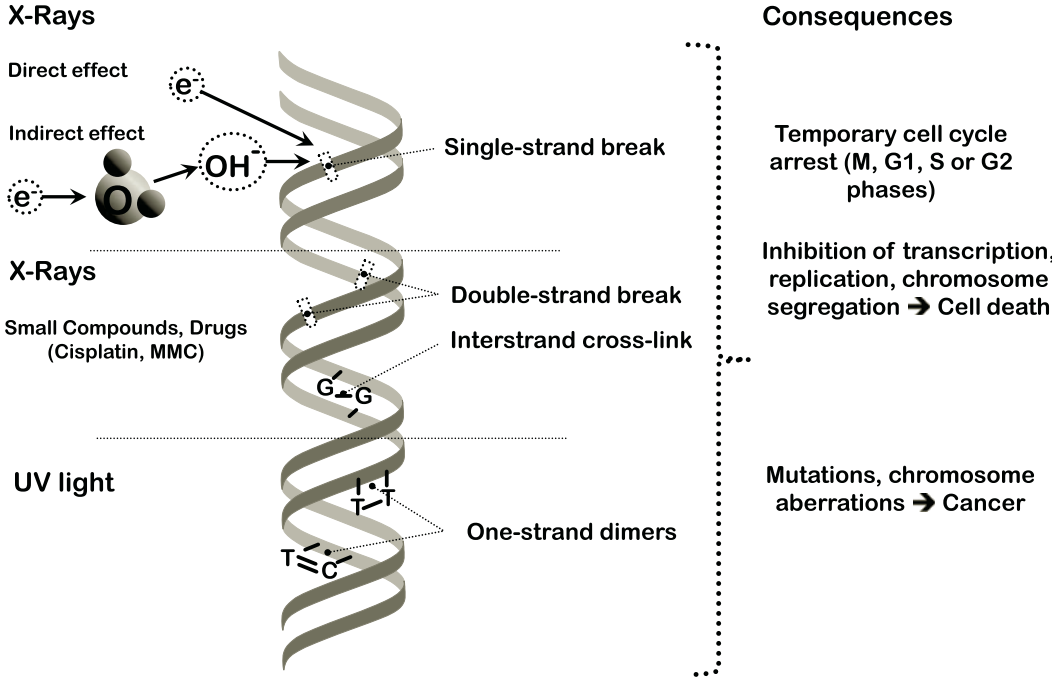



**FIGURE 9.3**
Radiation effects on DNA.

Nevertheless, in the majority of cases, X-rays act indirectly when they induce water radiolysis and consequent production of OH- hydroxide — highly active free radicals — that

then form hydrogen peroxide and cause single-strand DNA breaks.

We should also note that the effects described above are also very similar to the case of irradiation by protons, taking into account that, in this instance, the electrons and *bremsstrahlung* photons are produced when the protons lose their energy to electrons of the absorbing medium.

Several minor DNA defects — such as one-strand dimer formations (caused by UV light — see bottom part of Fig. 9.3) or single-strand breaks could be corrected by the internal cell mechanism, *base excision repair*, which involves endonucleases and DNA polymerase. Nevertheless, some errors can also appear at this step (for example, wrong nucleotide(s) insertion could cause a future coding mismatch).

*Double-strand breaks represent the most detrimental damage produced in DNA by ionizing radiation.*

Double-strand breaks are more difficult to repair and thus represent the most detrimental damage produced in DNA by ionizing radiation. When a cluster of complex DNA damage events occur together, for example from dense ionization at the terminal end of an energetic electron track, there is the highest probability of either mis-repair or failure to repair damage. A loss of entire parts of the chromosomes containing tumor-suppressor genes — an amplification of oncogenic-potential regions — could lead to carcinogenesis, the creation of tumor cells. Unrepairable damage can trigger the *apoptosis mechanism*, where defective cells undergo "cell suicide" and are, thus, eliminated. Mis-repair of DNA with cross-linked or mutated chromosomes may not be immediately lethal, but will lead to further DNA damage or cell death when the cell next attempts division — *mitotic catastrophe*.

Various examples of the consequences of DNA irradiation are shown on the right side of Fig. 9.3. They include temporary arrest at different stages of the cell cycle (mitosis (M), both check-points (G1 and G2), or DNA replication/synthesis (S)). The irradiation effects also may include inhibition of transcription and a variety of cell division abnormalities, all of which could result in a defective cell death.

*In molecular biology, a small molecule drug (or a small compound) is a low molecular weight organic compound that may help to regulate biological processes.*

The positive effects of ionizing radiation are caused by the same mechanisms, since what could be detrimental for a normal cell could also be destructive for a malignant one.

In order to further enhance the destructive effect of X-rays on malignant cells, some concomitant agents such as *small molecule oncology drugs* could be involved. The upper limit for a small molecule's weight is approximately 900 daltons, which allows rapid diffusion across cell membranes so that they can reach intracellular sites of action.

*Dalton (Da) is the standard unit defined as one twelfth of the mass of an unbound neutral atom of $C_{12}$ in its nuclear and electronic ground state. In other words, one Da is approximately equal to a mass of a proton or neutron.*

The effects of irradiation and the use of concomitant agents can lead to an increased number of difficult-to-repair inter-strand cross links, as shown in Fig. 9.3. The apoptosis mechanism, launched by the strong synergetic effect of these two techniques, leads to a more efficient elimination of malignant cells.

## 9.3 Conventional proton therapy facilities

The design of conventional proton therapy facilities is dictated by several factors, such as the size of the proton source, the number of treatment beamlines and the size of the gantry.

A generic layout of a proton therapy facility is shown in Fig. 9.4. Typically, either a cyclotron or a synchrotron is used as a source for a proton facility. In the event that ions are used for therapy (e.g., carbon, which gives certain advantages), a cyclotron can be used to inject the beam into a synchrotron.

The beamline distribution system then directs the accelerated beam into the treatment rooms. The beam can have a fixed location in the room, or can be brought into the room via a *gantry*, which allows flexibility in the direction of the arriving beam.

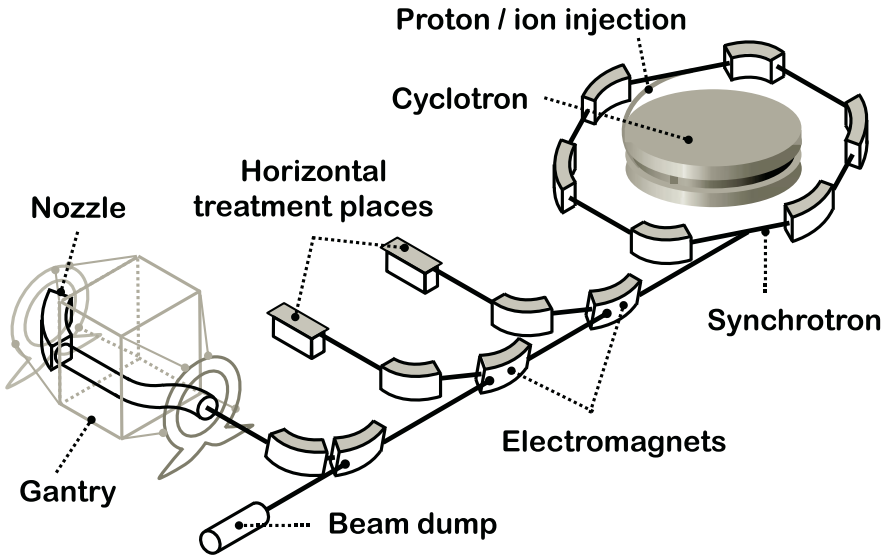



**FIGURE 9.4**
Generic proton or heavy-ion therapy facility.

A typical facility would usually have several treatment rooms with fixed beams and one or more rooms with gantries. The gantry is a complicated mechanical device that needs to rotate the entire section of a beamline while maintaining its precise (usually sub-mm) alignment. The energy of the protons or ions and the achievable strength of the bending magnets of the gantries would typically make the gantry on the order of ten meters in size and a hundred of tons in weight. The use of superconducting magnets in the gantries is possible and would typically reduce the weight of the gantry by around a factor of four.

The overall size of the proton/ion treatment facilities is defined by the size of the accelerator and beamlines, but, most notably, by the size and the height of the gantries. The

cost of the components and the size of the necessary building are the prevailing limiting factors — there are presently several tens of facilities operating worldwide, while the generally accepted estimation states that there should be one proton treatment facility for every 10 million people.

### 9.3.1 Beam generation and handling at proton facilities

The typical elements of the proton therapy system are illustrated in Fig. 9.5.

First, the proton beam coming from the source passes through a scatterer, which increases the divergence of the beam. The transverse beam shape is then adjusted by a multi-leaf steel collimator to match the shape of the target volume. The leaf of the collimator can be mechanically controlled on a sufficiently short time scale. Instead of an adjustable multi-leaf collimator, a fixed-shape brass collimator can be used, tailored for a particular case.

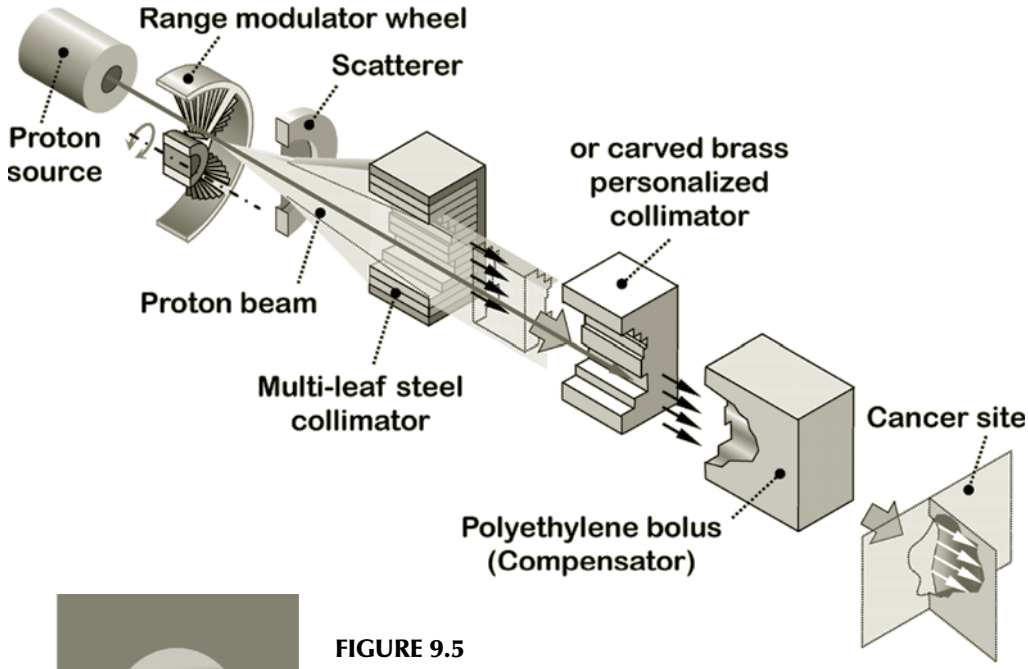



**FIGURE 9.5**
The elements of the proton therapy beamline.

**FIGURE 9.6**
Pencil beam scanning.

Another element of the system is called the bolus, or range-shifter, and serves to compensate for the depth difference of different regions of the target volume. The compensator is usually made from polyethylene and reduces the energy of the beam to fit the required penetration depth at each point.

The elements as described above are, in particular, suitable for the beam coming from standard cyclotrons, the en-

ergy of which usually cannot be changed easily. In the case of synchrotrons, where the beam energy can be finely scanned, the *pencil beam scanning* approach can be employed, as shown in Fig. 9.6. In this case, the beam should be small enough so that one could "paint" the desired target volume transversely, and simultaneously control the depth by appropriately adjusting the beam energy.

### 9.3.2 Beam injectors in proton facilities

So far, the most widely used accelerator for proton therapy facilities has been a cyclotron.

A standard cyclotron is shown in Fig. 9.7. A constant magnetic field (often arranged to decrease with the radius to ensure transverse focusing) houses the electrodes (dees) where oscillating voltage is applied.

The standard cyclotron is intended for CW operation, and this is in fact one of its main advantages. The frequency of accelerating voltage $\omega_0$ and the magnetic field $B_0$ are constant and are connected via the equation for the time of flight around the orbit

$$\omega_0 = \frac{v}{R} = \frac{qB_0}{m\gamma} \tag{9.2}$$

where $m$ is the mass at rest and $\gamma$ is the relativistic factor that can change during acceleration.

Since the condition $v = \omega_0 R$ must be valid up to the moment of extraction, the final energy of the cyclotron is fixed by its geometry and cannot be changed. Applications of the standard cyclotron to proton therapy thus have to rely on the use of range-shifters to adjust the penetration depth.

In addition to the fixed final energy, standard cyclotrons also suffer from the relativistic effects. Once protons accelerate to several tens of MeV, the perfect relation between the revolution time and frequency of the field starts to break.

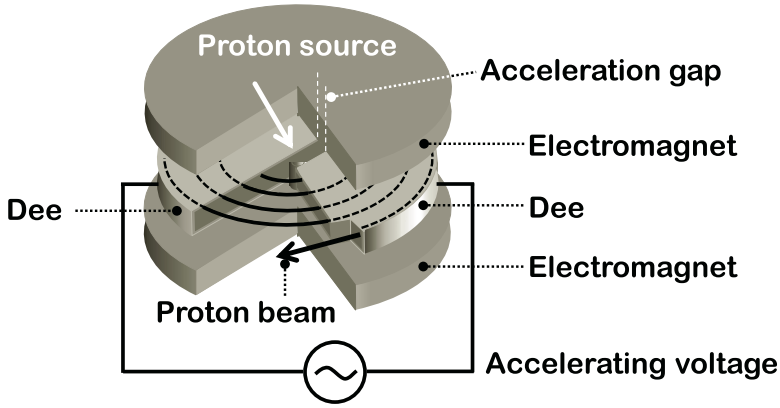



**FIGURE 9.7**
Schematic of a cyclotron.

In contrast to a standard cyclotron, a *synchrocyclotron* can have variable energies of the accelerated beam and it can also achieve higher energies — several hundreds of MeV.

In the synchrocyclotron, the relativistic effects are compensated by continuously decreasing the frequency of the accelerating voltage during acceleration so that

$$\omega = \frac{\omega_0}{\gamma(t)} \tag{9.3}$$

The time-varying accelerating frequency also means that only a bunch of a certain length can be in sync with the field — the synchrocyclotron therefore cannot accelerate CW current but can only produce pulsed beams.

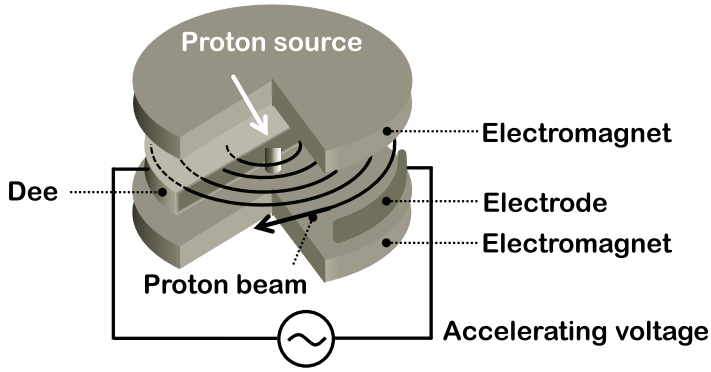



**FIGURE 9.8**
Schematics of a synchrocyclotron.

The electrode configuration in synchrocyclotrons is also different — only one dee remains, while the other electrode has a modified open shape as illustrated in Fig. 9.8. Adjustments to the final energy can be achieved by modifying the relations between the magnetic field and accelerating frequency, making it possible to accelerate protons to GeV energies. The pulsed mode of operating synchrocyclotrons, however, limits their *duty factor* and thus limits their intensity.

The variability of the final energy of a synchrocyclotron is an advantage that keeps attracting attention to their potential use as proton therapy machines. A notable recent development[1] includes a design of an iron-free superconducting synchrocyclotron, where dual nested solenoids are used to cancel the external fields of the device. The iron-free design makes the synchrocyclotron light and compact, and particularly suitable for mounting directly on a gantry.

*Compare this with Section 10.6.2 and analyze this design from the TRIZ point of view.*

Another type of cyclotron — the *isochronous cyclotron* — compensates the relativistic effects by allowing the magnetic

[1] A. Radovinsky et al., MIT report PSFC/RR-13-9, 2013.

field to increase with the radius of the orbit as

$$B = \gamma(R)\, B_0 \tag{9.4}$$

To increase the field with a radius in an isochronous cyclotron, shims attached to the poles can be used in such a way that the azimuthal fraction of the shims increase with the radius, as illustrated in Fig. 9.9.

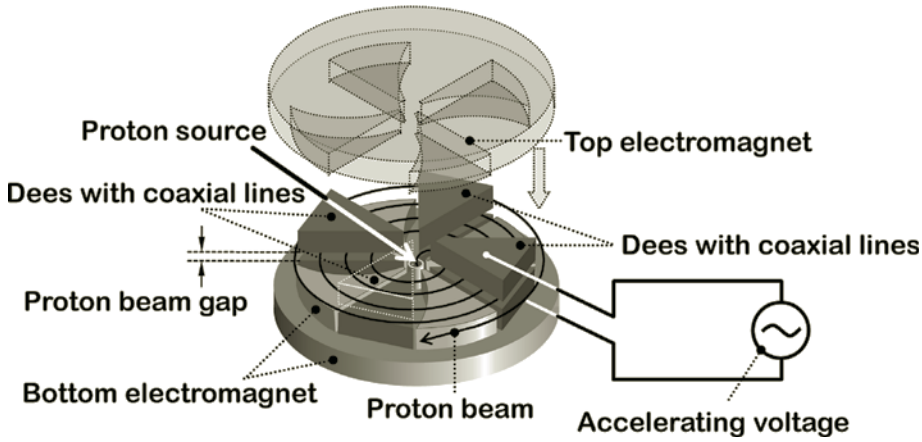



**FIGURE 9.9**
Schematics of an isochronous cyclotron.

The field increasing with the radius contradicts the requirements for weak focusing, and therefore strong focusing is arranged in isochronous cyclotrons by employing a spiral shape for the pole shims (often called *flutter* configuration), so that the regions of the field transition shown in Fig. 9.10 would introduce edge focusing on the beam.

The isochronous cyclotrons, as their acceleration frequency is constant, are suitable for CW operation. Proton energies of around a GeV can be achieved by compensating for the relativistic effects. The final energy of isochronous cyclotrons is fixed, as its adjustments require modification of the poles.

CW operation and the ability to reach higher energy made isochronous cyclotrons very popular for proton therapy applications. Superconducting isochronous cyclotrons have also been developed, with the goal to increase the compactness of proton therapy devices.

Synchrotrons are naturally also a possible choice as an accelerator for use in proton therapy. Their advantage is that they can provide variable energies; however, there are also a number of disadvantages. First of all, the size: synchrotrons are larger than cyclotrons. Second, synchrotrons are pulsed machines, while proton therapy benefits from a slow dose delivery. Therefore, slow extraction methods, such as those based on excitation of nonlinear resonances, may need to

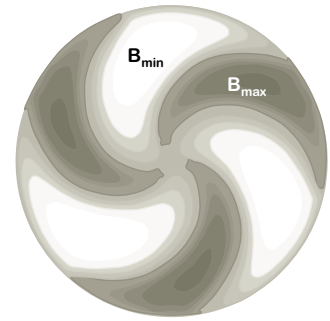



**FIGURE 9.10**
Example of a field profile in an isochronous cyclotron.

be used. Overall, synchrotrons are a viable choice and they are being used in many proton therapy facilities around the world.

## 9.4 Plasma acceleration of protons and ions — motivation

The motivation for developing laser plasma acceleration techniques for beam therapy can be defined by the following factors. On the one hand, there are advantages to using protons instead of X-rays for certain tumors, particularly in pediatric cases, where use of protons significantly reduces the probability of tumor recurrence or side effects due to lower dose for non-target tissues. On the other hand, widespread use of proton therapy facilities is limited by their overall cost.

Proton therapy systems require 250 MeV beams (or above 330–350 MeV when protons are also used for diagnostic imaging). Such systems, especially the beam delivery gantries, are large and expensive (see Fig. 9.4). The cost and size arguments are even more pronounced for heavy ion therapy systems based on the use of carbon ions, which create certain therapeutic advantages.

Meanwhile, laser plasma acceleration has demonstrated rapid progress, delivering several GeV in energy of quasi monoenergetic electron beams and more than 100 MeV proton beams. The progress is due to advances in lasers, where a CPA laser beam of a few hundred TW or around a PW (corresponding, e.g., to 400 J energy in 400 ns or 30 J in 30 fs), when focused to a 5 $\mu$m spot, can create an intensity on the order of $10^{25}$ W/m$^2$ — which is suitable for proton plasma acceleration. Such lasers, while still bulky and expensive, are rapidly improving and will become more efficient and compact in the future.

Ultimately, the desire to create compact laser plasma acceleration proton therapy systems is one of the main motivations for developing plasma acceleration of protons.

## 9.5 Regimes of proton laser plasma acceleration

In this section, we will very briefly describe several different regimes (mechanisms) of laser-driven proton acceleration.

We will start with a discussion of an already classical mechanism called *sheath acceleration*, or TNSA — *target normal sheath acceleration*. As we will see, this classical method provides protons with too large an energy spread and poor scaling with laser power.

We will then follow up on more recent and promising mechanisms that rely on *radiation pressure acceleration*, specif-

ically the mechanisms of *hole-boring* and *light-sail acceleration*. The latter has favorable scaling with laser power and promising perspectives for achieving near-monoenergetic beams.

We will briefly touch upon other mechanisms that have been identified, such as *shock acceleration* and the *relativistic transparency regime* — also called the *break-out afterburner*. Various academic review papers[2,3,4] can be consulted to gain deeper insights into the discussed areas.

Once more, we would like to stress that the mentioned mechanisms of plasma acceleration represent a gradual improvement of our understanding of this very complicated phenomenon. Different mechanisms compete, and often more than one mechanism is active in a particular case. As we strive to sharpen our understanding of these mechanisms, further significant progress in this area will be inevitable.

### 9.5.1 Sheath acceleration regime

Sheath acceleration (TNSA — target normal sheath acceleration regime) is illustrated conceptually in Fig. 9.11. Here, the laser pulse is focused on a thin metal foil that creates plasma. The plasma electrons quickly become relativistically hot and leave the foil, creating a sheath of charge, which then pulls out the ions and protons from the plasma.

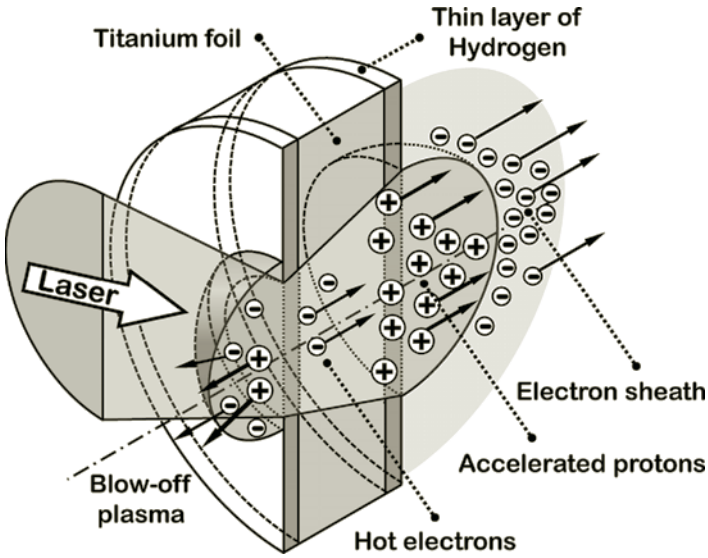



**FIGURE 9.11**
Sheath laser acceleration of protons.

[2] A. Macchi et al., Rev. Mod. Physics, 85, 751 (2013).
[3] H. Daido et al., Rep. Prog. Phys. 75, 056401 (2012).
[4] M. Borghesi et al., Fusion Science and Technology, 49, 412 (2006).

The surface of the foil is typically contaminated (which is, in this case, useful) by a thin layer of hydrogen that becomes the source of protons that are most readily accelerated. Acceleration of protons and ions happens on both sides of the foil. Still, the side opposite the laser irradiation side exhibits sharper boundaries of the electron sheath, providing higher energies and better beam quality of the accelerated protons.

Experimental investigations of the TNSA mechanism using lasers with intensities above $10^{19}$ W/cm$^2$ demonstrated[5] that multi-Mev ion acceleration from the rear surface of thin foils is possible.

The scaling rules for the TNSA mechanism can be derived by taking into account that a Debye sheath will form from hot electrons on the rear of the foil, and the potential difference $U$ through this sheath will be of the order of the electron temperature $T_e$ multiplied by the Boltzman constant: $U \approx k_B\, T_h$. Considering that the temperature of electrons is related to laser intensity $I$ and wavelength $\lambda$ as $T_e \propto \sqrt{I\lambda^2}$, one can obtain a scaling for the maximal energy of accelerating protons as

$$W_{max} \approx k_B\, T_e \propto \sqrt{I\lambda^2} \tag{9.5}$$

In the above equation, we have assumed in the first approximation that the maximum energy of the protons is equal to the value of the potential difference $U$. In fact, energies several times larger than that are expected, taking into account that the sheath is expanding and protons are "surfing" on the expanding potential.

In some of the first experiments,[6] proton energies close to 20 MeV were obtained. The spectrum of protons was, however, very broad and there were not many protons at the large energy side of the spectrum. Qualitative behavior of the TNSA proton spectrum is shown in Fig. 9.12. Even though later experiments demonstrated that the proton energies close to 100 MeV were possible (see review[7]), these disadvantageous qualities of the spectrum remained.

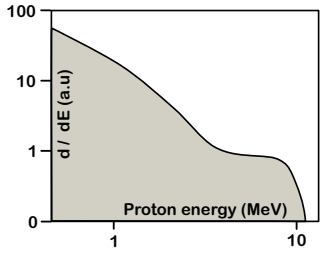



**FIGURE 9.12**
TNSA spectum, qualitative behavior.

Some of the characteristic properties of TNSA-produced beams include a large divergence (of several degrees) combined with micron-scale beam size. Such a beam may formally have a low emittance; however, it will quickly filament and the emittance will increase if the beam is not captured in an appropriate focusing system. The TNSA mechanism can produce $10^{11}$ to $10^{13}$ protons per shot, but not many of those protons will be at the high-energy edge of the spectrum.

Scaling rules predicted by TNSA models allow us to make projections (with large uncertainties) towards reaching a proton energy of around 200 MeV. These projections have

[5] Maksimchuk et al., PRL, 84, 4108 (2000); Snavely et al., PRL, 85, 2945 (2000).

[6] E.L. Clark et al., PRL, 84, p.670. (2000)

[7] M. Borghesi et al., Plasma Phys. Control. Fus., 50, 124040 (2008).

shown that the required laser intensities are on the order of $10^{21}$ W/cm$^2$.

Various ways to improve TNSA have been considered, starting from the brute force method to increase the laser intensity, to more subtle methods involving the enhancement of the laser energy transfer to electrons and increase of the electron density. Reduction of foil thickness and reduced mass targets, enhanced coupling by a conically shaped target and use of nano-particle structured targets, among others, have all been used with some degree of success.

Despite the mentioned improvements, the TNSA mechanism still has a major disadvantage due to the shape of the spectrum of accelerated particles — the number of particles at the high end of the spectrum remains very low.

### 9.5.2 Hole-boring radiation pressure acceleration regime

The radiation pressure acceleration mechanism is based on the effects equivalent to *radiation pressure*, which light exhibits on a mirror when it reflects from its surface, as illustrated in Fig. 9.13. The radiation pressure for perfect reflection is proportional to the laser intensity $I_L$ as

$$P_L = \frac{2\, I_L}{c} \tag{9.6}$$

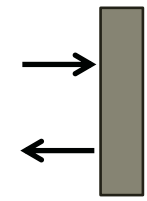


**FIGURE 9.13**
Radiation pressure acceleration concept.

Similar pressure can be applied to a thin foil, upon whose surface plasma quickly forms. The radiation pressure effect is transmitted into the plasma by electrons via the ponderomotive force. Displaced electrons produce space charge that creates a steady pressure, which in its turn transfers the effect to the ions.

Two versions of radiation pressure mechanisms have received distinct names — the *hole-boring* and the *light-sail* mechanism.

In the case of hole-boring, the space charge due to electrons acts on ions that are pushed into the overdense plasma, initially compressing the foil and then pushing a region of the foil forward, as illustrated conceptually in Fig. 9.14.

The derivation of the approximate scaling rules for the

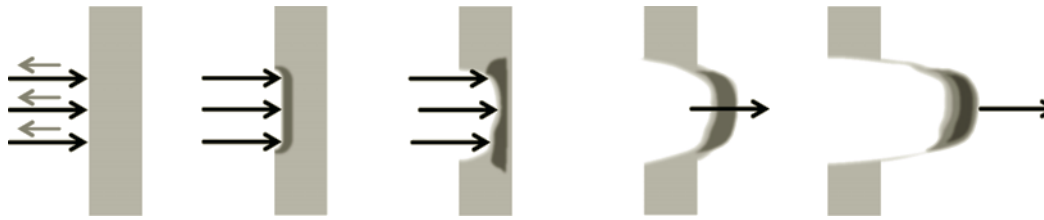


**FIGURE 9.14**
Hole-boring radiation pressure laser acceleration of protons.

hole-boring mechanism can be obtained by including the radiation light pressure into the fluid equation of motion, and transferring it to the reference frame where the shock is stationary (i.e., where the time derivatives are zero). This will yield $\rho u^2 = P_L$ and therefore the velocity of ions can be estimated as

$$u \approx \sqrt{\frac{I_L}{\rho c}} \tag{9.7}$$

Consider an example of a laser with $I_L = 10^{21}$ W/cm$^2$ shining on an aluminum target with $\rho = 2.7\,10^3$ kg/m$^3$. The resulting velocity of ions in this case is $u \sim 0.01c$ or about 3 mm/ps. The energy of the protons associated with this shock is 60 keV. By "bouncing" the stationary protons of the shock front, the protons can gain twice the velocity or four times the energy. Reducing the density of the foil could lead to a further increase of the energy of accelerated ions — the latter can be achieved by using gas targets.

The most attractive feature of the hole-boring radiation pressure mechanism is that the resulting proton beam has been demonstrated[8] to have a nearly monochromatic peak at the maximum energy.

### 9.5.3 Light-sail radiation pressure acceleration regime

In the *light-sail*[9] regime, the radiation pressure mechanism is taken to the extreme when the foil is so light that it starts to accelerate immediately as a whole, as shown schematically in Fig. 9.15.

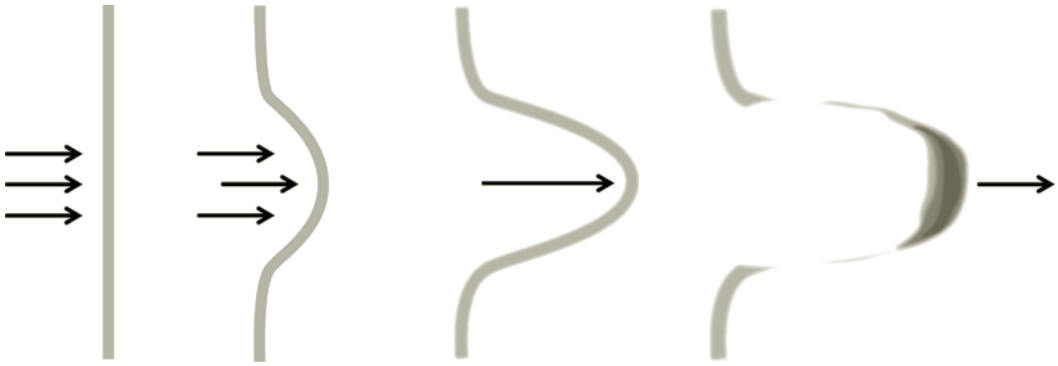


**FIGURE 9.15**
Light-sail radiation pressure laser acceleration of protons.

[8] C.A. Palmer et al., Phys. Rev. Lett., 106, 014801 (2011).
[9] T. Esirkepov et al. Phys. Rev. Lett., 92, 175003 (2004).

In a simple model of light-sail acceleration, one can assume that an area $A$ is illuminated with an intensity $I_L$ and if the light reflection is perfect, the force acting on this region of the foil is given by

$$F = 2A\frac{I_L}{\rho c} = m\frac{dv}{dt} = Ad\rho\frac{v_i}{t} \tag{9.8}$$

where $m$ is the mass of this segment of the foil, $m = A\ d\ \rho$, with $d$ being the foil thickness and $\rho$ the density. The velocity of ions will therefore grow proportionally to the laser intensity and laser pulse duration:

$$v_i \approx \frac{2I_L\tau}{\eta\ c} \tag{9.9}$$

where $\eta = \rho\ d$ — the areal density. As we can see, the energy scaling in the light-sail regime is more favorable than in the hole-boring case, as in the non-relativistic case of light-sail acceleration the velocity of ions is proportional to $I_L$ and the energy is thus proportional to $I_L^2$.

The light-sail radiation pressure mechanism cannot be completely decoupled from the competing TNSA mechanism. The electron heating — which was ignored in the simplified picture above — may cause foil deterioration. A possible improvement for the latter involves use of a circularly polarized laser, which reduces the effects of TNSA and foil heating.

The scaling of a light-sail mechanism for 10 PW pulses with laser intensity of $10^{22}$ W/cm$^2$ predicts[10] that GeV proton beams with good near-monochromatic spectral characteristics can be produced.

A Rayleigh–Taylor instability of the shape of the foil can develop during acceleration, resulting in deterioration of the resulted spectrum. However, this instability has been shown[11] to be stabilized by simultaneous acceleration of multi-species ions.

Overall, the proton energy scaling as $(I_L\ \tau/\eta)^2$ for the light-sail radiation pressure regime is the most favorable and promising.

### 9.5.4 Emerging mechanisms of acceleration

Various other mechanisms have been suggested that describe the behavior of laser plasma acceleration of ions in certain parameter ranges.

In particular, the *break-out afterburner* regime has been described[12] as the mechanism based on the appearance of the

[10] B. Qiao et al., Phys. Rev. Lett., 102, 145002 (2009).

[11] B. Qiao et al., PRL, 105, 1555002 (2010), T. Pu Yu et al., PRL, 105, 065002 (2010).

[12] Yin et al., Phys. Rev. Lett. 107, 045003 (2011).

*relativistic transparency* of an initially solid target. As a result, this leads to the enhancement of ion acceleration via the TNSA mechanism.

Shock acceleration is a mechanism[13] based on the appearance of a high Mach number electrostatic shock in the overdense plasma created by the sharp front of the laser pulse. The propagating electrostatic shock will then reflect the plasma ions to the doubled velocity of the shock, resulting in an appearance of monochromatic proton peaks observed in experiments using overdense gas jet targets.

As we can see, while the number of forthcoming theoretical and experimental tasks and questions is still very high in this area, waiting for an inquisitive mind, there are mechanisms of proton plasma accelerations which, given the rate of the laser technology progress, can already yield the practically useful techniques for creating beam therapy facilities based on proton plasma acceleration.

## 9.6 Glimpse into the future

Creating a more compact and affordable design of a proton therapy facility is a formidable task, one that is attracting the attention of many research teams worldwide. The plasma acceleration community, together with laser and conventional acceleration communities, are joining forces to solve this task.

Techniques of beam control and energy selection, along with methods of capturing divergent and chromatic beams, all developed alongside conventional accelerators and combined with novel opportunities enabled by modern and future lasers, should help us to find a way to create a viable design and a prototype system that can be used efficiently in practice.

In Chapter 10, we will briefly review various methods of beam control and manipulation. Some of these methods can be particularly applicable to the design of a plasma proton acceleration facility.

[13] L. Silva et al., PRL 92, 015002 (2004); D. Haberberger et al., Nature Phys., 8, 95 (2012).

## EXERCISES

9.1 *Chapter materials review*.
Discuss the advantages and challenges of particle beam therapy in comparison to X-ray therapy.

9.2 *Chapter materials review*.
Discuss the key requirements for the laser and the target that may result in 200 MeV mono-energetic beams of protons in laser plasma acceleration.

9.3 *Mini-project*.
Discuss and develop a plan to create a 250 MeV proton source based on plasma acceleration, aiming for it to be applied in the medical field. Select approximate laser parameters and target parameters, and discuss their requirements. Discuss and select a method for energy monochromatization or energy collimation/selection. Describe why you selected these particular values of certain parameters (for target or laser, collimation or monochromatization system, etc.).

9.4 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this chapter (e.g., isochronous cyclotron) in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

9.5 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 10 Advanced Beam Manipulation, Cooling, Damping and Stability



We have arrived at the second-to-last chapter. The scope of this chapter is a mixture of beam and pulse manipulation topics, selected based on the degree of the synergy they exhibit in four areas that we previously discussed. The first three areas are accelerators, lasers and plasma (see Chapter 4), and the fourth area is the methodology of inventiveness.

We will start this chapter with the methods and techniques for creating short bunches of charged particles and also short laser pulses. We will follow with a discussion of various topics of beam stability and ways to enhance the intensity of both the particle beams and the laser pulses. We will conclude with an overview of cooling methods, as well as local disturbance correction methods, aiming to connect these and other techniques to *Accelerating Science TRIZ*.

## 10.1 Short and narrow-band

The topic of bunch and pulse compression is an area that shows strong synergy between accelerators and lasers. This is a broad area that includes techniques of bunch or pulse compression as well as techniques for creating short bursts of radiation. We will also pay attention to some effects that may prevent the formation of ultra-short bunches — notably *coherent synchrotron radiation*. Connected to the topic of short pulse or bunch creation is the theme of narrow-band radiation, where techniques developed in optics and lasers merge with the accelerator-based techniques.

### 10.1.1 Bunch compression

Short electron bunches are typically produced by RF photo guns where the cathode is illuminated by a short laser pulse. Such photo injectors produce relatively long pulses (a few ps), while for many applications, such bunches would be too long and femtosecond bunches would be required instead.

Some compression of the beam may be performed already in the photo injector, via the technique of *velocity bunching*. In this case, by properly adjusting the moment when the laser illuminates the cathode with the RF field in the gun, one can create correlation between the energy of electrons and the longitudinal position within the bunch. The difference of

DOI: 10.1201/b18696-10

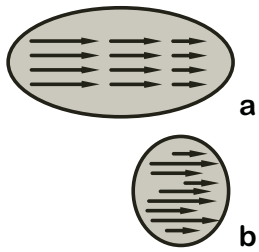



**FIGURE 10.1**
Velocity bunching. Initial beam (a) and compressed beam (b).

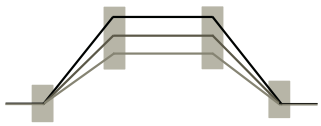


**FIGURE 10.2**
Four-magnet chicane.

energy for beams that are still weakly relativistic can create a difference in longitudinal velocities. If the faster particles are arranged to be at the tail of the bunch (as illustrated in Fig. 10.1), they can catch up, resulting in a shortening of the bunch.

Because velocity bunching is based on the velocity's dependence on energy, it can work only for weakly relativistic beams and this, in fact, is the method's main shortcoming. Weakly relativistic electron beams can, in particular, suffer strongly from space charge effects, limiting the degree of compression.

Therefore, achieving ultra-short electron bunches is usually done at higher energies, when the beam is relativistic and space charge effects are less severe. Shortening the bunch length in this case is usually achieved with a magnetic compression system. When discussing bunch compressors, we will build on the foundation established in Chapter 4, further expanding it towards an analytical description of the process.

A typical arrangement that can compress the bunch is a beamline made of four bending magnets of opposite polarity arranged as shown in Fig. 10.2 — called a *chicane*. In this chicane, the time of flight (or equivalently the path length) is different for different energies.

In order to exploit the dependence of the time of flight (or path length) on the particle's energy, we need to introduce an energy–time correlation within the bunch. This correlation can be created using the electric field of an RF cavity, properly phased with the beam.

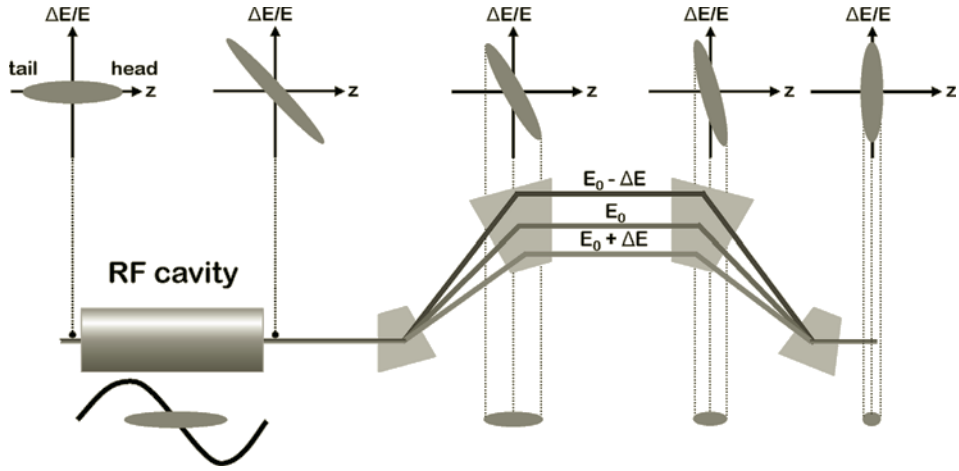



**FIGURE 10.3**
Energy–time correlation and bunch compression.

The RF cavity will create an energy chirp along the bunch — the necessary condition for bunch compression to work; see Fig10.3. The chirp is phased in such a way that the particles in the tail have a higher energy and will therefore travel

on a straighter path through the chicane, catching up with the synchronous particle, and resulting in bunch compression.

The bunch compression process can also be described analytically. Let's use the linear transfer matrices to evaluate the evolution of the longitudinal position and relative energy offset $(z, \delta)$ as the particle propagates through the RF cavity and then the bunch compressor.

The first step is to evaluate how the coordinates $(z, \delta)$ change in the RF cavity. Passing through the RF cavity (assuming the cavity is thin), the longitudinal coordinate does not change, while the energy change depends on the initial position $z_0$ as follows:

$$\begin{aligned} z_1 &= z_0 \\ \delta_1 &= \delta_0 + \tfrac{eV_{RF}}{E_0} \cos\ \tfrac{\pi}{2} - k_{RF} z_0 \end{aligned} \tag{10.1}$$

The above equations are equivalent to the following matrix transformation expressed in linear approximation in $(z, \delta)$ as follows (refer to Eq. 2.58 and Eq. 2.59 for definitions of the indexes of the coordinate vector and of the transfer matrix):

$$\begin{pmatrix} z_1 \\ \delta_1 \end{pmatrix} \approx \begin{pmatrix} 1 & 0 \\ R_{65} & 1 \end{pmatrix} \cdot \begin{pmatrix} z_0 \\ \delta_0 \end{pmatrix} \tag{10.2}$$

where

$$R_{65} = \frac{eV_{RF}}{E_0} \sin(\varphi_{RF}) k_{RF} \tag{10.3}$$

The next step is to take into account the bunch compressor itself. In the chicane, the particle coordinate will change according to the following general expression (the higher-order terms are defined as in Eq. 2.60):

$$\begin{aligned} z_2 &= z_1 + R_{56}\delta_1 + T_{566}\delta_1^2 + U_{5666}\delta_1^3 \dots \\ \delta_2 &= \delta_1 \end{aligned} \tag{10.4}$$

which can be linearly approximated as

$$\begin{pmatrix} z_2 \\ \delta_2 \end{pmatrix} \approx \begin{pmatrix} 1 & R_{56} \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ \delta_1 \end{pmatrix} \tag{10.5}$$

The full transformation is given by multiplying the matrices of each element, which can be computed to be given by the following:

$$\begin{pmatrix} z_2 \\ \delta_2 \end{pmatrix} \approx \mathbf{M} \cdot \begin{pmatrix} z_0 \\ \delta_0 \end{pmatrix} \qquad \mathbf{M} = \begin{pmatrix} 1 + R_{65}R_{56} & R_{56} \\ R_{65} & 1 \end{pmatrix} \tag{10.6}$$

Dependence of the final bunch length on the initial beam length is therefore given by the following equation

$$\sigma_{z_2} = |1 + R_{65}R_{56}| \sigma_{z_0} \tag{10.7}$$

We can see that, in order to achieve maximal compression, we

should adjust the value of $R_{65}$ (energy chirp induced by the RF cavity) in such a way that

$$R_{65}R_{56} \approx -1 \tag{10.8}$$

in which case the *compression factor* $C$ will be the largest and the final bunch length

$$\sigma_{z_2} = \frac{\sigma_{z_0}}{C} \quad \text{where} \quad C = 1/|1 + R_{65}R_{56}| \tag{10.9}$$

will be minimal.

Let us now discuss the minimal achievable bunch length. While in the linear approximation, the condition $R_{65}R_{56} = -1$ suggests that $C$ is infinite and the resulting final bunch length is zero, this cannot be achieved in practice.

First of all, taking second and higher-order terms into account will give us non-zero final bunch lengths for any initial parameters.

Secondly, there is a limitation caused by the longitudinal emittance of the beam. The linear transformation of the $(z, \delta)$ coordinates preserves the longitudinal beam emittance according to the *Liouville's theorem* (it's often said that the corresponding transformation is *symplectic*). The longitudinal emittance is given by the following:

$$\varepsilon = \sqrt{\sigma_z^2 \sigma_\delta^2 - \sigma_{z\delta}^2} \tag{10.10}$$

which tells us that the minimum reachable bunch length is limited to the product of the beam energy spread $\sigma_\delta$ times the matrix element $R_{56}$.

Furthermore, additional limitations to the achievable compression come from the effects associated with the beam's high peak current that we have neglected in the linear approximation denoted above. These effects include longitudinal space charge, wakefields and *coherent synchrotron radiation* (CSR). When taken into account, these effects can produce serious degradation to the quality of the beams and limit the achievable minimum bunch length.

### 10.1.2 CSR — coherent synchrotron radiation

We will now evaluate the *coherent synchrotron radiation effect* in a back-of-the-envelope fashion, following the example of Chapter 3.

The characteristic frequency of SR

$$\omega_c = \frac{3}{2}\frac{c\gamma^3}{R}$$

defines the higher edge of the SR spectrum while radiation is emitted at any frequencies $\omega$ below $\omega_c$.

Let us assume that we have a bunch with $N$ electrons

and its bunch length is $\sigma$. For certain frequencies $\omega$ of synchrotron radiation, its wavelength $\approx c/\omega$ can be longer than the bunch length $\sigma$; the beam at these frequencies will radiate coherently, as illustrated in Fig. 10.4.

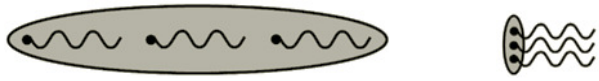


**FIGURE 10.4**
Incoherent radiation (left) and coherent radiation (right).

The two regimes can thus be defined as $c/\omega < \sigma$ — incoherent radiation — and $c/\omega > \sigma$ — coherent radiation.

For frequencies where c/$\omega$ is longer than $\sigma$, the beam radiates coherently, increasing radiated power proportionally to the square of the number of particles involved, i.e., as $N^2$. As the typical bunch population N is $\sim 10^9$ to $\sim 10^{10}$, this effect increases the radiated power tremendously.

Assuming that the spectrum of incoherent SR is given by $I(\omega)$, the CSR spectrum can be approximated as

$$I_b(\omega) = I(\omega) N \left[ 1 + N \exp\left( - \frac{\omega\sigma}{c}^2 \right) \right] \qquad (10.11)$$

A typical spectrum described by the above equation is shown in Fig. 10.5.

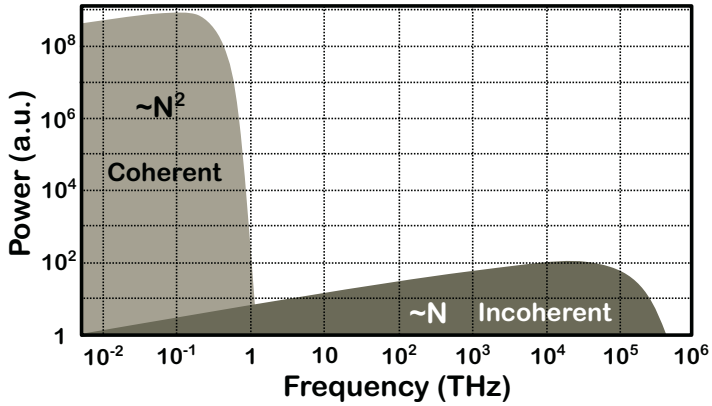



**FIGURE 10.5**
Qualitative comparison of the spectrum of coherent synchrotron radiation in comparison with the spectrum of incoherent SR.

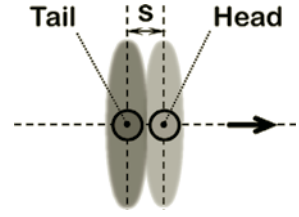



**FIGURE 10.6**
Two-particle model of the beam and its field.

### 10.1.3 CSR effects on the beam longitudinal phase space

We will now make a back-of-the-envelope estimate[1] of the force that CSR exerts on the beam. We will use an approxi-

[1] Following derivations from Ya. Derbenev et al., Tesla-FEL Report 1995–05.

mation of the *two-particle model*. In this instance, the beam is represented just by two particles — at the head and in the tail of the bunch, as illustrated in Fig. 10.6, where the particles — as well as its field — are shown qualitatively. As we will see, the CSR effects are essentially caused by the possibility for the tail field to overtake the head particle while the beam is moving on a curved trajectory.

Let us thus define the bunch as a two-particle system with the length between the head and the tail $s$ equal to *rms* bunch length of the initial beam $\sigma$. The EM fields of the relativistic bunch are distributed primarily in transverse directions, as indicated in Fig. 10.6.

Assuming that the beam is moving along the curved trajectory shown in Fig. 10.7, we can determine the moment at which the field of the tail, radiated at point A, and moving on the straight line, will catch up with the head of the beam. The condition for the catch-up (overtake) to happen in point B can be expressed as follows:

$$Arc\,(AB) - s = |\mathbf{AB}| \tag{10.12}$$

which can be rewritten as

$$R\,\theta - s = 2R\sin(\theta/2), \quad \text{which gives} \quad s \approx R\,\theta^3/24 \tag{10.13}$$

Thus the *overtaking distance* $L_0$ can be estimated as

$$L_0 = |\mathbf{AB}| = R\,\theta \approx 2\,\left(3sR^2\right)^{1/3} \tag{10.14}$$

Knowing the overtaking distance, we can determine the characteristic transverse distance $r$, which is equal to the distance between the head particle at the point of overtake and the axis of propagation of the SR fields that travel along the initial trajectory (Fig. 10.7).

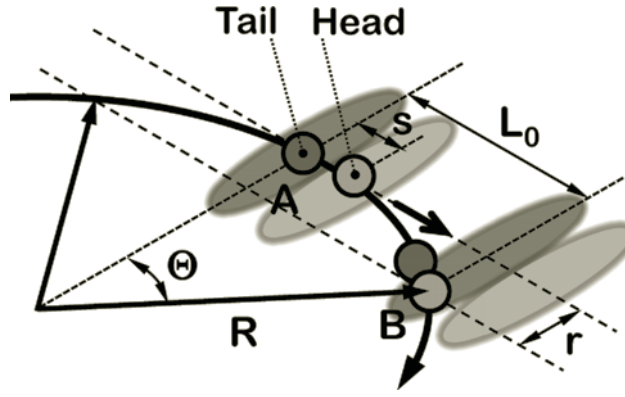



**FIGURE 10.7**
Illustration of the tail field overtaking the head of the bunch in the mechanism of coherent synchrotron radiation.

The transverse characteristic distance can be estimated as follows:

$$r = L_0\,\theta/2 = 2\ 9s^2R^{\ 1/3} \tag{10.15}$$

We can now estimate the field of the bunch tail (radiated at point A), acting on the head at point B.

Assume that the beam is uniform and has the linear change density $eN\lambda$ where $\lambda = 1/\ell_b$ and $\ell_b$ is the length of the bunch.

The values of the transverse field at the characteristic distance for the linear charge density beam are as follows:

$$E_\perp = H_\perp \approx \frac{2N_e\lambda}{r} \tag{10.16}$$

In order to find the longitudinal field acting on the head, we need to multiply the transverse field by the angle: $E_\perp \cdot \theta$. The longitudinal force acting on the head can thus be estimated as

$$F_\parallel = eE_\perp \cdot \theta = \frac{2Ne^2\lambda\theta}{r} = \frac{2Ne^2\lambda}{(3sR^2)^{1/3}} \tag{10.17}$$

Let us now assume that $s = \ell_b = 3^{1/2}\sigma$ (the latter assumes that the bunch distribution is Gaussian). The estimate for the longitudinal force thus becomes

$$F_\parallel \approx \frac{2Ne^2}{3R^{2/3}\sigma^{4/3}} = \frac{2N\,r_emc^2}{3R^{2/3}\sigma^{4/3}} \tag{10.18}$$

which is a rather accurate back-of-the-envelope estimate.

Accurate derivations of the CSR effects for a realistic Gaussian bunch can show that the longitudinal force acting

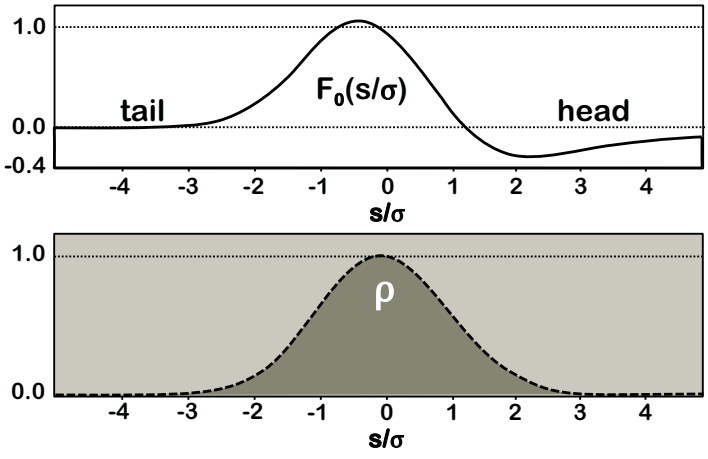



**FIGURE 10.8**
Shape function $F_0$ (top plot) of coherent synchrotron radiation for a bunch with Gaussian density profile (bottom plot).

on the beam is very close to the estimate Eq. 10.18 and contains an extra term: the shape function $F_0$, which depends on the position of the particle within the bunch.

$$F_{\parallel} \approx \frac{2N\, r_e mc^2}{3R^{2/3}\sigma^{4/3}}\, F_0 \tag{10.19}$$

The shape of the CSR shape function is shown in Fig. 10.8. As we can see, its amplitude is close to one, confirming the back-of-the-envelope estimate. The shape function changes its sign. While the head of the bunch slightly accelerates due to the CSR effect, the major part of the bunch decelerates.

The effects of CSR are particularly important in bunch compressors and can prevent achieving ultra-short bunches. CSR can cause bunch instability and microbunching, and can therefore deteriorate the longitudinal phase space of the beam.

To conclude, one should note that the mechanism of CSR's creation — depicted in Fig. 10.7 — suggests that, in certain parameters, there may be a "cure" for CSR, because the vacuum chamber where the beam and fields propagate can partially shield the fields and reduce the CSR effects.

### 10.1.4 Short laser pulse and Q-switching techniques

The methods of generating short laser pulses often rely on so-called *Q-switching techniques*. In this case, the laser gain medium is placed in the optical cavity (the quality factor Q of this cavity can be controlled). In the initial moment, the Q factor is set to a low value, and pumping of the laser medium will then build up a large population inversion in the medium.

Once a sufficient inversion population is achieved, the Q factor of the laser cavity is suddenly increased, which results in quick buildup of the light in the cavity and an avalanche of stimulated emission. A giant and short pulse is thus emitted from the laser cavity. This technique is illustrated in Fig. 10.9.

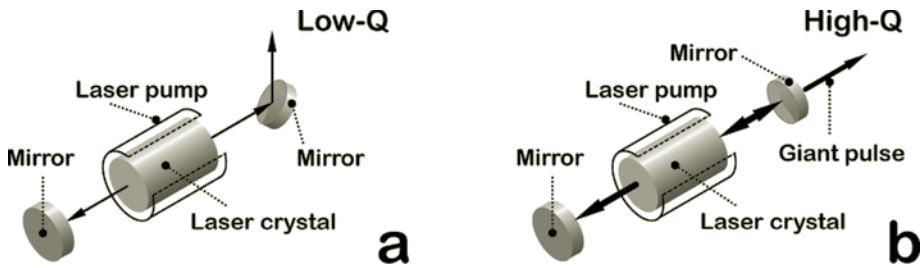



**FIGURE 10.9**
Q-switching technique. In step one (a) the pump builds up large inversion in the gain media. In step two (b) the laser cavity switches from low to high-Q.

### 10.1.5 Q-switching methods

The methods used to switch the Q factor in the laser cavity can be divided into two categories: passive and active. Active methods may involve mechanical effects, e.g., rotation of the laser cavity mirrors or applying an EM field to a substance inserted into the cavity (which changes its optical properties under the influence of the field). Passive methods involve special mirrors that can change the Q factor reflectivity when the power level of radiation reaches a certain level.

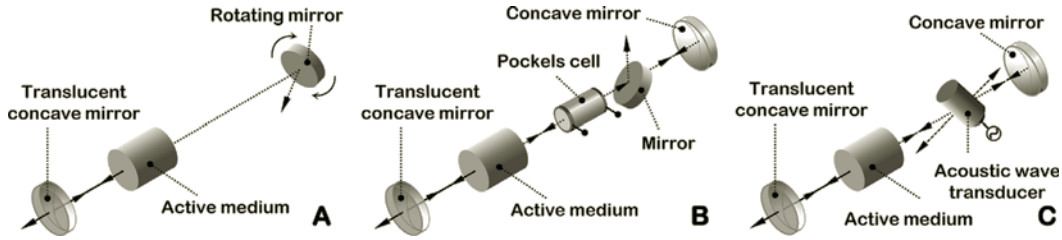



**FIGURE 10.10**
Examples of active Q-switching methods. Rotating mirror (A), Electro-optic (B) and Acousto-optic (C).

Some of the active methods are presented in Fig. 10.10. The first active Q-switching method is mechanical and involves the use of a rotating mirror. In this case, *lasing* will happen only at the moment when the mirror is parallel to the other mirror. A rotating cylinder — to which many mirrors are connected — is sometimes used because it generates many pulses for each rotation of the cylinder.

The speed of the mechanical active Q-switching methods is obviously limited by the mechanical strength of the movable objects, induced vibration and other factors. Active methods that minimize mechanical motion can avoid these limitations. The second active method is based on the use of the *acousto-optic* effect — the Bragg diffraction of light from the planes of a varied refractive index created in a crystal by the applied sound wave. When the acoustic wave transducer is switched on, a certain fraction of light diffracts away from the main light path in the laser cavity, which can then be used to change the cavity's Q value. The third active method is based on the *electro-optic* effect — the dependence of optical properties on objects such as absorption or refraction (called *Pockels effect*) on the applied electric field.

Similar to the active methods, various passive methods of Q-switching exist. Two examples shown in Fig. 10.11 include a saturable absorber and a SESAM. The saturable absorber uses the fact that certain absorbers become transparent when they reach saturation and cannot absorb radiation any further. At the moment when they become transparent, increasing the Q factor of the cavity can lead to lasing.

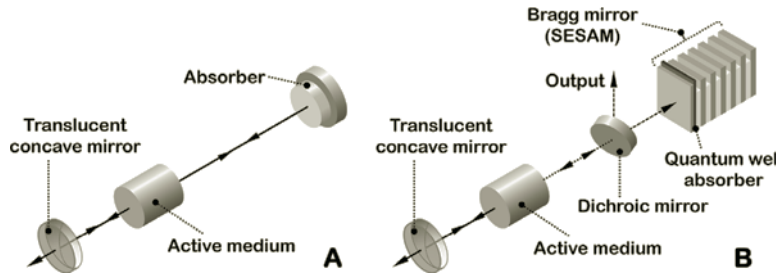



**FIGURE 10.11**
Passive Q-switching — saturable absorber (A) and SESAM (B).

A saturable absorber can be built based on doped YAG crystals or GaAs, on media with immersed *quantum dots*, or semiconductors (called *semiconductor saturable absorber mirrors* — SESAM). Another possible design of a passive Q-switch is based on a *thin-film absorber* where a quarter-wave plate is combined with a *thin-film polarizer* (TFP), serving as a hold-off polarizer in the Q-switched laser cell.

### 10.1.6 Regenerative amplifiers

The technology employed in Q-switching techniques is used in *regenerative laser amplifiers* (often called "regens"), which are designed to generate short, powerful laser pulses. In a "regen," a laser amplifier is placed inside of a Q-switched optical cavity (see Fig. 10.12) and it operates as follows.

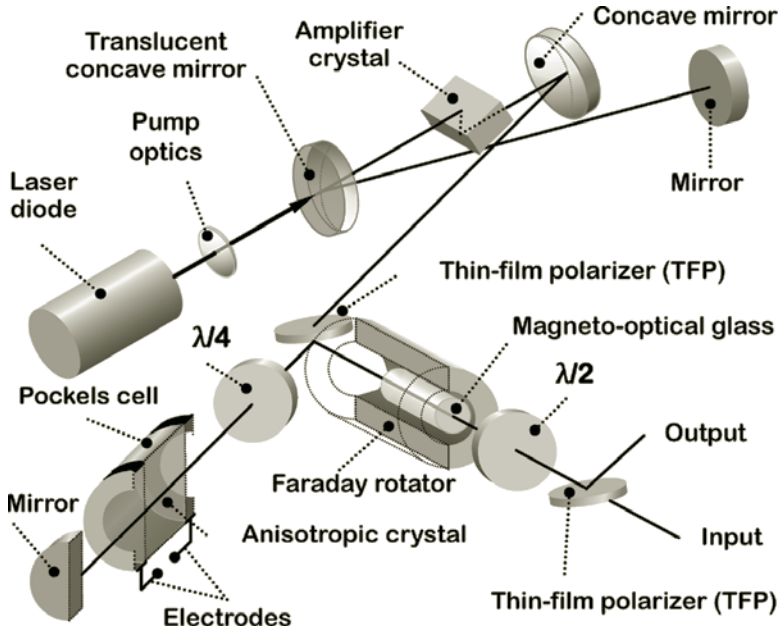



**FIGURE 10.12**
Schematics of a regenerative amplifier.

First, the gain medium (amplifier crystal in Fig. 10.12) is pumped by the laser diode. A short initial pulse from the "master oscillator" is then injected into the input of the system through the thin-film polarizer and the Faraday rotator, both of which are switched on for this short moment in order to let the input pulse pass.

The initial pulse then undergoes many round trips through the system, its power level getting amplified each time it passes the gain medium until it reaches a high level. Finally, another switch is powered (Pockels cell in Fig. 10.12) and the amplified pulse is released from the system.

### 10.1.7 Mode locking

*Mode locking* is the technique for achieving short laser pulses. In this method a rapid modulator is installed in the laser cavity, one that can open for short moments exactly in sync with the pulse's round-trip time around the cavity; see Fig. 10.13. Therefore, all the generated and amplified photons will be clustered only within those short moments when the modulator is open.

The amount of time of an average round trip in a laser resonator can be on the order of a nanosecond, while the modulator can open for periods of tens of femtoseconds. Correspondingly, the peak power in the mode-locked lasers can be of many orders of magnitude higher than the average power.

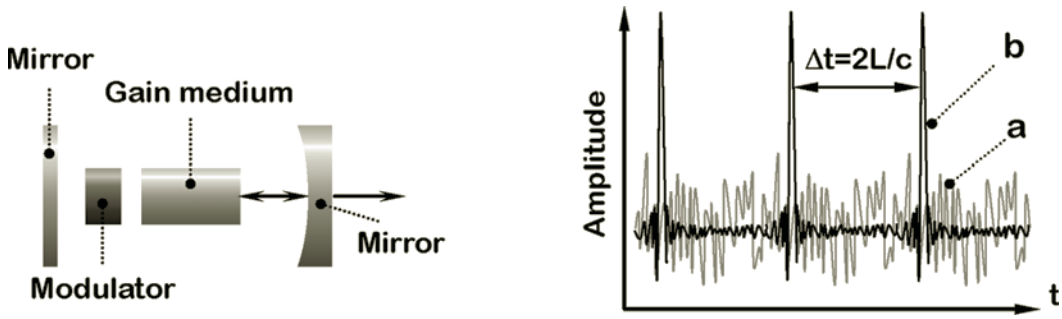



**FIGURE 10.13**
Mode-locked laser (left) and the laser output (right) in the normal (a) and mode-locked (b) regimes.

The name of the technique — mode locking — is a term derived from the analysis of this problem in the frequency domain. The laser cavity has a certain bandwidth that can support many longitudinal modes — all of which can coexist. In a normal laser, the phases between different modes are random. The mode locking created by the modulator would, in fact, create a certain fixed relation between the phases of all modes. In this case, there will be only one point in the laser cavity at each moment of time where the modes' electric

fields will add together constructively — which is equivalent to the case of a single short pulse travelling inside the laser cavity.

### 10.1.8 Self-seeded FEL

Various techniques used in optics and particularly in laser systems are ideologically similar to techniques employed for creating short pulses of FEL radiation. The *self-seeding* technique described in this section has certain similarities with the mode-locking method described in the previous section.

Recall that, in the SASE regime, the FEL lasing starts from noise, and also remember that the length of the slice that contributes to radiation is much shorter than the length of the electron bunch.

This means that each individual slice can generate radiation at slightly different wavelengths near the resonant wavelength, and, moreover, the amplitude of the radiation coming from each slice can be slightly different, as illustrated in the $P(\lambda)$ spectrum on the left side of Fig. 10.14.

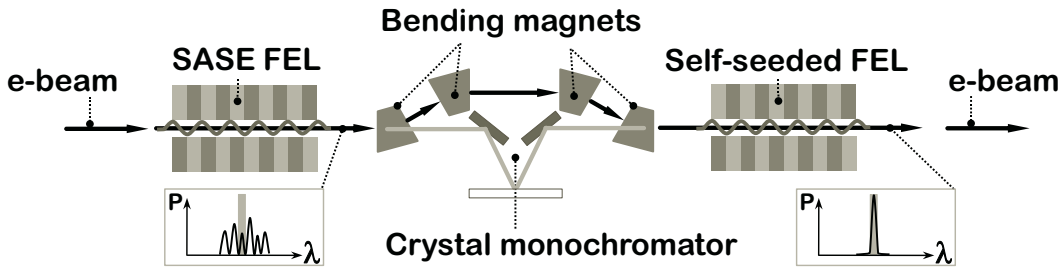



**FIGURE 10.14**
Self-seeded FEL.

The output structure of the FEL can be considerably improved using the self-seeding approach. In this event, the FEL is split into two parts. Before entering the second part, radiation generated in the first part is passed through a crystal monochromator. In order to have the timing maintained, the beam is simultaneously passed through a four-bend chicane.

The second part of the FEL is thus seeded with narrow spectrum radiation, which continues to be amplified along the way in the undulator; the resulting output spectrum of FEL then contains a narrow peak, as shown in Fig. 10.14 on the right side.

## 10.2 Laser–beam interaction

Interaction of laser light with electron beams in wigglers offers a wide range of techniques that can be used to manip-

ulate the electron beam or to improve the properties of FEL radiation.

We recall from Chapter 8 that the light of a resonant harmonic will be exponentially amplified in an undulator, taking the energy from the electron beam. The inverse can also be true — sending laser light at the resonant harmonic into the undulator will affect the electron beam (inverse FEL), creating variations of its energy.

The inverse FEL principle is the basis for many techniques used to manipulate electron beams, create short pulses of radiation, generate higher harmonics, create two-color FEL pulses and many others uses.

### 10.2.1 Beam laser heating

Short laser pulses can be used to improve the properties of short pulses of FEL radiation. The first example we will consider is the *laser heater*.

The generation of femtosecond-short X-ray pulses requires fs-short electron bunches. However, as we saw earlier in this chapter, the creation of short bunches in bunch compressors can be complicated due to CSR effects that can cause instability and microbunching.

Instabilities can often be suppressed (thanks to decoherence — see discussion of Landau damping in Section 10.3.4 later in this chapter) if the beam has sufficient spread of its relevant phase space coordinates (the energy spread, in the case of CSR instability).

The issue that has been observed in FELs is that the beam energy spread coming from a photocathode gun is extremely small, and insufficient for suppressing CSR instability.

A method has been developed at SLAC to introduce additional uncorrelated energy spreads into the beam: laser heating (see Fig. 10.15).

In this case, the laser light of the resonant wavelength co-propagates with an electron beam through a wiggler. This system acts as an inverse FEL and introduces an additional

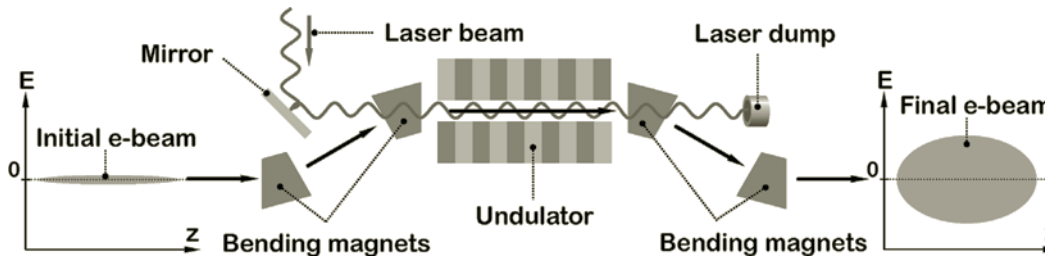



**FIGURE 10.15**
Laser heater.

energy spread to the beam, sufficient to cure the CSR instability.

### 10.2.2 Beam laser slicing

Beam laser slicing is a technique that selects a femtosecond short portion of radiation from a much longer initial electron pulse.

In this method,[2] a very short laser pulse overlaps with the center of a longer bunch in the undulator or wiggler; see Fig. 10.16. The laser wavelength $\lambda_L$ matches the undulator resonance condition

$$\lambda_L = \frac{\lambda_W}{2\gamma^2}\left(1 + \frac{K^2}{2}\right)$$

and the interaction of the light with the beam in the undulator will therefore produce modulation of energy in the short beam slice overlapped with the laser pulse.

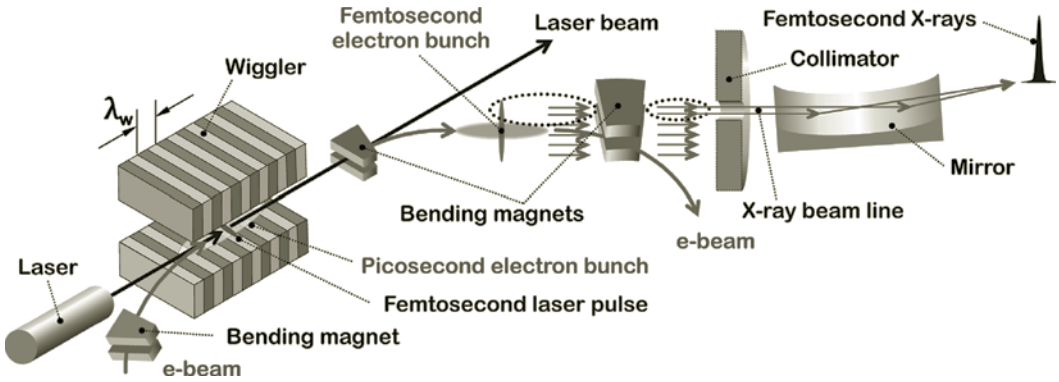



**FIGURE 10.16**
Beam laser slicing.

The electron beam coming out from the undulator will therefore have a short fs region with an increased energy spread. The energy spread of this short region can be converted to spatial variation using a dispersive beamline section. The resulting beam can then pass through a bend to generate synchrotron radiation and, following that, the SR corresponding to the short portion of the electron beam can be transversely separated from the rest of the pulse using collimators.

The method described above can be especially suitable for ring-based SR sources, where the natural length of the electron bunch is a picosecond long. The application of the laser slicing technique can therefore help in generating femtosec-

[2] A. A. Zholents and M. S. Zolotorev, Phys. Rev. Lett. 76, 912 (1996).

ond synchrotron radiation pulses, even for third-generation sources.

### 10.2.3 Beam laser harmonic generation

*Harmonic generation* is the technique that can help to produce X-ray photons of much higher energies. We recall from Section 8.3.3 that odd higher harmonics can also resonate and can thus be generated in an FEL. An FEL would normally generate primarily the first harmonic. In order to primarily generate a higher-order harmonic, external seeding is required. There are, however, no conventional lasers of appropriately short wavelengths that can be used for such seeding.

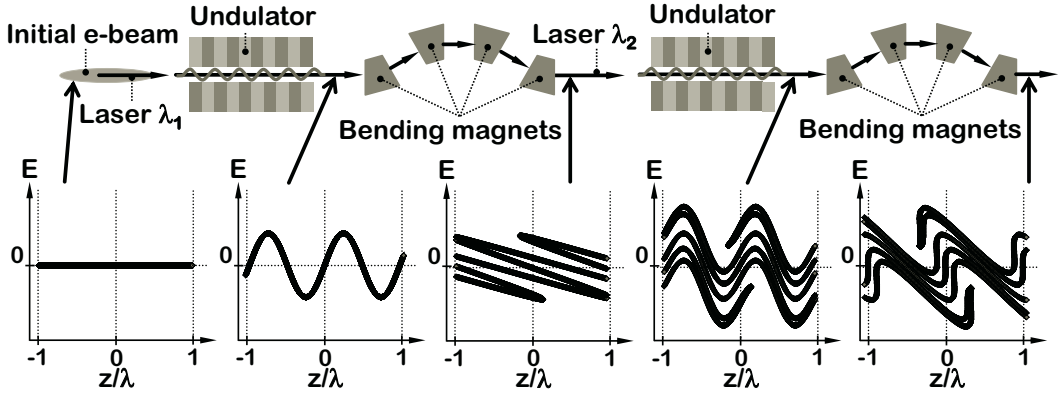



**FIGURE 10.17**
Echo-enabled harmonic generation scheme — EEHG.

A technique invented[3] by G. Stupakov is currently solving the problem: *echo-enabled harmonic generation* (EEHG). The term *echo* came from plasma physics and refers to the phenomenon of a spontaneous appearance of a wave with wavenumber $k_3$ in plasma, at a certain time after the initial excitation waves with wavenumbers $k_1$ and $k_2$ would decay via the collisionless *Landau damping* mechanism.

It is useful, however, to discuss the mechanism of EEHG without referring to the physics of plasma echo. The EEHG technique is based on laser-beam interaction in wigglers and distortion of the $z-E$ phase space in four-bend magnetic chicanes; see Fig. 10.17. Here, the first wiggler and laser with wavenumber $k_1$ create sine-like modulation of the $z-E$ phase space.

The first chicane applies $z = z + R_{56} \cdot \Delta E/E_0$ transformation that deforms the sine-like phase space into diagonally distorted lines. The second wiggler and laser $k_2$ modulate the beam again, and the second chicane creates another distor-

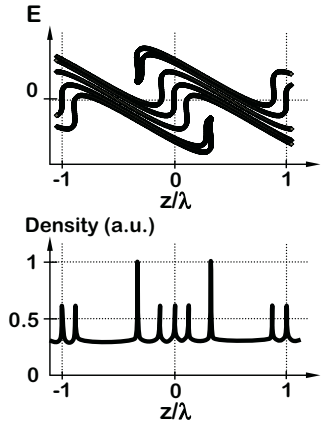



**FIGURE 10.18**
Phase space (top) and density profile (bottom) of an EEHG-modulated beam.

[3] G. Stupakov, PRL 102, 074801 (2009).

tion, which, if properly adjusted, produces repeating vertical lines in the $z - E$ plane (see Fig. 10.18), which correspond to spatial density modulation of the beam at high harmonics of the lasers.

Spatial harmonics that can be created with this method are defined by $k = nk_1 + mk_2$, where the integers $n$ and $m$ can be large, granting the potential for seeding at very short wavelengths.

## 10.3 Stability of beams

Stability of particle beams or laser pulses is usually one of the most important requirements of any design. Any practical realization of a beamline or laser system has imperfections: static (caused by misalignments) or dynamic (caused, e.g., by vibrations) positioning errors. In this section, we will look at just a few selected examples from this wide topic.

### 10.3.1 Stability of relativistic beams

*Head–tail effects* are often the cause of instabilities and deterioration of bunch properties, and occur when an initial offset of the bunch head creates much larger oscillations in the bunch tail.

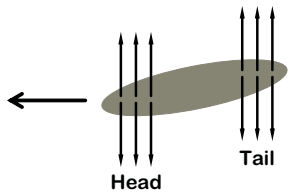



**FIGURE 10.19**
Fields of the bunch and head–tail effects.

However, we need to take into account that the fields of a relativistic bunch are mostly transverse, as illustrated in Fig. 10.19. Therefore, in the free space, the motion of the tail of the bunch would be independent of the motion of the head, and in particular as to whether the head has offset and/or oscillations.

For the head–tail instability to develop, it is necessary to have an "agent" to carry the information about the offset from the bunch head to the bunch tail.

This so-called agent can be, for example, the fields induced by the bunch in the surrounding accelerating structures. The field of the opposite colliding bunch can also play the role of this agent. We will study our examples starting with the latter.

### 10.3.2 Beam–beam effects

Modern electron–positron colliders aiming at the highest luminosity[4] require tiny beam sizes at their *interaction region* (IR). An example of the beam sizes for a 500 GeV CM $e^+e^-$ ILC collider project is shown in Fig. 10.20. Note that these beams are extremely flat, to ensure that the energy losses of the beams caused by synchrotron radiation during beam collision are significantly suppressed.

[4]International Linear Collider Technical Design Report, 2013.

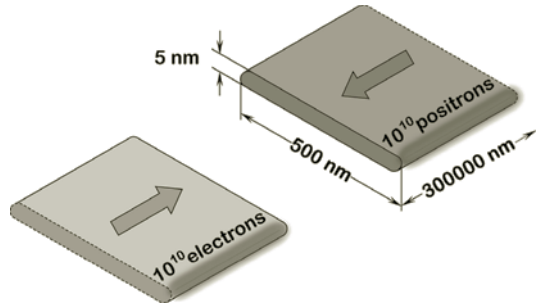



**FIGURE 10.20**
Flat beam collision in an IR of a typical linear collider.

For a beam with uniform density distribution and transverse sizes $\sigma_{x,y}$, the maximum field at the boundary of the beam can be estimated using the Gauss theorem $\int Eds = 4\pi Q$, which thus gives for this maximum field:

$$E \approx \frac{eN}{(\sigma_x \sigma_z)} \tag{10.20}$$

where we take into account that, for the beam with uniform distribution in the longitudinal direction, the surface integral $\int ds$ turns into the contour integral $\int d\ell$ with charge $Q$ taken per unit of length.

Taking the contour integral in the above derivation along the contour inside of the beam, we conclude that the field inside of the beam grows linearly. The fields outside of the beam can be estimated in the same manner. It is important to note that, if the beam is very flat, the integral would be almost independent of the offset from the beam in the $y$ direction until the point when the vertical offset reaches values comparable with $\sigma_x$.

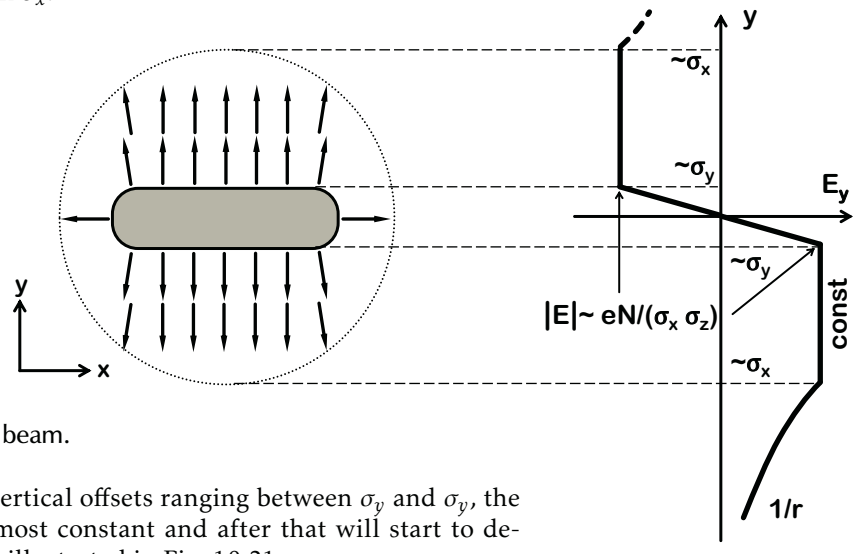



**FIGURE 10.21**
Fields of the flat beam.

Therefore, for vertical offsets ranging between $\sigma_y$ and $\sigma_y$, the field will be almost constant and after that will start to decrease as $1/r$, as illustrated in Fig. 10.21.

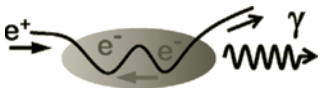



**FIGURE 10.22**
*Beamstrahlung.*

Fields of the beam cause synchrotron radiation in the particles of the opposite beam. As previously mentioned, this radiation is called *beamstrahlung* (Fig. 10.22). It results in an increase of the energy spread of the beams and smears out the luminosity spectrum. The produced SR photons create large numbers of $e^+e^-$ pairs that create *background* in the experimental detector. *Beamstrahlung* energy losses can be estimated using the formulas from Chapter 3. We will leave these estimations for your exercises while we return to the transverse dynamic effect.

The transverse fields of the flat beam estimated above produce an angular kick for a particle travelling through the opposite colliding bunch. For Gaussian transverse beam distribution, and for a particle near the axis, the total angular kick of the particle after collision is estimated as:

$$\Delta x' = \frac{dx}{dz} = -\frac{2Nr_e}{\gamma\sigma_x\ \ \sigma_x+\sigma_y}\cdot x$$

$$\Delta y' = \frac{dy}{dz} = -\frac{2Nr_e}{\gamma\sigma_y\ \ \sigma_x+\sigma_y}\cdot y \qquad (10.21)$$

We can now introduce the notion of the *Disruption parameter*.

$$D_x = \frac{2Nr_e\sigma_z}{\gamma\sigma_x\ \ \sigma_x+\sigma_y} \quad \text{and} \quad D_y = \frac{2Nr_e\sigma_z}{\gamma\sigma_y\ \ \sigma_x+\sigma_y} \qquad (10.22)$$

As follows from Eq. 10.21, the disruption parameter is inversely proportional to the focusing distance $f_{beam}$ corresponding to the focusing field of the bunch: $D_y \sim \sigma_z/f_{beam}$.

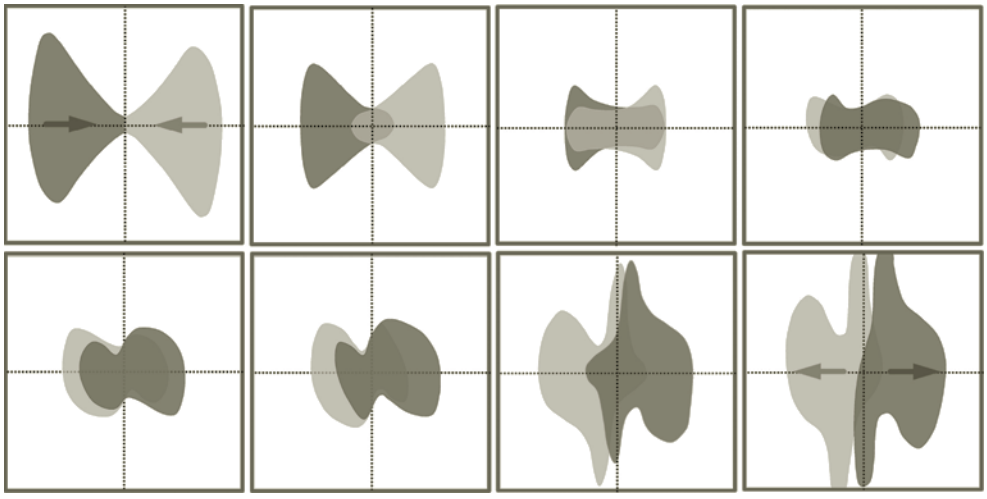


**FIGURE 10.23**
Consequent moments of high-disruption beam collision.

A low-disruption parameter $D \ll 1$ means that the bunch acts as a thin lens. Its high values ($D \gg 1$) link to the instance when particles oscillate in the field of the opposite colliding bunch. The number of oscillations is approximately equal to $\sqrt{D/(2\pi)}$ and if $D$ is around or bigger than $\sim 20$, there will be a couple or more oscillations during collision.

Imagine now that the beams had small transverse offsets before the collision. The beams would attract each other and start to oscillate. As the collision developed further, the new portions of the beam would already have larger initial displacements. The oscillations would grow in an unstable manner, limited only by the finite length of the bunch.

An example of beam–beam collisions with $D_y \sim 24$ and with an $0.1\sigma_y$ initial offset between the beams is shown in Fig. 10.23. We can see that the second half of the collision is indeed noticeably disrupted.

Collisions in high disruption regimes have both their challenges and their advantages. As we can note in Fig. 10.23, the middle part of the collision shows that the beams are nicely focused on each other. The beam–beam focusing forces the beams to squeeze tighter and thus give an additional enhancement to the luminosity, which is an advantage. However, beam–beam instability may develop, and therefore the luminosity enhancement would be compromised by a higher sensitivity to the initial offsets.

We will later discuss one more collision scenario where the disruption regime may become useful — in the method aimed to overcome the *hourglass effect* (see Section 10.6).

### 10.3.3 Beam break-up and BNS damping

The pursuit of high-charge and high-current beams leads to challenges caused by various imperfections. One particular issue is caused by *wakefields*.

The interaction of the charged beam with the accelerating structure, or the vacuum chamber in general, can generate

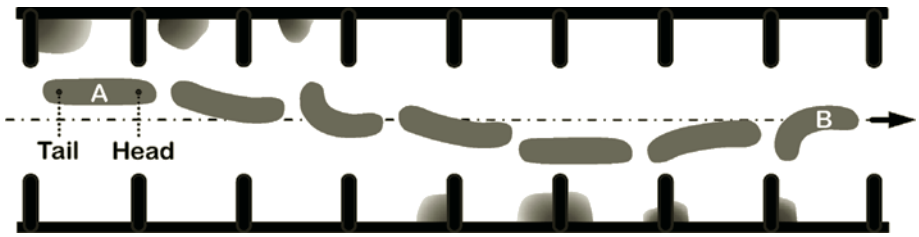



**FIGURE 10.24**
Beam break-up instability of a single beam. Fields left by the bunch are shown qualitatively. Beam evolution from the initial unperturbed shape (A) to the final BBU-distorted shape (B).

electromagnetic fields — wakefields — that can act back on the bunch itself.

In the RF accelerating structure, these fields can build up resonantly and disrupt the bunch itself. This is called the *beam break-up* (BBU) instability and it can happen either for a single bunch or in a multi-bunch case.

The single-beam break-up instability is illustrated in Fig. 10.24. The cure for the single-bunch BBU is called *BNS damping*, according to the names of its creators.[5]

The mechanism of how BNS damping mitigates single-bunch BBU is explained in the following: assume that the bunch has an offset with respect to the center of the accelerating cavity. The bunch head will excite a transverse dipole wakefield $W$ (proportional to the offset) that will cause transverse deflection of the tail, which can consequently result in BBU.

We can note that the wake $W$ acting on the tail is an additional *defocusing*. The BNS damping recipe is thus to introduce additional *focusing* to the bunch tail, which would then compensate the wakefield $W$ acting on the tail.

In order to compensate the wake $W$ acting on the tail, one needs to decrease the energy of the tail in such a way that the effectively increasing focusing via lenses in the accelerator channel exactly cancels out the wakefield's defocusing effect. The necessary energy difference between the head and the tail of the bunch is achieved for the BNS damping by placing a bunch off-crest in the RF pulse, which then creates corresponding optimal BNS energy spread over the bunch ($E - z$ correlation), as illustrated in Fig. 10.25.

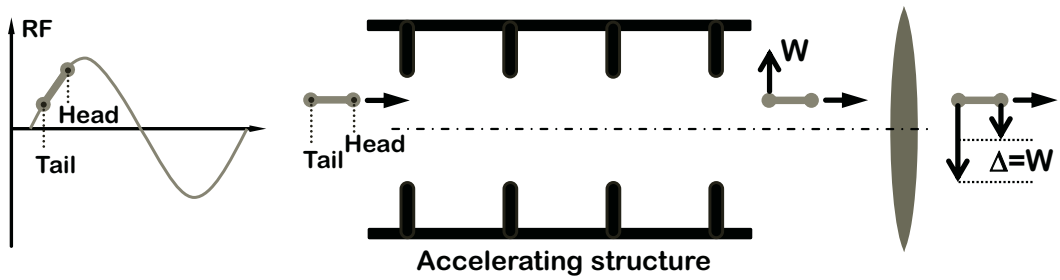



**FIGURE 10.25**
BNS damping method.

BBU can also occur in trains of bunches, wherein accumulated wakefields act on the next bunch and the following bunches in the train, enhancing their oscillation. A possible method to cure the multi-bunch BBS is to enhance *decoherence* of transverse modes that the beam excites. This can be

[5]V. Balakin, A. Novokhatsky and V. Smirnov, in Proc. of the 12th Int. Conf. on High Energy Accelerators, Fermilab, 1983.

achieved via gradual variation of the parameters (such as iris radii) of the individual cells in the accelerating cavity.

In this case, as qualitatively shown in Fig. 10.26, the time structure of excited modes will be slightly different, and the total sum wakefield will quickly lose coherence, minimizing the kick on the following bunch. Often, in addition to variation of accelerating cells, additional passive damping elements are inserted into the accelerating structures to absorb particularly the higher-order modes. Such structures are called *damped detuned structures* and can be suitable for accelerating trains of high-charge bunches with separation between bunches as minimal as just a few RF cycles.

Before moving on to the next section, we would like to make a few remarks. We have seen in this section that BNS damping prevents BBU instability by maintaining the conditions for coherent motion of the bunch head and tail (i.e., preventing relative oscillations of the head with respect to the tail). It is helpful to look at the word *damping*, in "BNS damping" (or the word *echo*, in "EEHG") broadly. The method to avoid BBU instability in multi-bunch cases involves making use of decoherence.

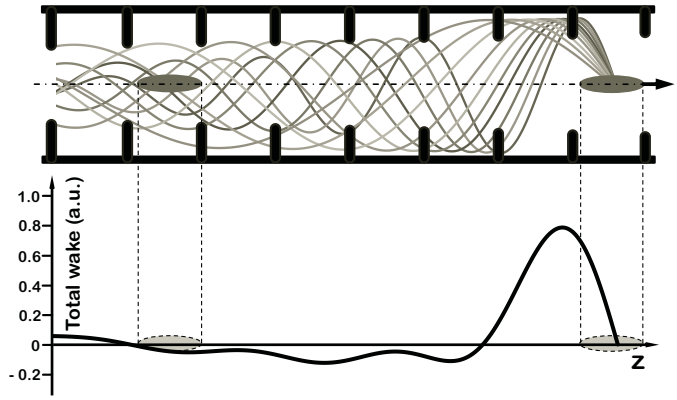



**FIGURE 10.26**
Detuned structure as a cure for multi-bunch BBU instability.

We will now look at one more mechanism — Landau damping, which is used both in plasma and accelerators, but has, in some cases, a different meaning within these two areas.

### 10.3.4 Landau damping

Landau damping[6] is the mechanism first discovered in plasma. It describes a collisionless damping of collective plasma oscillations. This mechanism can be illustrated via the following approximate analogy.

[6] L. Landau, J. Phys. USSR 10 (1946).

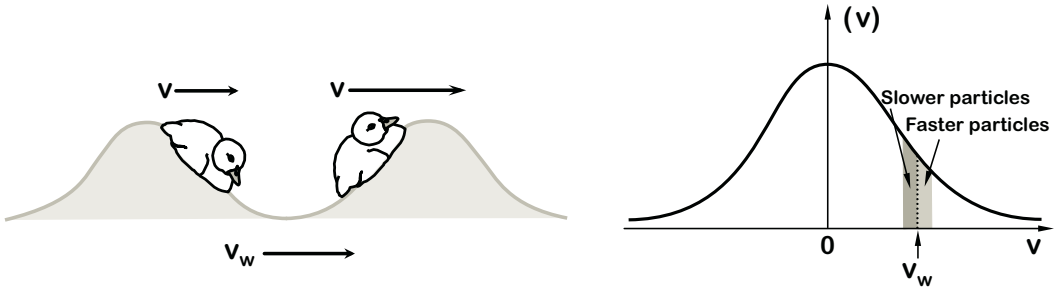



**FIGURE 10.27**
For illustration of Landau damping mechanism.

Let us consider an ocean wave that travels with phase velocity $v_w$. Assume that a duck floats along the wave with the velocity $v$, which is very close to $v_w$; see Fig. 10.27.

The duck would soon be "trapped" in the ocean wave, which means that if the duck was initially moving faster than the wave ($v > v_w$), it would slow down and thus the wave would gain energy from the duck. In the opposite case, if the duck was initially slower ($v < v_w$), the wave energy would be transmitted to the duck.

The overall damping of the wave can therefore occur if the distribution of velocities of ducks (or particles in plasma) decreases for larger values of velocity — which is indeed typically the case, as shown in Fig. 10.27 for Maxwellian distribution. As there are fewer faster particles than slower particles, the collective wave in plasma will be damped by the Landau mechanism.

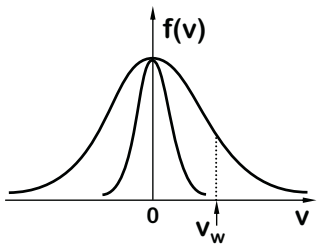



**FIGURE 10.28**
Velocity spread and Landau damping.

It is important to note that, for the Landau damping to be possible, the distribution function should have a nonzero number of particles (ducks) at the wave velocity $v_w$. Therefore, if the initial spread of velocities of ducks or particles was not sufficient, increasing it as shown in Fig. 10.28 would help to enhance Landau damping.

Landau damping is a very important mechanism for accelerators too, as it helps to provide beam stability for either transverse or longitudinal motion. Increasing the spread of certain beam parameters, as was just mentioned above, is often also done in accelerators, as a possible way to provide improved stability.

In particular, increasing the energy spread in circular accelerators would result in, via the nonzero momentum compaction factor, the spread of revolution periods, thus helping the longitudinal stability of the beam. Increasing the energy spread of the beam with a *laser heater* may help in coping with CSR-caused microbunching.

Spread of betatron frequencies (caused by energy spread and nonzero chromaticity) is often also helpful to the stability of the beams, as well as an additional spread introduced by the so-called *Landau octupoles* — the magnets often inserted

into the beamline specifically for the purpose of enhancing the tune spread and thus helping the beam stability.

We would like to mention here that increasing the spread of tunes or energy often helps accelerators simply because it enhances the decoherence, preventing particles from assembling into collective motion. The term *Landau damping* is sometimes used in cases where the term *decoherence* would be more appropriate.[7] Details of a particular physics setting define which of those two mechanisms is acting in each situation.

### 10.3.5 Stability and spectral approach

Beam or laser pulse stability issues are often analyzed based on a spectral approach. We will omit detailed and rigorous definitions here and focus only on a few highlights.

The disturbance — e.g., the position of a particular magnetic lens or optical element of a laser (we will define it as $x(t)$) — often has the same characteristics as a *random process*. In this case, the notion of the *power spectral density* $p(f)$ (dubbed "power spectrum") should be used instead of the usual Fourier spectrum.

The key property of the power spectrum is that its integral is equal to the variance $\sigma$ of the signal $x$:

$$\sigma^2 = \langle x^2 \rangle = \int_{-\infty}^{+\infty} p(f)\, df \tag{10.23}$$

where we assume that the average of $x$ is zero: $\langle x \rangle = 0$.

Disturbance to our beam or laser line often comes from certain frequencies. The power spectrum is useful here as it can identify the contribution from a particular frequency range according to the following (Fig. 10.29):

$$\sigma^2(f_1 < f < f_2) = \langle x^2 \rangle = \int_{f_1}^{f_2} p(f)\, df \tag{10.24}$$

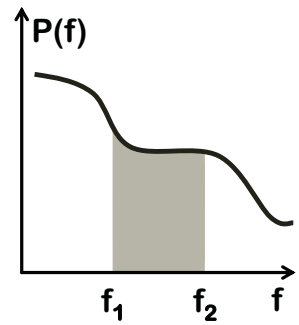



**FIGURE 10.29**
Power spectrum.

Quantitative analysis of the influence of disturbances often requires the knowledge of the expected relative displacement of a beamline element during a specific time duration $\tau$. The associated variance (contribution to it from a frequency range $[f_1; f_2]$) can also be calculated from the power spectrum:

$$\langle [x(t+\tau) - x(t)]^2 \rangle_t = \int_{f_1}^{f_2} p(f)\, 2[1 - \cos(\omega t)] df \tag{10.25}$$

You may notice that the description given above is not sufficient for analyzing beamlines that have many elements distributed in space. In addition to time, we also need to take spatial information into account.

[7] Werner Herr, CERN Accelerator School, 2013.

Therefore, our analysis should involve a spectrum that depends on frequency as well as on wavenumber — a *two-dimensional power spectrum*[8] $P(\omega,k)$.

The 2D spectrum (also called PWK spectrum) can evaluate expected relative displacements of two beamline elements, separated by a certain distance $L$ and after a certain time interval $T$. Assuming that at $t = 0$ the beamline was perfectly straight, the variance of the relative misalignment after a time $T$ of two points separated by the distance $L$ is given by

$$\sigma^2(T,L) = \int_{-\infty}^{+\infty} P(\omega,k)\, 4[1-\cos(\omega T)]\,[1-\cos(kL)]\frac{d\omega}{2\pi}\frac{dk}{2\pi} \quad (10.26)$$

Since the formula Eq. 10.26 can predict the stability of any two elements in our beamline, we can also evaluate the stability of the entire beamline by properly taking into account all of the elements.

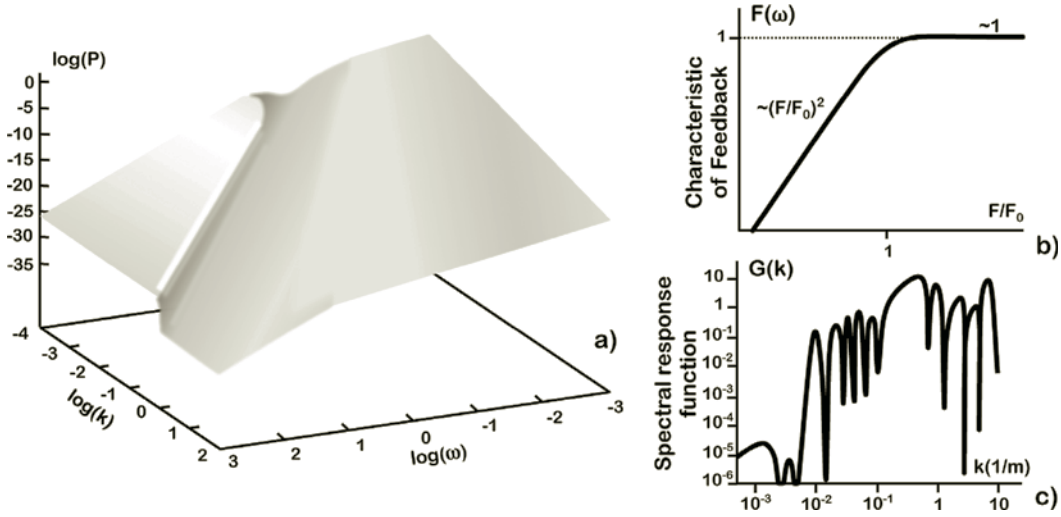



**FIGURE 10.30**
Examples of power spectrum $P(\omega,k)$ (left), spectral response function $G(k)$ and characteristic function of the feedback $F(\omega)$.

Combining the coefficients that determine how displacement $x$ of an individual beam element or laser line contributes to, for example, displacement of the beam/light at the focus $x_{out}$, we can construct a so-called *spectral response function* $G(k)$. The variance of the relative misalignment of the output beam/light after a time $T$ depends on $G(k)$ as

$$\sigma^2(T) = \int_{-\infty}^{+\infty} P(\omega,k)\, 2[1-\cos(\omega T)]\, G(k)\frac{d\omega}{2\pi}\frac{dk}{2\pi} \quad (10.27)$$

The variance of the misalignment defined above in Eq. 10.27 depends on the observation time $T$ and typically

[8] A. Seryi and O. Napoly, Phys. Rev. E. v.53, 5323, 1996.

grows with $T$. To stabilize the beam/light at the focus, a feedback would usually need to be applied, which can be described by the *characteristic function of the feedback* $F(\omega)$. In this case the variance of the relative misalignment of the output beam/light is a constant, and is given by

$$\sigma^2 = \int_{-\infty}^{+\infty} P(\omega,k)\, F(\omega)\, G(k) \frac{d\omega}{2\pi}\frac{dk}{2\pi} \tag{10.28}$$

Examples of a 2D power spectrum as well as a spectral response function and characteristic of feedback for a particular beamline (final focus system beamline) are shown in Fig. 10.30.

A particularly useful and insightful approximation for a 2D spectrum is the one related to the *ATL law*, which was first suggested[9] to describe space–time motion of ground (earth), on which the focusing elements of a beamline are installed.

For this "ATL law," the variance of the misalignment for elements separated by distance $L$ and observed after time $T$ (described above by Eq. 10.26) is simply

$$\sigma^2(T,L) = A \cdot T \cdot L \tag{10.29}$$

where $A$ is the parameter of the model that could be estimated, for example, from ground motion measurements of our site (where the elements of beamline are installed) or perhaps of our optical table (where the laser elements are mounted).

The 2D power spectrum corresponding to the "ATL law" is given by

$$P(\omega,k) = \frac{A}{\omega^2\, k^2} \tag{10.30}$$

The 2D spectrum shown in Fig. 10.30 includes the ATL component; however, it is included in a corrected way, at higher frequencies and wavenumbers, to ensure integrability of the equations.

The term "law" in "ATL law" should be taken cautiously here, as this is certainly a model, an estimate that is applicable in certain ranges of parameters and certain situations. It is, however, amazing how many different situations and wide parameter ranges[10] there are in which this approximation can be found valid.

## 10.4 Beam or pulse addition

The topic of beam or pulse addition is one where there is an important difference between the area of accelerators —

[9] B. A. Baklakov et al., INP Report 91–15, 1991.
[10] V. Shiltsev, Phys. Rev. Lett. 104, 238501 (2010).

where charged particles are used — and lasers, where we deal with light.

In this section, we will briefly discuss the methods of laser pulse addition in the optical cavities in comparison with injection methods used in accelerators, return to lasers to touch on a promising technique to coherently combine fiber laser signals, and finish with resonant plasma excitation by a train of laser pulses.

### 10.4.1 Optical cavities

Optical cavities consisting of two or more mirrors are a widespread element used in practically any laser or optics system.

Optical cavities, in the event when one of the mirrors is semi-transparent, are efficient tools for adding laser pulses together for their accumulation to higher intensities.

Various arrangements of optical cavities are possible, and some examples are shown in Fig. 10.31.


**FIGURE 10.31**
Examples of optical cavities. Plane-parallel (A), concentric/spherical (B) and confocal (C) configurations.

An optical cavity suitable for hosting interaction between an electron beam and a laser (for Compton sources in particular) can be composed of two or more mirrors.

One of the important concerns is the cavity's stability and the light pattern inside of it, and also how it reacts to small deviations in its parameters.

It has been found that, while two-mirror systems are pos-

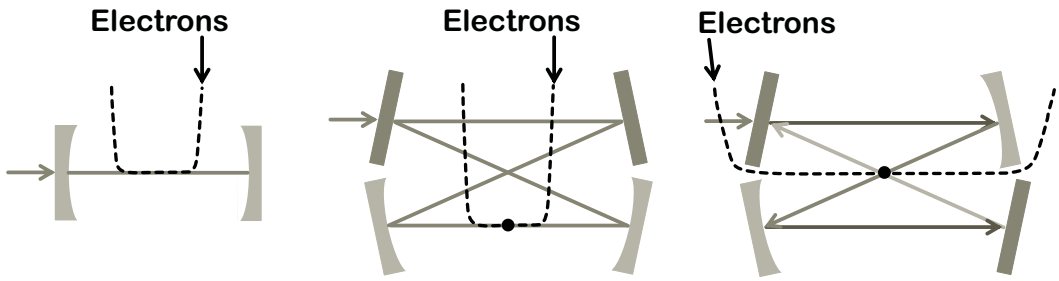



**FIGURE 10.32**
Examples of four-mirror optical cavity suitable for electron beam–laser interaction.

sible, the four-mirror optical cavities are more practical,[11] and are especially more impervious to disturbances.

Moreover, the four-mirror cavities can also be more suitably arranged when one needs to pass the electron beam around the cavities, as shown in the third case presented in Fig. 10.32.

There is an essential difference between the accumulation of pulses of light and the accumulation of bunches of particles, as we will discuss momentarily.

### 10.4.2 Accumulation of charged particle bunches

The accumulation of bunches of charged particles (e.g., at injection) obeys *Liouville's theorem* (see its full definition in the margin).

Liouville's theorem: *In the vicinity of a particle, the particle density in phase space is constant if the particle moves in an external magnetic field or in a general field in which the forces do not depend upon velocity.*

The consequence of the theorem is the conservation of emittance $\varepsilon$, i.e., the area $p\,dq$ of the phase space beam portrait is constant (during accelerating, $\gamma\varepsilon$ is constant).

Beam cooling methods (SR, electron, stochastic) do not contradict Liouville's theorem, as in these cases the cooling forces acting on the particle depend on its velocity.

One important consequence of the theorem is that we cannot stack bunches of charged particles on top of each other in the phase space. If we would like to accumulate them to increase the total intensity, we must inject the bunches into the neighboring areas of the phase space (Fig. 10.33).

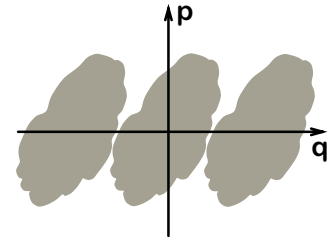



**FIGURE 10.33**
Phase-space stacking.

Phase-space stacking and accumulation of bunches during injection can be done in various ways. Fig. 10.34 shows an example of transverse stacking.

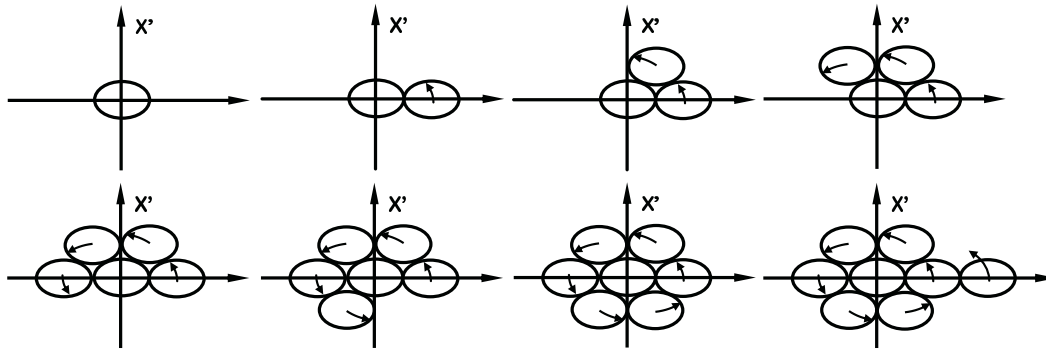



**FIGURE 10.34**
Transverse phase-space stacking. Consecutive moments.

Longitudinal stacking is also possible, as illustrated in Fig. 10.35. Fast-switching bending magnets or electrostatic deflectors — *kickers* — are used to perform bunch stacking during injection.

[11] A. Variola et al., ThomX — Conceptual Design Report, LAL, 2010.

There is one particular injection mechanism that can overcome limitations imposed by Liouville's theorem — the *charge-exchange injection*. In this case, $H^-$ ions are injected into the beamline through a thin foil. Ion interactions at the foil strip away two electrons and the ions are therefore converted (at the foil) to protons.

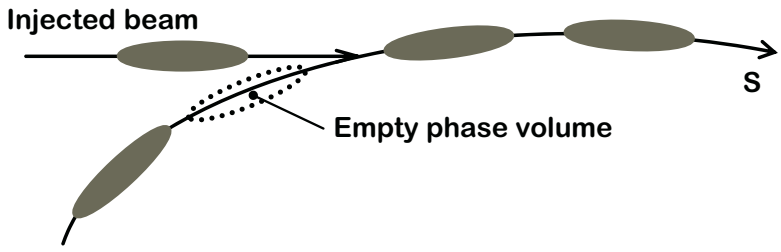



**FIGURE 10.35**
Longitudinal phase-space stacking.

This charge-exchange injection can be applied many times into the same phase space. Since the foil is inserted into the accelerator on the way to the main circulating beam (see Fig. 10.36), the number of injections is limited by the scattering of the circulating beam on this foil.

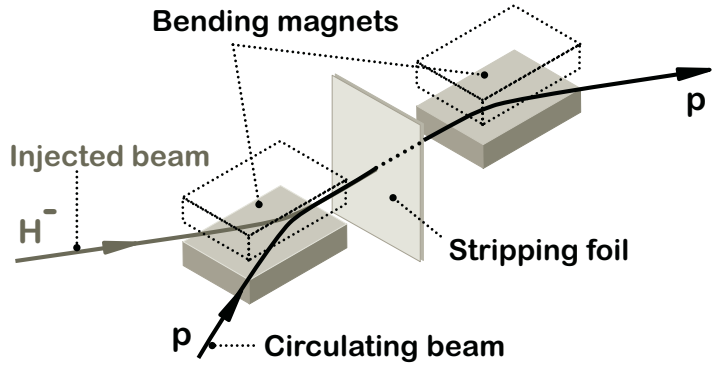



**FIGURE 10.36**
Charge-exchange injection.

### 10.4.3 Coherent addition of laser pulses

Returning to the laser theme, we would like to briefly mention a technique recently suggested[12] to coherently add laser pulses.

*Refer to Sections 4.2.3 and 4.2.4 for discussion of fiber laser repetition rate and TRIZ inventive principles.*

The technique is aimed in particular to fiber lasers that can produce pulses with high efficiency (tens of percent) and high repetition rates (tens of kHz and higher). The idea of

[12]Gerard Mourou et al., Nature Photonics 7, 258–261 (2013).

the technique, which is presently under development, is to coherently combine the output of many fiber lasers, with the goal to produce laser systems with high repetition rates, high energy and high efficiency. The concept is illustrated in Fig. 10.37.

The technique of coherent fiber laser pulse combination is striving to solve the deficiency of present high-peak power lasers (such as Ti:Sapphire or Nd:glass lasers) both of which have high peak powers but very low repetition rates (ranging from one shot per hour to around a Hz) and low efficiency (less than 0.01%).

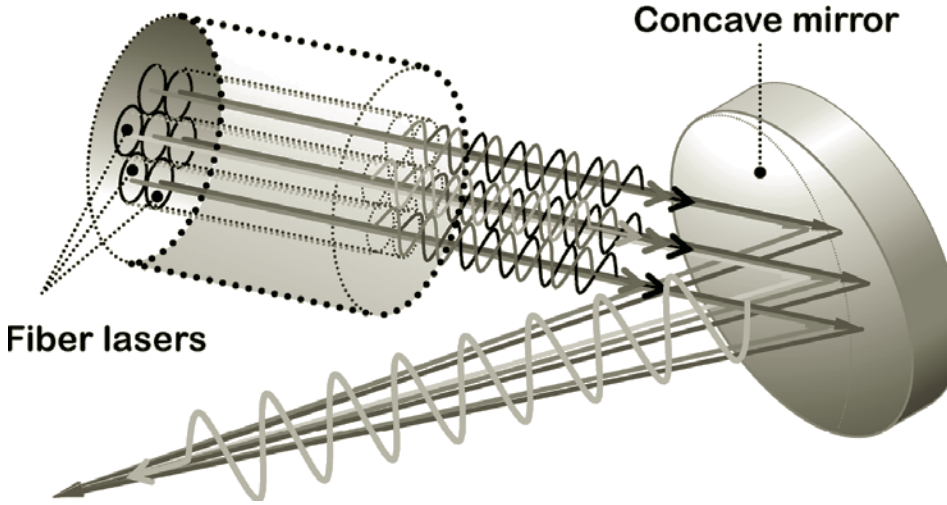



**FIGURE 10.37**
Concept of fiber laser coherent combination of pulses.

This technique, once developed, can serve as an efficient laser driver for laser plasma acceleration based light sources, which can be applied to the field of medicine, among others.

### 10.4.4 Resonant plasma excitation

We will mention one more accelerator technique that is currently being developed — *resonant plasma excitation*.[13]

In this case, the amplitude of plasma oscillation accumulates slowly, while plasma oscillations are excited by a train of low-energy laser pulses distanced at plasma wavelengths. Such a technique is particularly suitable for use in conjunction with fiber lasers because it can provide the driving laser trains with high efficiency and high repetition rates.

This technique, once finalized, can also offer the path to efficient laser-driven plasma acceleration based light sources (which have applications beyond accelerator physics).

[13] S.M. Hooker et al., J.Phys. B47 (2014) 234003.

## 10.5 Cooling and phase transfer

We will now examine several topics related to beam cooling and phase space manipulation.

### 10.5.1 Beam cooling methods

The three main methods of beam cooling (apart from SR) are shown conceptually in Fig. 10.38.

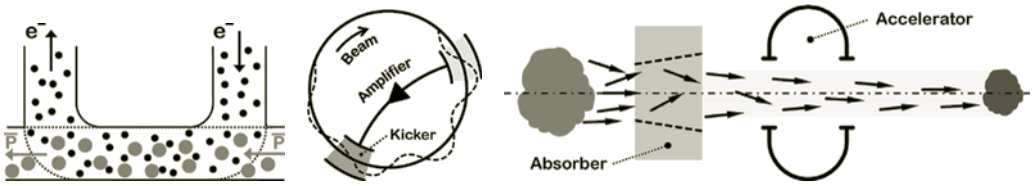



**FIGURE 10.38**
Electron cooling, stochastic cooling and ionization cooling concepts.

In addition to electron and stochastic cooling, which we have already discussed, we add *ionization cooling* — which is based on subjecting particles to ionization losses in an absorber, and then restoring the longitudinal component of the particles' momentum.

Stochastic and electron cooling are well-developed techniques. In particular, stochastic cooling was essential for the discovery of W and Z Bosons (C. Rubbia and S. van der Meer, 1984 Nobel Prize in Physics). Electron cooling was most recently used for improving the operation of the Tevatron collider.

Conversely, ionization cooling is a concept still under development. Conceptually simple, it is quite challenging technologically. However, this technique might be the only way to reduce the emittance of short-lived particles such as muons.

### 10.5.2 Electron cooling, electron lens and Gabor lens

In order to stimulate our TRIZ-inspired discussion, we will now draw a parallel between the three different techniques, which have completely different purposes but have, nevertheless, similarities.

Electron cooling, as we discussed, uses an electron beam co-propagating with the proton beam with equal velocities in order to enable their energy exchange. The electron beam is guided by a magnetic field. It is essential for the e and p velocities to be equal. The conceptual layout of an electron cooler is shown in Fig10.39.

An electron lens[14] is geometrically a very similar device wherein the proton (or antiproton) beam passes through the electron beam and receives additional focusing due to the fields of the electron beam; the lens aims to mitigate variation of betatron tunes between different bunches of proton beams. As the main effect of the lens comes from the electric field of the electron bunch, the direction of the e-beam travel is not essential (usually, counter-propagating beams are used). The e-lens concept is shown in the same Fig. 10.39.

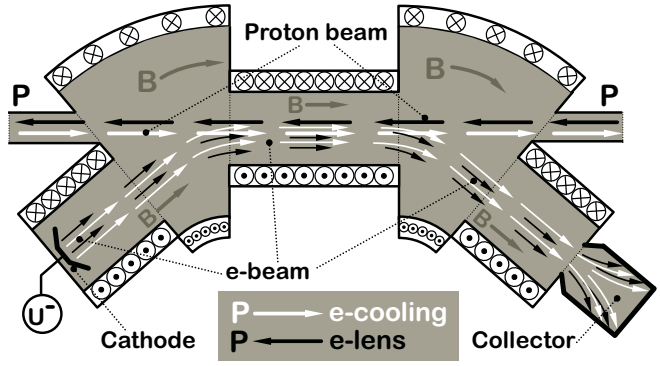



**FIGURE 10.39**
Electron cooling or electron lens.

We will also mention one more arrangement involving an electron beam and a magnetic field — the *Gabor lens*. Suggested[15] in 1947, this technique recently attracted attention[16] in connection to its use for laser plasma ion acceleration, for energy selection of accelerated protons.

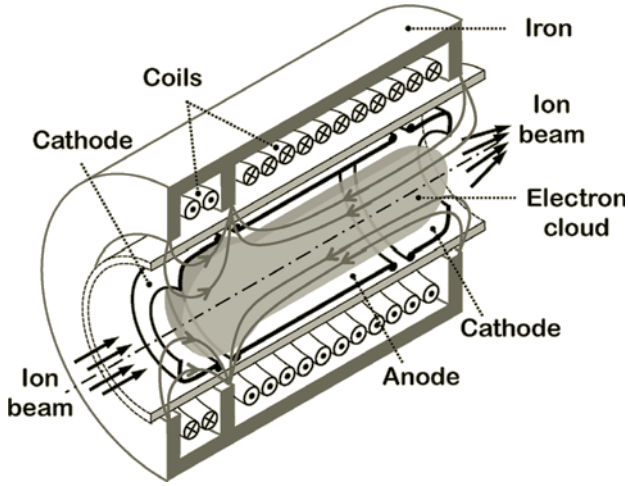



**FIGURE 10.40**
Conceptual schematic of a Gabor lens.

[14]V. Shiltsev et al., PR-ST-AB 2, 071001 (1999).
[15]D. Gabor, Nature 160, 89–90 (19 July 1947).
[16]J. Pozimski et al., Laser and Particle Beams, v. 31, 04, 2013, pp. 723–733.

A conceptual schematic of a Gabor lens is shown in Fig. 10.40. In this lens, the electron beam is formed by a cathode with a hole in the center. Electrons trapped in the center of the lens are attracted to the anode and are contained by the magnetic field.

A Gabor lens can potentially accumulate a large electron charge. In a steady state, the electrons rotate around the axis and the electrostatic repulsion — together with the centrifugal force — will balance the radial Lorentz force produced by the magnetic field. Using this assumption, the maximum density of electrons in the Gabor lens can be estimated as:

$$n = -\frac{B^2}{8\pi m_e c^2} \tag{10.31}$$

which can be high enough to be considered for use as a lens created by the radial field of the stored electrons.

The electrons stored in the Gabor lens move longitudinally in both directions. In Fig. 10.41, we summarize the relationships between the velocities of the proton (antiproton, ion) beams and the velocities of electrons in these three devices.

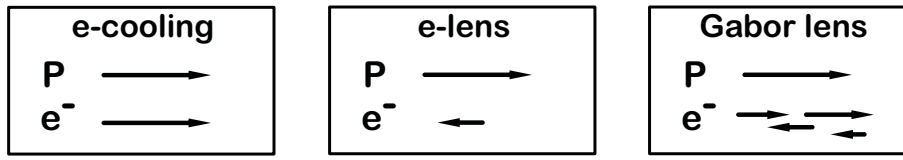



**FIGURE 10.41**
Relations of velocities of proton and electron beams in different configurations: electron cooling, electron lens, Gabor lens.

In the above, we have considered three systems that involve electron beams or stored electrons. These systems are aimed at different applications and details of their functions are very different. We assume, however, that it is useful to look at these systems simultaneously, as discussion of analogies between these systems is helpful for our TRIZ analysis (which runs in parallel with the main accelerator-laser-plasma story of this book).

### 10.5.3 Laser cooling

We conclude this section on cooling methods with laser cooling of ions. The method relies on the *Doppler frequency shift*, which is the key to laser cooling.

The cooling process involves the following steps (see Fig. 10.42). First, the laser photon coming from a certain direction hits the atom and is absorbed. The excited atom then re-emits a photon into a random direction.


**FIGURE 10.42**
Laser cooling steps. Absorption of a photon by an atom (a); excited state of the atom (b); emission of a photon (c).

The laser is in resonance with the atoms only when they are moving towards the laser, but not if they are moving sideways or away as shown in Fig. 10.43. This condition ensures the eventual cooling of the ions.

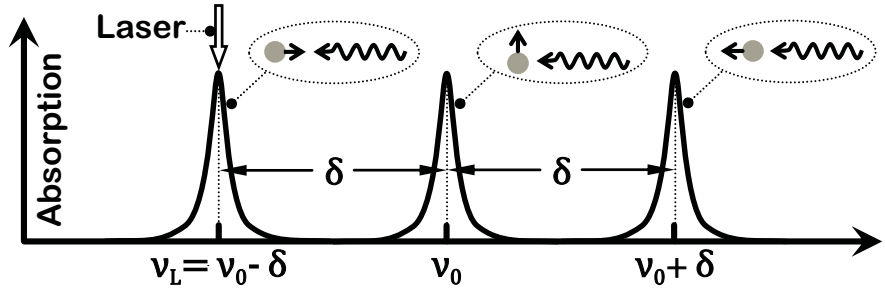



**FIGURE 10.43**
Relation between laser wavelength and Doppler shifted resonance absorption of an atom moving in different directions.

## 10.6 Local correction

In accelerators — as well as in any other technical fields — any unwanted disturbances are corrected either locally or non-locally. The universality of the approaches can also be observed with help from TRIZ, which includes relevant inventive principles (principle of *preliminary anti-action*).

Non-local correction is often used when it is not possible to correct the disturbance at its origin. The general issue with this approach lies in its non-locality — a correction needs to propagate and be properly preserved from the point of preliminary correction to the point where it needs to act.

Local corrections, if they can be used, are often superior, as they correct the disturbance at the origin. In this section, we will discuss several examples of local corrections.

### 10.6.1 Final focus local corrections

A *final focusing system* (FF) of any collider is aimed to produce a small beam size at the *interaction point* (IP). The final lenses of the system (usually arranged in a *final doublet* — FD) are the strongest and produce the largest chromaticity. The values of chromaticity are usually very large (if left uncorrected,

they increase the beam size at IP by orders of magnitude) and need to be compensated.

In a traditional final focus design, the chromaticity correction is performed in dedicated optical sections upstream from IP. In these sections, sextuple magnets are installed in dispersion sections and create some chromaticity there, which will then propagate to the FD, where they will cancel the FD chromaticity.

The performance of such an FF with non-local chromaticity compensation is often limited, as various additional disturbances (e.g., the SR-generated energy spread), which occur between the chromatic correction section and the FD, can alter the cancellation conditions.

An alternative design[17] exists: a final focus with local chromatic correction, shown conceptually in Fig. 10.44.

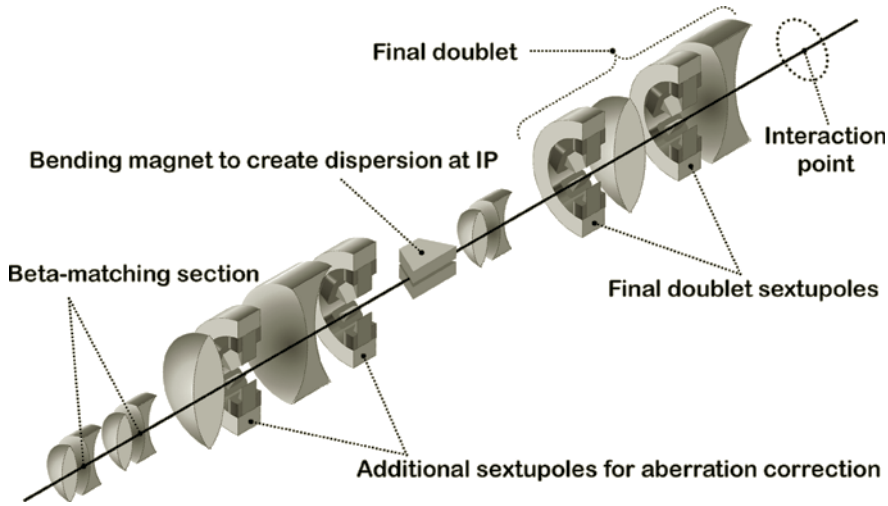



**FIGURE 10.44**
Final focus with local chromaticity correction.

In this design, chromaticity is canceled locally by two sextupoles interleaved with FD, while a bend upstream generates the dispersion (which is necessary for the sextuples to generate chromaticity) across the final doublet. The value of dispersion in the FD is usually chosen so that it does not increase the beam size in the FD by more than 20 to 30% for a typical energy spread of the beam.

Local chromatic compensation by FD sextuples needs to compensate the FD chromaticity without introducing other unwanted aberrations. In a vertical plane, the chromaticity compensation is straightforward, and therefore the vertical FD sextupole should be tuned to completely cancel the vertical FD chromaticity.

However, the FD horizontal sextupoles also introduce hor-

[17] P. Raimondi and A. Seryi, Phys. Rev. Lett., 86 (17), 3779, (2001).

izontal second-order dispersion. Second-order dispersion is produced by FD quadrupoles as well as FD sextupoles, while the latter produce just half as much of second-order dispersion as the FD quadrupole. Therefore, if we tune the FD horizontal sextupole to cancel the FD chromaticity, half of the uncompensated second-order dispersion will remain.

A solution to this issue consists of allowing the beta-matching section (shown schematically in the beginning of the beamline in Fig. 10.44) to produce as much horizontal chromaticity as the final doublet, so that the horizontal FD sextupoles will run twice as strong and simultaneously cancel the second-order dispersion and horizontal chromaticity.

Geometric aberrations to the final doubled sextupoles are cancelled in this design by two more sextupoles placed in phase with the FD sextupoles and upstream from the bend.

The final focus system can be reversed and the IP considered as an entrance point that captures a strongly diverging beam, such as the one coming out of the laser plasma accelerating bubble. This beam will also have an energy spread and associated chromaticity effects in the capture optics. A very small but divergent beam corresponds to a small beta function at the origin; therefore the chromatic effect for such a beam should be significant. The local chromaticity correction approach can therefore be applicable also to the laser-plasma capture optics.

### 10.6.2 Interaction region corrections

The interaction region of linear colliders exhibits several design inventions that can illustrate several TRIZ inventive principles.

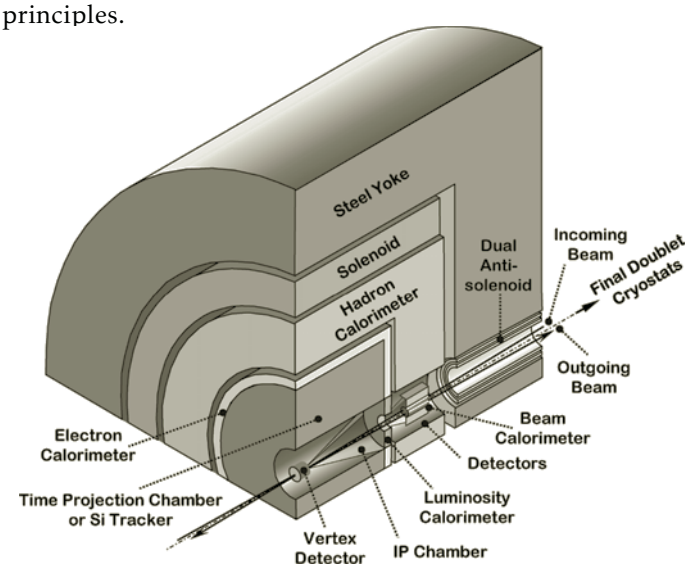



**FIGURE 10.45**
Conceptual layout of experimental detector and beamlines in the interaction region of a linear collider.

A typical layout of a detector and beamlines is shown in Fig. 10.45. One can see here that an anti-solenoid is inserted into the beamline. This anti-solenoid is needed to cancel the beam coupling effects that the main solenoid produces on the beam.

The problem with the anti-solenoid is that a huge force will be exerted on it because of the main solenoid. A solution implemented[18] in IR design is to use a dual anti-solenoid, where the outer coil with current $I_2 = -I_1 (R_1/R_2)^2$ cancels the external field; this makes it force-neutral (see Fig. 10.46).

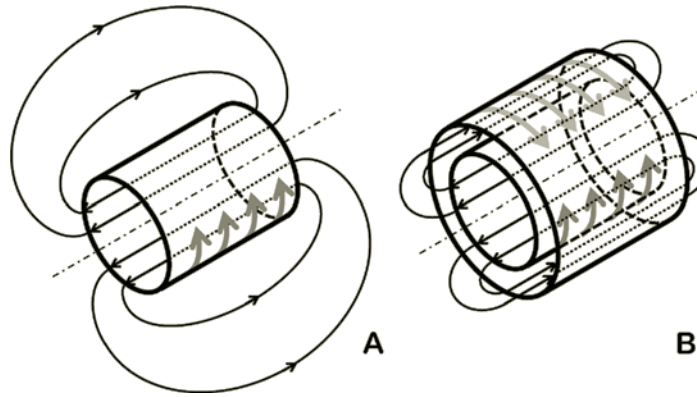



**FIGURE 10.46**
Standard solenoid (A) and interaction region dual solenoids (B).

From the point of view of TRIZ, this dual solenoid is an example of application of the nested doll and system–anti-system inventive principles.

### 10.6.3 Travelling focus

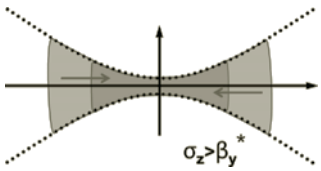



**FIGURE 10.47**
Hourglass effect.

The travelling focus regime[19] of beam collisions employs beam–beam focusing strengths in order to overcome the hourglass effect.

The *hourglass effect* (see Fig. 10.47) prevents reduction of the beta function at the IP to values lower than the length of the bunch, as further increases in focusing distort the bunch (into an hourglass shape) without increases to the luminosity.

In a travelling focus regime, the beam is focused to $\beta_y \ll \sigma_z$ and the location of focus dynamically changes during the collision. The focal point of the colliding bunches is made to coincide with the location of the head of the opposite bunch. This helps to optimally use additional focusing due to beam–beam forces and keeps the beams properly focused on each other during the entire collision.

[18] B. Parker, ca. 2002. A similar solution is used in nuclear magnetic resonance (NMR) scanners.
[19] V. Balakin, ca. 1991.

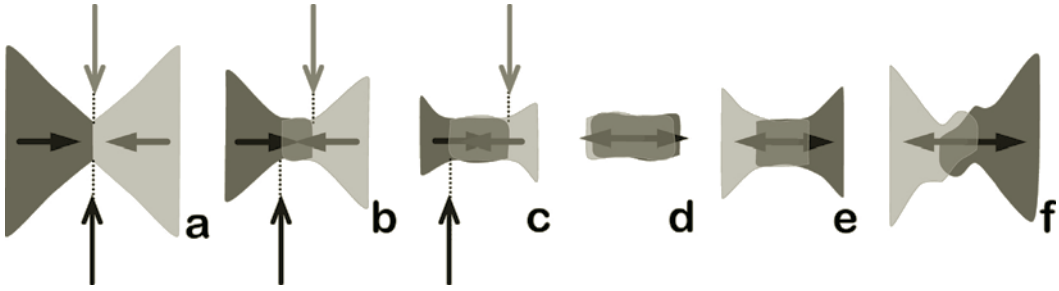



**FIGURE 10.48**
Travelling focus collisions.

Fig. 10.48 shows a simulation of traveling focus. The arrows show the position of the focus point during collision. This method hasn't yet been experimentally tested. One of the particular difficulties to this method consist of an increased sensitivity to imperfections, i.e., to the initial beam offset. This example gives us, however, additional information for a TRIZ-like analysis of the local correction approaches.

### 10.6.4 Crabbed collisions

The interaction region shown in Fig. 10.45 involves a collision of beams with a certain nonzero crossing angle $\theta_c$.

With crossing angle $\theta_c$, the projected $x$-size is

$$\sqrt{\sigma_x^2 + \theta_c^2 \sigma_z^2} \approx \theta_c \sigma_z$$

Taking the IP beam parameters shown on Fig. 10.20 and taking $\theta_c \approx 15$ mrad, we can conclude that the projected horizontal size is equal to several micrometers — which is several times larger than the nominal size. This is illustrated in the upper part of Fig. 10.49, where incomplete overlaps of the beams are apparent.

The crossing angle collision will result, therefore, in a substantial (by several times) reduction in luminosity, unless this effect is locally compensated.

Compensation of the crossing angle effect can be achieved by giving the bunch a z-correlated kick in such a way that the beam starts to rotate in the horizontal plane and arrives at the IP properly overlapping with the opposite beam — as shown at the bottom part of Fig. 10.49.

The z-dependent kick on the bunch can be produced by a special transverse mode cavity called a *crab cavity*, which does not disturb the central particle of the beam, but kicks the head and the tail particles in opposite directions in the x-plane. The design of a crab cavity and its fields are shown in Fig. 10.50.

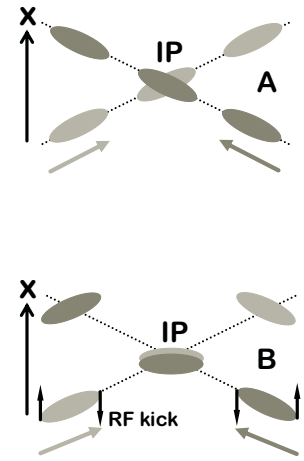



**FIGURE 10.49**
Collisions of the beams with crossing angle at the IP. Normal (A) and crabbed (B) collisions.

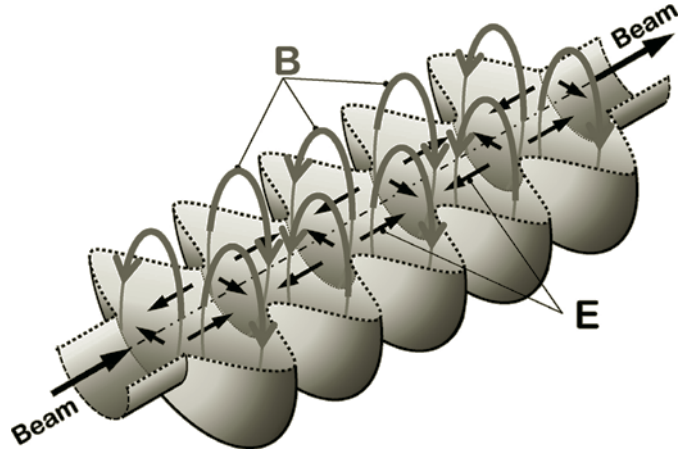



**FIGURE 10.50**
Crab cavity and its fields.

### 10.6.5 Round-to-flat beam transfer

In the final section of this chapter we will discuss round-to-flat or flat-to-round beam transformation, suggested by Y. Derbenev.[20]

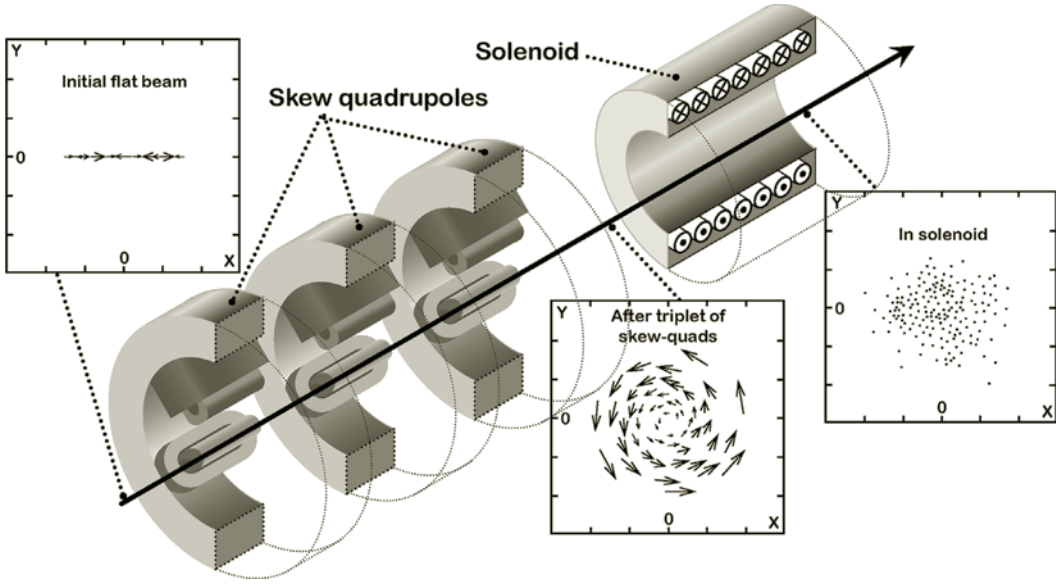



**FIGURE 10.51**
Beamline magnetic elements and phase-space portraits of the beam subjected to flat-to-round beam transformation. Initial flat beam, vortex, parallel beam in the solenoid.

[20]Ya. Derbenev, Adapting Optics for High Energy Electron Cooling, University of Michigan Report No. UM-HE-98–04, 1998.

This method is primarily intended for the beams that have significant asymmetries between phase planes. In particular, flat beams are often generated in accelerators. For example, SR rings naturally have $y$ emittance much smaller than the $x$ emittance.

It has been shown by Derbenev that a triplet of skew quadrupoles can transform such a flat beam to a vortex, as illustrated in Fig. 10.51.

The beam with vortex-like phase space distribution can be transformed further, noting that an edge of the solenoid also creates a vortex-like phase space portrait for an initially parallel beam.

Therefore, sending the vortex beam created by the skew triplet into the solenoid with appropriate strength will result in creating a round beam with zero angles — as shown in the final step in Fig. 10.51.

The transformation described above can be used in many accelerator systems. A particular example[21] uses this transformation to reduce space charge effects of flat beams in damping rings with long straight sections by employing flat-to-round and round-to-flat transformations at the beginnings and exits of each section, respectively. The round beam in straight sections will have lower space charge effects, improving the beam stability.

[21] R. Brinkmann, Ya. Derbenev and K. Floettmann, Phys. Rev. ST Accel. Beams 4, 053501 (2001).

## EXERCISES

10.1 *Chapter materials review*.
Describe the typical approaches for preventing beam break-up instability in linacs. Discuss if these approaches can be used in other situations considered in this book.

10.2 *Chapter materials review*.
Using the formulas in Chapter 3, derive an estimate for SR energy losses (*beamstrahlung*) for colliding $e^+e^-$ bunches. Justify the need to use flat beams in a high energy $e^+e^-$ collider. Estimate the energy losses for beam parameters shown in Fig. 10.20.

10.3 *Chapter materials review*.
For the same beams as in the previous exercise, estimate the number of emitted *beamstrahlung* photons per particle in the case when the oncoming positron bunch has either a 3 nm or 30 nm vertical offset with respect to the electron bunch.

10.4 *Mini-project*.
Define approximate parameters of a 1 GeV compact SR source ring aimed at 10 keV X-rays, based on top-off, on-orbit and on-energy injection by a laser plasma acceleration system. Discuss the research steps required for implementing such a concept.

10.5 *Analyze inventions or discoveries using TRIZ and AS-TRIZ*.
Analyze and describe scientific or technical inventions described in this chapter (e.g., coherent combination of laser pulses) in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

10.6 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you already know about accelerator science, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

# 11 Inventions and Innovations in Science



At last, we have reached the final chapter.

We have seen, throughout this book, that the three technical fields we discussed — accelerators, lasers and plasma — sometimes have a lot in common, things such as scientific approaches and solutions, similar inventions and technological breakthroughs. We have seen several examples when the TRIZ, or AS-TRIZ, inventive principles can be retrospectively identified in various scientific inventions.

In this final chapter, we would like to return to the topic we began to discuss in the very first chapter of this book — the methodology of inventiveness.

We will start by reviewing the updated Accelerating Science TRIZ principles.

Following that, we will then look critically at some of the suggested AS-TRIZ inventive principles, and discuss whether these are indeed principles or rather trends of the evolution of scientific systems. In this discourse, we will recall the definition of the laws of the evolution of technical systems as described by the standard TRIZ principles. After observing the parallels between the evolution of radar and lasers, we will conclude that some of the suggested AS-TRIZ inventive principles should be redefined as the laws of the evolution of technical/scientific systems.

We will then continue to consider the inventive principles and the laws of the evolution of systems by taking a very general approach, and will look at various fields of science and even beyond those.

The next topic we will discuss is the way to take these scientific inventions and apply them in practical settings. We will look at the linear and two-dimensional models of innovation and discuss the notion of the Pasteur quadrant in application to accelerator science in particular.

The well-known challenge of "crossing the Valley of Death" in technological innovation will be our next topic, and we will suggest a possible method for crossing this so-called valley in regards to the field of sci-tech, with help from a selection of compact light source projects.

In the next section we will offer a possible approach of using TRIZ to teach future generations of science students. In fact, the method we have used throughout this book has not been the ready-to-use standard TRIZ; instead, we immersed

DOI: 10.1201/b18696-11

ourselves in the proactive process of developing an extension of TRIZ — Accelerating Science TRIZ.

And finally, we will aspire to apply these developed principles to new, challenging projects and their unsolved problems.

## 11.1 Accelerating Science TRIZ

First of all, we hope that the reader has already guessed the play of words in the term "Accelerating Science TRIZ." The word "accelerating" does not refer here to accelerators — the devices that help particles reach higher energies. Instead, the word "accelerating" in AS-TRIZ implies that science itself can be accelerated via the application of these inventive principles.

Throughout this book we have seen several recurring inventive principles that can be added into the tables of the AS-TRIZ extension. Shown below are the updated tables where a number of parameters for the AS-TRIZ matrix of contradiction (Table 11.1), as well as several inventive principles (Table 11.2) for the AS-TRIZ, are included.

**TABLE 11.1**
Updated table of emerging AS-TRIZ parameters

| No. | Parameter |
|---|---|
| 1. | Energy |
| 2. | Rate of energy change |
| 3. | Emittance |
| 4. | Luminosity |
| 5. | Brightness |
| 6. | Intensity |
| 7. | Efficiency |
| 8. | Power |
| 9. | Integrity of materials |
| 10. | Time duration or length |
| 11. | Spatial extent |
| 12. | Sensitivity to imperfections |
| 13. | Cooling rate |
| 14. | ... |
| ... | ... |
| 21. | ... |

**TABLE 11.2**
Updated table of emerging AS-TRIZ inventive principles

| No. | Principle |
|---|---|
| 1. | ... |
| 2. | ... |
| 3. | Undamageable or already damaged |
| 4. | Volume-to-surface ratio |
| 5. | Local correction |
| 6. | Transfer between phase planes |
| 7. | From microwave to optical |
| 8. | Time–energy correlation |
| ... | ... |
| 21. | ... |

In addition to the principles of "undamageable or already damaged" and "volume-to-surface ratio" mentioned already in Chapter 1, we have included the inventive principle related to local correction (based on various examples from the focusing systems in beams or lasers), transfer between

phase planes (based, for example, on Derbenev's transformation), transition from microwave to optical (recall radar and CPA, or conventional and laser acceleration), and using time–energy correlation (laser pulse or bunch compression).

The reader can perhaps already add other inventive principles to this list — in fact, one of the purposes of keeping the unfilled spaces in the tables above is to fill them in, during the proactive process of learning.

## 11.2 Trends and principles

Standard TRIZ describes not only the inventive principles, but also the laws (trends) of the evolution of technical systems. Knowledge of these evolution laws can help to create innovative inventions.

There are just a few laws/trends of the evolution of technical systems and they can be described in broad terms. The inventive principles, however, are more concrete, as they are unique approaches that help to solve contradictions — there are several dozens of these.

The lists of additional AS-TRIZ inventive principles described in Table 11.2 should not be interpreted as cast in stone. Compiling such tables should instead be considered as a process of proactive learning parallel to the application of the TRIZ method to science. A critical analysis is therefore appropriate.

With this notion in mind, we would now like to look more carefully at the inventive principle "from microwave to optical" that we introduced in Table 11.2, and discuss whether this is indeed a principle or rather a general law (trend) of the evolution of technical systems.

### 11.2.1 TRIZ laws of technical system evolution

Standard TRIZ defines three types of laws of technical system evolution — *static*, *kinematic* and *dynamic*.

The three static laws are as follows. The TRIZ law of the *completeness of the parts of the system* states that a system should have the following four parts: an engine, a transmission, a working unit and a control element. The law of *energy conductivity of the system* says that every technical system is a transformer of energy and the energy should circulate freely and efficiently through these four main parts of the system. The law of *harmonizing the rhythms of the parts of the system* suggests that the frequencies and periodic motions of these parts should be in sync with each other.

The three kinematic laws are defined in standard TRIZ as follows. The law of *increasing the degree of ideality of the system* states that the "ideality," which is a qualitative ratio between

all desirable benefits of the system and its cost or other harmful effects, should have the tendency to increase. The law of *uneven development of parts of a system* suggests that the different parts of a technical system will evolve differently, leading to new technical and physical contradictions. And finally, the *law of transition to a super-system* states that a system which has exhausted the possibilities of further significant improvement is included in a super-system as one of its parts.

Finally, the two dynamic laws are as follows. The law of *transition from macro to micro level* suggests that the development of working organs proceeds initially on a macro and then more and more on a micro level. And the law of *increasing involvement of fields in the system* states that the fields evolve from mechanical fields to electro-magnetic fields.

We note that the phrase in the previous paragraph, "the fields evolve from mechanical fields to electro-magnetic fields," can be extended and clarified as meaning that, within the electro-magnetic spectrum, the evolution should follow from RF frequencies to optical frequencies.

We can conclude from this comparison and discussion that "from microwave to optical," which we initially identified as an inventive principle, is in fact better suited to be defined as a general trend of technical/scientific system evolution.

Let's highlight this conclusion by using an example of the invention of CPA in connection to radar.

### 11.2.2 From radar to high-power lasers

In 1935, Robert Watson-Watt developed the RDF (Radio-Direction Finding), which later become known as RADAR (RAdio Detection And Ranging). Development of the radar device took place at Ditton Park in England, which later became known as the Appleton Laboratory and, after merging with the Rutherford Laboratory, became the Rutherford Appleton Laboratory — currently the home of many accelerators and lasers (and located just a half-hour drive from Oxford — the place where this book was written).

The operation of radar involves chirped pulse amplification. The positive influence of radars on the span of the 20th century cannot be discounted. Indeed, we would not have been party to the tremendous and revolutionary advances in laser technology had it not been for the chirped pulse amplification — CPA, developed around 1985 by D. Strickland and G. Mourou.[1]

The development of CPA stimulated an exponential growth of available laser powers. From around 1990 to the

[1] D. Strickland and G. Mourou, Compression of amplified chirped optical pulses. Opt. Commun. 56, 219–221 (1985).

early 2000s, the peak laser power increased by more than two orders of magnitude, reaching PW levels.

High-power lasers, in their turn, enabled progress in many technical and scientific areas. For instance, nuclear physics, the production of Giga–Gauss magnetic fields, fusion science, studies of material properties in extreme conditions, laboratory astrophysics, studies of turbulence, nonlinear quantum electrodynamics and many other fields benefitted from the development of CPA lasers.

The example of radar and CPA once more asserts TRIZ's founding proposal that the same problems and solutions appear again and again, but in different disciplines. It also illustrates the general trend of technical/scientific system evolution (i.e., the transition from the microwave part of the EM spectrum to the optical range).

### 11.2.3 Modern laws of system evolution

The laws/trends of the evolution of systems were defined by standard TRIZ several decades ago. The laws are very broad and will likely remain valid for a long time. Still, it is tempting to look at these laws from a present-day perspective and, initially, see whether the laws need to and/or can be updated, and, secondly, attempt to apply these laws to other areas of science, and also to seemingly unrelated areas.

In order to illustrate the tendencies that can be redefined as the laws of the evolution of scientific systems, let us recall the example of tumor therapy wherein the effects of a proton beam can be combined and synergistically enhanced by using small compounds. The latter can be created specifically for a particular patient, based on a genetic analysis of their tumor.

This example shows, first of all, the trend of using more than one technique for achieving synergy and a superior overall effect. And second, this illustrates the universal trend of developing solutions and services tailored to specific persons.

The reader can certainly suggest other examples that illustrate these or other trends. The fascinating question of whether the noted trends are indeed modern general laws of scientific system evolution is better left for further discussion outside of the framework of this text.

## 11.3 Engineering, TRIZ and science

There are many well-founded connections between science and engineering. We will take this time to discuss several particular aspects of these connections while highlighting their link to TRIZ methodology.

### 11.3.1 Weak, strong and cool

Recall our discussion relating to the focusing in accelerators that is needed to keep particle trajectories near the center. The first accelerators had weak focusing with spatial periods greater than the perimeter of the accelerator. The trajectories of particles in a weak focusing accelerator had large transverse excursions, requiring extremely large vacuum chamber apertures.

Strong focusing was suggested as a better way to focus particles by involving the usage of a sequence of focusing–defocusing elements. This example in itself may serve as an illustration of the TRIZ inventive principle of system and antisystem.

*Strong FODO focusing may illustrate TRIZ inventive principle system–antisystem.*

When strong focusing was suggested, it was not totally clear if it would work in practice. The realization of the first strong focusing proton accelerator was led by Sir John Adams, who had the courage to cancel (in October, 1952) the already approved 10 GeV weak focusing accelerator for a totally innovative 25 GeV *Proton Synchrotron*.

The risk was worth it — in 1959, Sir John Adams announced that CERN's Proton Synchrotron had just reached 24 GeV and had surpassed Dubna's Synchrophasotron world record of 10 GeV.

Another example of scientific risk-taking relates to the development of electron beam cooling (which was especially necessary for antiparticles such as antiprotons). The method was suggested by G.I. Budker, the founder and first director of the Institute of Nuclear Physics in Novosibirsk. Budker was the mind behind many inventions in the field of physics, including the concept of electron cooling.

However, when electron cooling was first proposed by Budker, the general consensus from the scientific community was "brilliant idea, but unfortunately non-realistic." Luckily, Novosibirsk scientists did not listen to these predictions and successfully constructed the first e-cooler around 1974 and have built many more since then.

There is one more parallel between these two stories. Sir John Adams had a unique combination of scientific and engineering abilities, whereas Lev Landau once called Budker a "relativistic engineer." As it happens, the art of inventiveness (TRIZ) originated from engineering.

### 11.3.2 Higgs, superconductivity and TRIZ

In the first chapter, we discussed the connection between the TRIZ inventive principle of the nested doll (*matreshka*) and Bryusov's poem describing the world of an electron and particle detectors, which are arranged like a *matreshka* doll — a system within another system.

The particle detectors are the devices that (in tandem

with accelerators) helped to discover[2] the Higgs boson — the essential building block of the *Standard Model* of particle physics.

A particular connection we would like to highlight is the relationship between Higgs and superconductivity. In a recent article,[3] A. Pashkin and A. Leitenstorfer reminded us that "... the theoretical proposal of the Higgs mechanism was actually inspired by ideas from condensed matter physics ... In 1958, Anderson discussed the appearance of a coherent excited state in superconducting condensates with spontaneously broken symmetry... On page 1145 of this issue, Matsunaga et al. report direct observation of the Higgs mode in the conventional superconductor niobium nitride (NbN) excited by intense electric field transients."

The above passage shows us once again that that the general conclusion of TRIZ that "the same problems and solutions appear again and again but in different disciplines" is applicable to science, too.

### 11.3.3 Garin, *matreshka* and Nobel

One of the first inspirations and predictions of a device similar to today's laser may have appeared in a 1926 novel by Aleksey Tolstoy, *The Hyperboloid of Engineer Garin*. In that story, a device was described that was capable of producing a ray of light of immense power. It is fascinating to note that technical drawings were included in this novel — a significant attraction for curious readers (even despite the fact that the drawings referred to non-existing materials).

*The focusing mirror in Garin's hyperboloid was made from "shamonite" — an extremely durable material imagined by the writer. Fictitious and non-existing in early 20th century, such a material can perhaps be created in the 21st century thanks to the advent of new engineered materials.*

The adventures described in Tolstoy's novel were extraordinary. This fictional and powerful ray was responsible for many astounded remarks. One such was, "Can you imagine what opportunities are opening now? Nothing in nature can withstand the power of the beam of light — buildings, forts, battleships, airships, rocks, mountains, the earth's crust — everything could be penetrated, destroyed, cleaved with my beam." These are the words Tolstoy put into the mouth of the story's protagonist — the engineer Garin.

The lasers developed later in the century[4] did not adhere to the design in *The Hyperboloid of Engineer Garin*, and were luckily used for peaceful purposes.

In connection to lasers and TRIZ, we would like to mention here one particular recent invention, *stimulated emission depletion microscopy* (STED), which was developed by Stefan W. Hell and his colleagues.[5]

The stimulated emission depletion microscopy allows for

[2] Francois Englert and Peter W. Higgs, Nobel Prize in Physics, 2013.
[3] A. Pashkin and A. Leitenstorfer, Science 345, 1121 (2014).
[4] C. Townes, N. Basov and A. Prokhorov, Nobel Prize in Physics, 1964.
[5] E. Betzig, S. W. Hell and W. E. Moerner, Nobel Prize in Chemistry, 2014.

an increase of resolution in optical microscopes by a factor of around five. It does this by applying two pulses of laser light to an object under study — an excitation pulse and, shortly afterwards, a de-excitation pulse.

A fluorescent dye introduced into the object is first excited and then depleted with the second pulse of an appropriately different wavelength. The key feature of this technique is the usage of the depleting laser pulse with a special spatial profile, i.e., its minimum intensity is located in the center, as illustrated in Fig. 11.1.

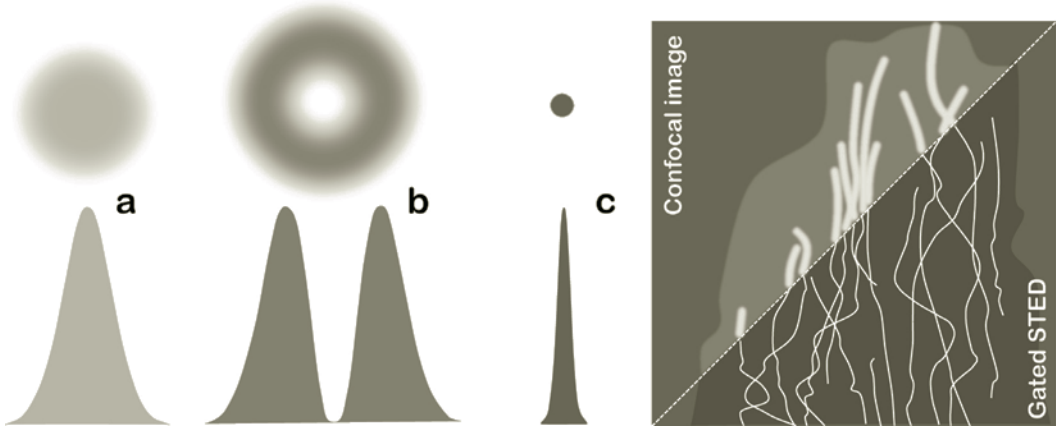



**FIGURE 11.1**
Stimulated emission depletion microscopy (STED) and TRIZ inventive principle of *matreshka* and system–antisystem. Excitation laser pulse (a), de-excitation pulse (b) and remaining fluorescence (c). Improvement of resolution of a protein imaging due to STED is shown qualitatively on the right.

The minimum intensity region of the second laser can be several times smaller than the wavelength. Therefore, detecting the remaining fluorescence spot will result in improving the imaging resolution to values several times lower than the laser wavelength.

Considering this invention within the framework we have discussed in this book, we can also note a remarkable connection to TRIZ. From the perspective of the theory of inventive problem solving, the STED method is an illustration of the use of the principle of system and antisystem (excitation and de-excitation laser pulse), perhaps combined with the inventive principle of the *matreshka* (one laser pulse is located geometrically inside of the other pulse).

We have already observed examples where system-antisystem and *matreshka* inventive principles combined to create novel systems (e.g., dual force neutral solenoids). This suggests that combinations of these or other inventive principles can be especially efficient in solving problems in various scientific fields.

## 11.4 Aiming for Pasteur quadrant

Science has always been a driver for the economy. This is a commonly accepted statement — however, the mechanisms of its impact are complicated. Still, their analysis is necessary, and not only from a philosophical point of view, but also in order to optimize the research priorities and define the strategies for technological innovation.

One of the attempts to analyze the model for research and technology transfer was done by Vannevar Bush, who, during World War II, was instrumental in reorganizing the research and science community according to the needs of that difficult time. Vannevar Bush's 1945 report[6] has since defined post-war scientific policy in the United States and in many other countries (supposedly, for decades to come).

In this report, Bush describes what would later be called a one-dimensional or linear model for research and technology transfer. Bush claims that research which is more basic is less applied and vice versa (illustrated in Fig. 11.2). According to Bush, applied research invariably drives out pure research if the two are mixed, and therefore, basic research must be completely isolated from *considerations of use*.


**FIGURE 11.2**
One-dimensional, linear model of research.

Correspondingly, the dynamic linear model of technology transfer resembles a pipeline wherein government funding stimulates basic research, which then in turn feeds applied research, and ultimately results in technology and product development, with society eventually benefiting from the process (see Fig. 11.3).

These views of the relationship between basic science and

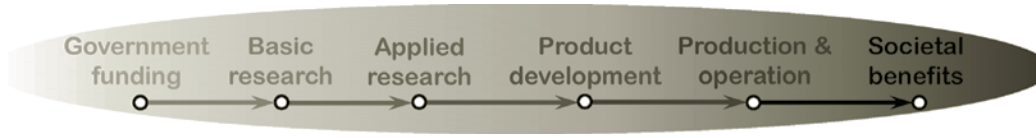



**FIGURE 11.3**
Dynamic linear model of technology transfer.

[6] *Science, the Endless Frontier: a Report to the President*, by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945.

technological innovation have since then been analyzed, criticized, and a new model has been developed.

The contradiction between these linear models and practice can be illustrated via the example of accelerator science and technology. The invention of "strong focusing" in the 1950s was a revolutionary change in accelerator technology. It enabled numerous applications. This invention may have come about as a result of a pure fundamental interest — however, it was developed as a result of the pursuit of a certain concrete goal, and was made possible due to certain technologies available at that time.

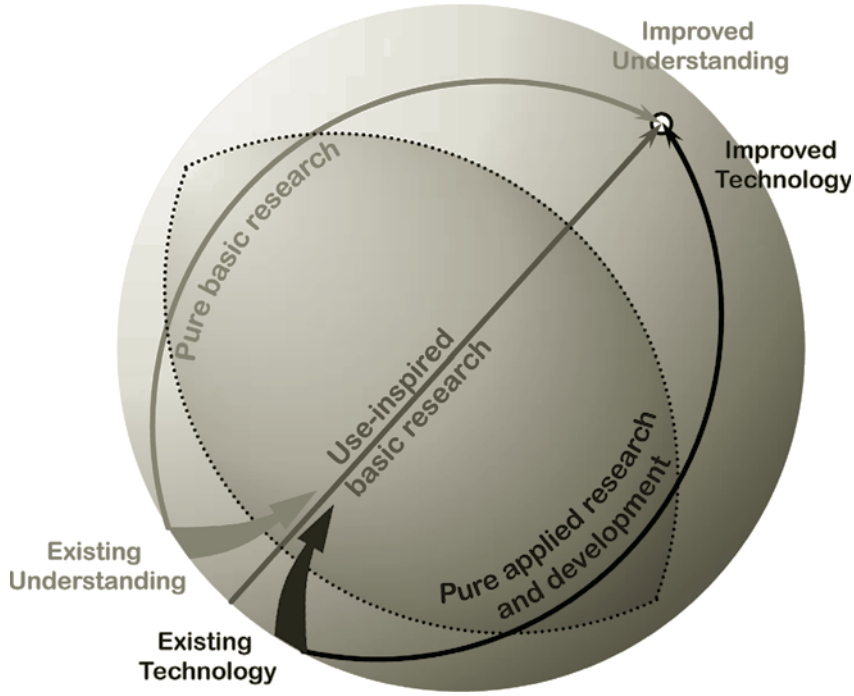



**FIGURE 11.4**
Revised dynamic model of research and technology transfer.

A new model of research and technology transfer was suggested by Donald Stokes, who worked on the Advisory Committee on Research for the National Science Foundation.

*"Research with consideration of use" is an effective way for science to provide an impact on our economy and society.*

In his report to the NSF, and in the book he subsequently published,[7] Donald Stokes argued against Bush's linear model and introduced the notion of use-inspired research — "research with consideration of use" — which redefined the paradigm of the relationship between basic science and technological innovation. A revised dynamic model suggested by Stokes is illustrated in Fig. 11.4.

[7] Donald E. Stokes, *Pasteur's Quadrant — Basic Science and Technological Innovation*, Brookings Institution Press 1997.

To portray his ideas, Donald Stokes suggested considering research on a two-dimensional plane, where the axes are fundamental knowledge impact and consideration of use.

Donald Stokes' plot is now called a *Pasteur quadrant* and it is illustrated in Fig. 11.5.

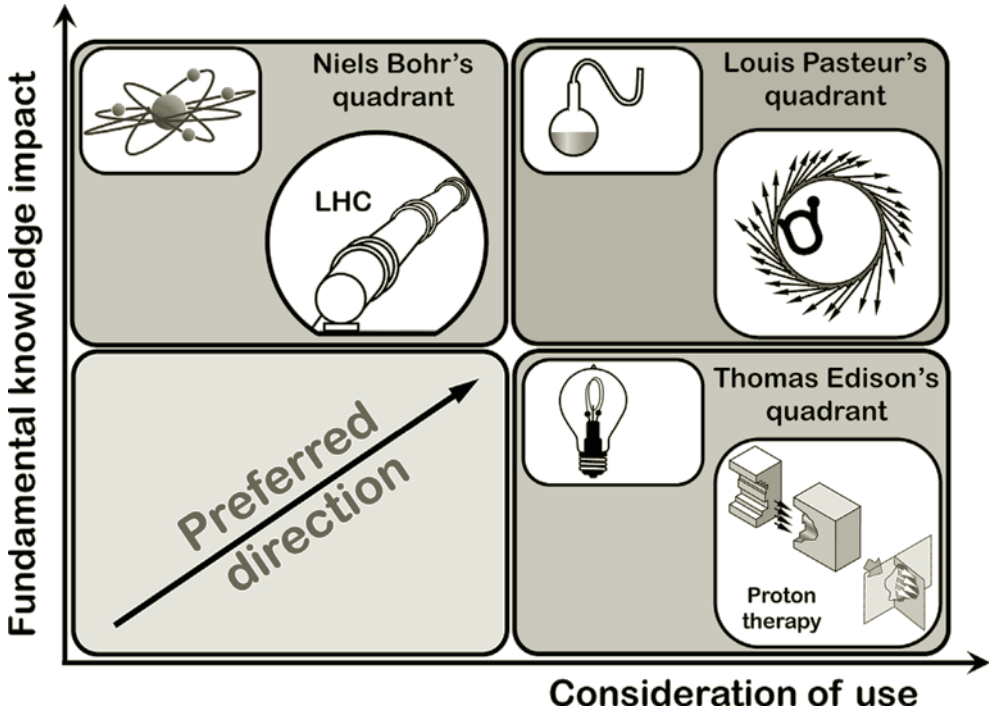



**FIGURE 11.5**
Pasteur quadrant and accelerator science.

A characteristic example of a purely fundamental scientific pursuit is the research works of Niels Bohr on the structure of nuclei, while another example of purely practical activity is Thomas Edison's developmental work on the filaments of light bulbs — as shown in Fig. 11.5.

The quantitative assessments of these examples — or of other research projects placed on this graph — can be done by evaluating the number of either the academic papers or patents, resulting from a particular research, as illustrated in Fig. 11.6.

Donald Stokes suggested, however, that an optimized approach should balance the fundamental pursuit of knowledge with consideration of use, which is illustrated by the works of Louis Pasteur — indicated in Fig. 11.5 as a preferred direction.

In the field of modern accelerator science and technology, characteristic examples for two sections of the quadrant are: colliders intended for the exploration of fundamental prop-

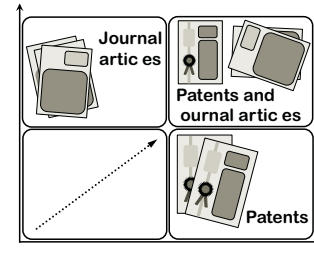



**FIGURE 11.6**
The units of quantitative assessments of research in the Pasteur quadrant.

erties of elementary particles (on one axis), and accelerator-based devices for medical applications (on the other axis).

The criteria suggested by Donald Stokes are universal and applicable to any scientific and technological area. By applying these criteria to accelerator science and technology, we can conclude that the preferred direction — which best balances the fundamental pursuit of knowledge with consideration of use — will be the direction to create novel light sources. Efforts made in this direction could potentially produce scientific instruments applicable to the investigation of protein structures or materials, which may also be almost directly applicable to the creation of new medicines or metals with controllable properties.

This analysis also shows that — as in many other disciplines — accelerator science and technology, in synergy with lasers and plasma physics, can truly span the entire range of directions — from pure fundamental science to pure applied development.

The research and technology innovation model is indeed not just linear, but at least two-dimensional, as Stokes outlined in his revised dynamic model. In the 21st century, the driving forces of technological innovation — as well as the global pact between science and society — are different than they were in the middle of the last century. The revised criteria — as illustrated in this section via the example of accelerator science — are universal, and can be applied to any discipline, which can help us to optimize the impact of our research investments on our economy and society.

## 11.5 How to cross the Valley of Death

Powerful beams of light, heat rays — now called lasers — are features of H.G. Wells' and Alexey Tolstoy's science fiction stories that have excited many generations of future inventors and scientists. These stories thrilled crowds of kids (and adults) as they rambled along back streets dreaming of having lasers in their pockets.

The curiosity and imagination of the younger generation is the fuel that enables the development of our civilization. The challenge for governments, educational institutions and societies is to understand how to nurture and later harness these attributes.

The first visible light lasers (with a wavelength of light of about half a micrometer) were typically big when they were created half a century ago, even huge, and certainly not pocket size. Now, miniature lasers are in CD players, bar-scanners in shops, laser-pointers — practically everywhere. However, lasers with a much shorter wavelength, in the Angstrom range (light in the X-ray spectrum), have only just become available. Called free electron lasers, they are a

kilometer long. Due to their short wavelength, X-rays are already indispensable for the analysis of protein structure, synthetic molecules, new materials and many other objects. The size and affordability of such X-ray lasers are major obstacles that hinder the widespread use of compact X-ray lasers. Creating compact X-ray lasers is the challenge that accelerator science now needs to confront.

Particle accelerators have already impacted many areas of our lives with their medical and industrial uses, as well as with their help in creating research instruments. Tens of millions of patients receive accelerator-based diagnoses and treatments each year, worldwide. The total annual market value for all products that are treated or inspected by accelerators is more than $500B.[8] Approximately 30% of the Nobel prizes in physics, as well as many in other areas, are directly connected to the use of accelerators.[9]

The ideas that enabled the use of accelerators in everyday life and industry were developed decades ago. New ideas will be essential for ensuring the future impact of this field.

Conventional accelerators, no matter how advanced they may be, are primarily based on the acceleration of particles in cavities — metal vessels shaped to resonate and create accelerating fields. The ability of metals to tolerate high electromagnetic fields is intrinsically limited. However, an accelerating wave can be created when gas is ionized and excited by an intense beam of particles or by a laser pulse, becoming plasma. Plasma is an indestructible medium and is able to withstand a thousand-times-higher accelerating gradient.

Accelerator science and technology is on the edge of a breakthrough brought on by synergy with laser and plasma physics. The most immediate outcome that this synergy will enable is the creation of novel, compact X-ray lasers and light sources. The direct collision of beam and laser light also opens up another opportunity for the creation of X-ray sources via the use of the Compton effect (when visible light photons are reflected from a relativistic electron beam and thereby decrease their wavelength down to Angstrom levels).

Science is indeed the driver of our civilization's progress. However, the journey from initial ideas and experimental demonstrations to widespread commercial applications is long and difficult. Various studies performed in different countries have all found a gap, a so-called "Valley of Death" in technology transfer. It is difficult to bridge the middle range of the technological readiness of ideas. On one end, the research institutions are usually not positioned to develop ideas into commercial applications, while on the other end, the risk is often too high for industry to pick up ideas that are too fresh and undeveloped.

[8] Accelerators for America's Future, Department of Energy, 2009.
[9] E. Haussecker and A. Chao, Physics in Perspective, 2011.

The challenge originates from the different motivations, methods and timescales of three key players: academic institutions, industry and investors.[10] Their corresponding aims and motivations — the front-end fundamental scientific results, the development of commercial devices in the foreseeable future, and optimization of investments versus risk/return factors — are often incompatible (Fig. 11.7).

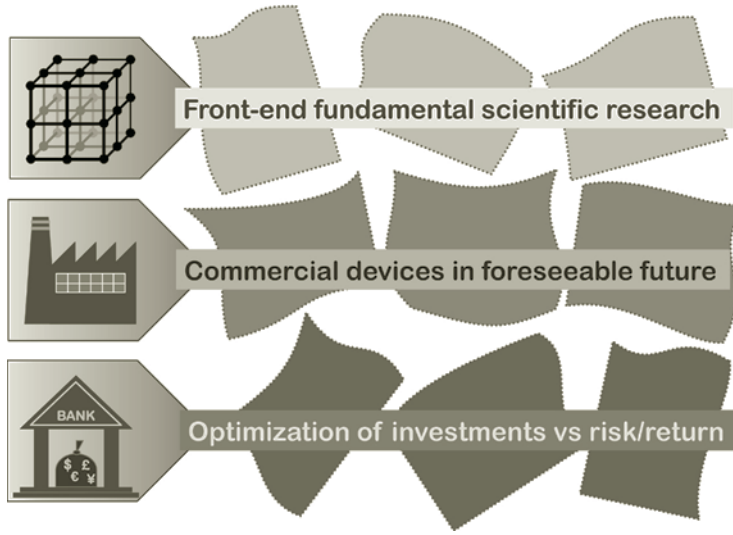



**FIGURE 11.7**
Academia–industry–investor puzzle caused by different motivations of the three participating groups.

Accelerators in synergy with lasers and plasma may, in fact, offer a solution for the academia–industry–investor puzzle through simultaneous, parallel work on a portfolio of three different types of compact X-ray light sources (Fig. 11.8).

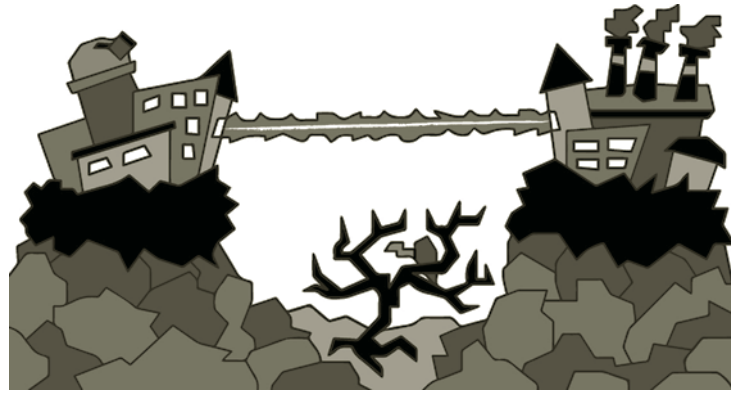


**FIGURE 11.8**
Working on a portfolio of compact X-ray light sources can help in crossing the "Valley of Death" between accelerator science and technological innovation.

[10] A government can be an investor.

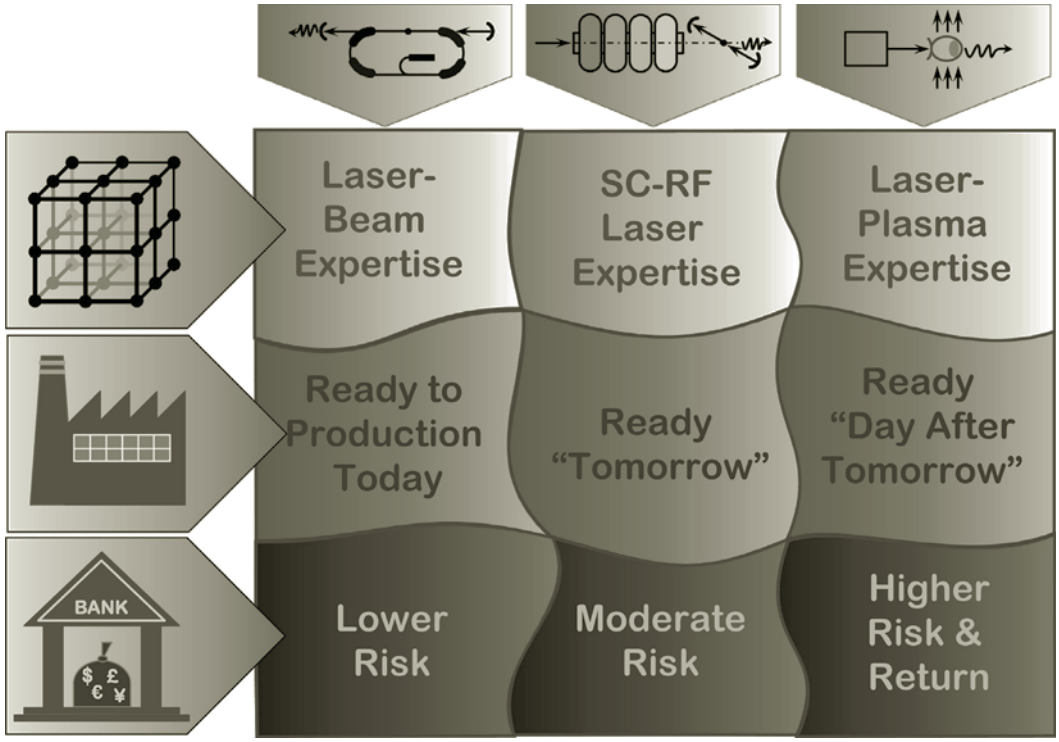



**FIGURE 11.9**
A solution to the academia–industry–investor puzzle — work on three designs of compact X-ray sources.

A solution to the academia–industry–investor puzzle involving work on three designs of compact X-ray sources is illustrated in Fig. 11.9.

Compton X-ray sources — the first type of source in the solved puzzle — are now actively being developed and are a lower-risk investment for industrial use. Yet a more challenging, but promising, Compton source requires superconducting acceleration to allow for a much higher electron beam current and X-ray brightness. This second option resides in the middle of the range for both the projected availability and risk/return.

*How to solve the academia–industry–investor puzzle? Working on a portfolio of compact light sources may give an answer.*

The most challenging, but also the most promising source is an X-ray source based on laser plasma acceleration — initially a betatron source, and ultimately, perhaps in less than a decade, a free electron laser. If we properly schedule the relative progression of the different stages of research and development for these three types of X-ray sources, we could balance the typical risks associated with the development of innovative products, and the opportunities they offer.

Such work on compact X-ray sources fits into to the preferred direction of the Pasteur quadrant. Moreover, research and development in this direction may result in the creation

of instruments that every university will aspire to have. The ability to obtain a compact X-ray laser would revolutionize contemporary science and technology yet again — not unlike how the spread of near-visible light lasers impacted science and industry during the 20th century.

## 11.6 How to learn TRIZ in science

In this second-to-last section, I would like to convey to the reader my personal impressions of the best ways to learn (and also teach) the methodologies of inventiveness in the science departments of universities. These impressions are based on the development of training programs involving TRIZ at the John Adams Institute for Accelerator Science — JAI — at Oxford University, Imperial College and Royal Holloway University of London.

The JAI is a center of excellence in the UK for advanced and novel accelerator technology. We focus on researching and developing accelerator science and techniques, as well as graduate training. Our cohort of PhD students numbers around three dozen, and on average we educate about six PhD experts per year. Like any other research organization or graduate school, we are motivated to turn the results of our research into having a positive societal impact, as well as endeavoring to find ways to be more inventive and innovative. For the latter, we recently (several years ago) started to look at TRIZ and developed a specific, scientifically oriented way to introduce this method to our PhD students.

Let us briefly review TRIZ and its connection to science graduate schools. TRIZ is the methodology of inventiveness that was specifically developed for engineering fields. It was created during the second half of the 20th century and gradually became one of the most powerful tools in the industrial world. TRIZ has become, according to Forbes, the bedrock of innovations of such companies as Samsung, as well as many others.

While the TRIZ method of industrial inventiveness was originally created for engineering, this methodology is universal and can also be applied to science. However, experience shows that knowledge of TRIZ is nearly nonexistent in the scientific departments of western universities.

Moreover, it is not rare to hear about unsuccessful attempts to introduce TRIZ into graduate courses of scientific departments of universities. Indeed, in many or most of these cases, the apparent reason for the failure was that the canonical version of TRIZ was introduced to science PhD students in the same way that TRIZ is taught to engineers in industrial companies. This appears to be a mistake, seeing as science students are rightfully more critically minded and justifiably sceptical about overly prescriptive step-by-step meth-

ods. (As we have seen above, TRIZ involves finding a pair of contradicting parameters in a problem, which then — using the TRIZ inventive tables created by TRIZ teams, based on the analysis of hundreds of thousands of past inventions — immediately leads to selecting just a few suitable inventive principles, narrowing down the choice and resulting in a much faster solution to a problem.)

The approach of introducing TRIZ to JAI graduate students is different, and takes into account lessons learned by its predecessors. Instead of teaching our graduate students the ready-to-use methodology, we are effectively taking them through the process of recreating parts of the TRIZ methodology by analyzing various inventions and discoveries from scientific disciplines, showing that these inventive principles can be efficiently applied to science. Moreover, in the process of this development, we often found that additional inventive principles, more suitable for scientific disciplines, can be introduced and added to standard TRIZ — we call this extension Accelerating Science TRIZ (the play on words is now apparent — the word "accelerating" is not referring to accelerators any more, but highlights that TRIZ can help to boost science).

The approach to teaching TRIZ described above has now been successfully introduced to JAI graduate students and was also successfully implemented at a course in the US-PAS (US Particle Accelerator School), and has also been introduced in the JUAS (European Joint Universities Accelerator School) and APPEAL school (a one-day to one-week course for high-school teachers that we conduct every year as part of our outreach activity).

TRIZ methodology is another way to look at the world. Combined with science it creates a powerful and eye-opening amalgam of science and inventiveness. This methodology is particularly helpful for building bridges of understanding between completely different scientific disciplines, and so is also naturally useful to educational and research organizations that endeavor to break barriers between disciplines.

Ultimately, my recommendations on learning and teaching TRIZ in universities can be summarized as follows. Introducing TRIZ courses/lessons to university students is only the first step. However, while teaching these TRIZ courses/lessons, avoid the canonical, ready-to-use version of TRIZ. Instead, take the students through the process of proactively adapting TRIZ for science. Arrange a sequence of relatively short TRIZ lessons throughout the entire duration of education in the university, so that students can gradually build up their ability to use TRIZ methods in order to build bridges of understanding between different disciplines that they will interact with during their education and research.

## 11.7 Let us be challenged

Many of the examples of TRIZ-like inventions in science that we considered in this book have already been made, and we only analyzed the applicability of TRIZ post-factum.

It is natural to wonder whether TRIZ and AS-TRIZ can actually help to inspire and create new scientific inventions and innovations, especially in regards to projects that continue to manifest many unsolved obstacles.

One example of such a project is a circular collider currently being considered as a successor to the Large Hadron Collider at CERN — *the Future Circular Collider* (FCC).[11] The FCC project is looking to create a 100 km tunnel infrastructure in order to house a 100 TeV CM proton–proton collider (the $e^+e^-$ and proton-e options are also being considered).

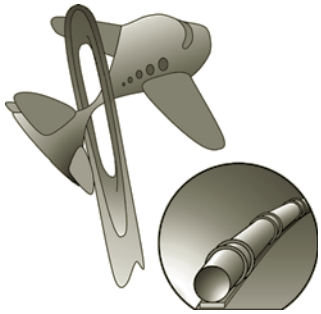


**FIGURE 11.10**
The FCC beam energy will be comparable to that of an airplane, while the beam will need to be focused at the interaction point to micron-scale size — an analogy with a plane passing through the needle's eye.

This project has many scientific and technical tasks and challenges that need to be solved. Notably, one issue is that the synchrotron radiation at these high energies starts to influence the proton beams. The total energy in each circulating proton beam is expected to exceed 8 GJ, which is equivalent to a kinetic energy of an Airbus-380 flying at 720 km/h. Not only does such a beam need to be handled safely in the bending magnets, but it also needs to be focused in the interaction region to a micron spot size — which is practically the equivalent of literally having to pass through a needle's eye

It remains to be seen if TRIZ and AS-TRIZ methodology can be applied to such a large-scale project as the FCC, as it brings a whole array of new, difficult and exciting challenges to the table. Nonetheless, it is certainly a project that can only flourish with the application of our knowledge and inventiveness.

[11] Future Circular Collider — http://cern.ch/fcc

**EXERCISES**

11.1 *Mini-project*.
A proposal for a Higgs and top factory is based on the design of an electron-positron collider (175 GeV for each beam) in a tunnel with a 50 km circumference. Assuming that the maximum power delivered by the RF system to each beam is limited to 50 MW, estimate the maximum current that can be stored in such a collider. Make an assumption about the focusing system at the interaction point, and evaluate beam emittances and the collider's luminosity.

11.2 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
A liquid anode X-ray tube is a contemporary technology that increases the photon flux. Analyze this technology in terms of the TRIZ and AS-TRIZ approach, identifying a contradiction and a general inventive principle that was used (could have been used) in this invention.

11.3 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
A liquid jet or liquid target is often used in high energy physics experiments, particularly for the production of antiprotons. A liquid jet allows withstanding the power of the incident proton beam. Analyze this technology in terms of the TRIZ and AS-TRIZ approach, identifying a contradiction and a general inventive principle that were used (could have been used) in this invention.

11.4 *Analyze inventions or discoveries using TRIZ and AS-TRIZ.*
Analyze and describe scientific or technical inventions described in this book in terms of the TRIZ and AS-TRIZ approaches, identifying a contradiction and an inventive principle that were used (could have been used) for these inventions.

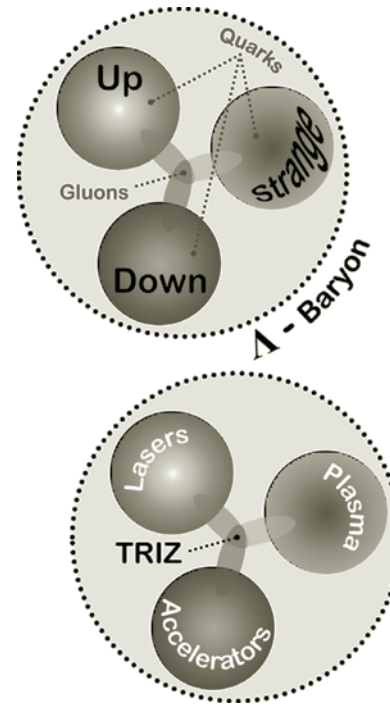
11.5 *Developing AS-TRIZ parameters and inventive principles*.
Based on what you have learned in this book, discuss and suggest the possible additional parameters for the AS-TRIZ contradiction matrix, as well as the possible additional AS-TRIZ inventive principles.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Final Words

Writing this book was an absolutely wonderful and enjoyable experience, worth every hour of long nights and weekends spent working on it. It has been about two years since I first had the inkling to write it, although the actual writing, from the first word of the first chapter to the last sentence, has taken about nine months.

The book is different from many (if not all) textbooks on accelerators, lasers and plasma because firstly, it looks at these areas together, and secondly, it mixes them with the methodology of inventiveness — the glue that pulls these areas close to each other just as gluons connect the quarks in a Λ-Barion, as illustrated below.




There are still many scientific topics I would have wanted to discuss and include in this book, but I had to resist this temptation, primarily based on the expectation that feedback from the readers will be invaluable for the work on possible future editions.

Andrei Seryi — Oxford, March 5, 2015

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Bibliography


Altshuller, G. (1999). *Innovation Algorithm: TRIZ, systematic innovation and technical creativity* (first ed.). Worcester, MA: Technical Innovation Center, Inc.

Amaldi, U. et al. (2013). *Elementary Particles*, Volume **21**: Subvolume C: Accelerators and Colliders , Group I: Elementary Particles, Nuclei and Atoms. Berlin, Heidelberg, New York: Springer-Verlag.

Badelek, B. et al. (2004). *The photon collider at TESLA*. International Journal of Modern Physics A **19**(30), pp.5097–5186. New Jersey, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai: World Scientific Publishing Co. Pte. Ltd.

Baklakov, B. and others (1991). *Investigation of seismic vibrations and relative displacements of linear collider VLEPP elements*. In Proc. of 1991 IEEE PAC. p.3273. San-Francisco, CA: Proc. of 1991 IEEE PAC.

Balakin, V. (1991). *Travelling Focus Regime for Linear Collider VLEPP*. In Proc. 7th ICFA Workshop on Beam Dynamics. Los Angeles, CA: Proc. 7th ICFA Workshop on Beam Dynamics.

Balakin, V., Nokhatsky, A. and Smirnov, V. (1983). *VLEPP: Transverse Beam Dynamics*. In Proceedings of the 12th Int. Conf. on HEACC. Batavia, IL: FERMILAB Print.

Bellan, P. (2006). *Fundamentals of Plasma Physics*. New York: Cambridge University Press Inc.

Blumenfeld, I. et al. (2007). *Energy doubling of 42 GeV electrons in a metre-scale plasma wakefield accelerator*. Nature **445**, pp.741–744. London: Macmillan Publishers Limited.

Borghesi, M. et al. (2006). *Fast Ion Generation by High-Intensity Laser Irradiation of Solid Targets and Applications*. Science and Technology **49**(3), pp.412–439. La Grange Park, IL: American Nuclear Society.

Borghesi, M. et al. (2008). *Laser-driven proton acceleration: source optimization and radiographic applications*. Plasma Phys. Control. Fusion (**50**, 124040). Bristol, Philadelphia, Beijing, Tokyo: IOP Publishing Ltd.

Bosco, A. et al. (2008). *A Two-Dimensional Laser-Wire Scanner for Electron Accelerators*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **592**(3), pp.162–170. Elsevier.

Bourgeois, N., Cowley, J. and Hooker, S. M. (2013). *Two-Pulse Ionization Injection into Quasilinear Laser Wakefields*. Phys. Rev. Lett. **111**(15, 155004). Ridge, NY: American Physical Society.

Boyd, T. and Sanderson, J. (2003). *The Physics of Plasmas*. New York: Cambridge University Press Inc.

Brinkmann, R. et al. (2001). *A low emittance, flat-beam electron source for linear colliders*. Phys. Rev. ST Accel. Beams **4**(5, 053501). Ridge, NY: American Physical Society.

Budker, G.I. et al. (1976). *Experimental studies of electron cooling*. Particle Accelerators **7**, pp.197–211. United kingdom: Gordon and Breach, Science Publishers Ltd.

Burrows, M. and Sutton, G. (2013). *Interacting Gears Synchronize Propulsive Leg Movements in a Jumping*. Science **341**(6151), pp.1254–1256. Washington, DC: American Association for the Advancement of Science.

Bush, V. (1945). *Science, the Endless Frontier: a Report to the President*. Science Education **29**(4), p.218. Washington, D. C.: U.S. Government Printing Office.

Chao, A. et al. (2013). *Handbook of Accelerator Physics and Engineering* (second ed.). New Jersey, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai: World Scientific Publishing Co. Pte. Ltd.

Chen, F. (1974). *Introduction to Plasma Physics*. New York and London: Plenum Press.

Clark, E. et al. (2000). *Measurements of Energetic Proton Transport through Magnetized Plasma from Intense Laser Interactions with Solids*. Phys. Rev. Lett. **84**(18), p.670. Ridge, NY: American Physical Society.

Daido, H. et al. (2012). *Review of laser-driven ion sources and their applications*. Rep. Prog. Phys. (**75**, 056401). Bristol, Philadelphia, Beijing, Tokyo: IOP Publishing Ltd.

Danilov, V. and Nagaitsev, S. (2010). *Nonlinear accelerator lattices with one and two analytic invariants*. Phys. Rev. ST Accel. Beams **13**(084002). Ridge, NY: American Physical Society.

Davidson, R. (2001). *Physics of Nonneutral Plasmas*. New Jersey, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai: World Scientific Publishing Co. Pte. Ltd.

Delayen, J. R. (2012). *Superconducting spoke cavities for electrons and high-velocity proton linacs*. In Proceedings of LINAC2012, 03 Technology, 3A Superconducting RF, TH1A03, pp.758–762. Tel Aviv, Israel: Proceedings of LINAC2012.

Dendy, R. (1990). *Plasma Dynamics*. New York: Oxford University Press Inc.

Derbenev, Y. (1998). *Adapting Optics for High Energy Electron Cooling*. In University of Michigan Report No. UM HE 98-04. Ann Arbor, MI: University of Michigan Report.

Derbenev, Y. et al. (1995). *Microbunch Radiative Head-Tail Interaction*. In Tesla-FEL Report. Hamburg, Germany: DESY Print.

Diver, D. (2013). *Plasma Formulary for Physics, Astronomy, and Technology* (second ed.). Weinheim, Germany: WILEY-VCH Verlag.

Edwards, D. et al. (2004). *An Introduction to the Physics of High Energy Accelerators* (second ed.). Weinheim, Germany: WILEY-VCH Verlag.

Eliezer, S. (2002). *The Interaction of High-Power Lasers with Plasma*. Bristol, Philadelphia, Beijing, Tokyo: IOP Publishing Ltd.

Esirkepov, T. et al. (2004). *Highly Efficient Relativistic-Ion Generation in the Laser-Piston Regime*. Phys. Rev. Lett. **92**(7, 175003). Ridge, NY: American Physical Society.

Gabor, D. (1947). *A Space-Charge Lens for the Focusing of Ion Beams*. Nature **160**, pp.89–90. London: Macmillan Publishers Limited.

Gadd, K. (2011). *TRIZ For Ingeneers: Enabling Inventive Problem Solving*. Chichester, West Sussex: John Wiley & Sons, Ltd., Oxford Creativity.

Gibbon, P. (2005). *Short pulse laser interactions with matter. An introduction*. London: Imperial College Press.

Haberberger, D. et al. (2012). *Collisionless shocks in laser-produced plasma generate monoenergetic*. Nature Physics **8**(1), pp.95–99. London: Macmillan Publishers Limited.

Haseroth, H. et al. (1996). *Laser Ion Source Development for Heavy Ions*. In Proceedings of the XVIII International linear Accelerator Conference Linac96, v.2, WE202. pp.570–574. Geneva, Switzerland: CERN, European Organization for Nuclear Research.

Haussecker, E. and Chao, A.W. (2011). *The Influence of Accelerator Science on Physics Research*. Physics in Perspective (PIP). **13**(2), pp.146–160. Bristol, Philadelphia, Beijing, Tokyo: IOP Publishing Ltd.

Herr, W. (2013). *New Tools for Non Linear Dynamics I, II*. In The CERN Accelerator School 2013, Trondheim, Norway. Geneva, Switzerland: CERN, European Organization for Nuclear Research.

Hidding, B.B. et al. (2012). *Beyond injection : Trojan horse underdense photocathode plasma wakefield acceleration*. In Advanced Accelerator Concepts. AIP Conference Proceedings., pp.570–575. Melville, NY: AIP.

Hooker, S. et al. (2014). *Multi-Pulse Laser Wakefield Acceleration: A New Route to Efficient, High-Repetition-Rate Plasma Accelerators and High Flux Radiation Sources*. J. Phys. B **47**, (234003). Bristol, Philadelphia, Beijing, Tokyo: IOP Publishing Ltd.

Hooker, S. and Webb, C. (2010). *Lasers physics (Oxford Master Series in Physics)*. New York: Oxford University Press Inc.

Huang, Zh. and Ruth, R.D. (1998). *Laser-electron storage ring*. Phys. Rev. Lett. **80**, pp.976–979. Ridge, NY: American Physical Society.

Jackson, J. (1999). *Classical Electrodynamics* (third ed.). Hoboken, NJ: John Wiley & Sons, Ltd.

Jaroszynski, D. et al. (2009). *Laser-Plasma Interactions*. Boca Raton, FL: CRC Press, Taylor Francis Group.

Kneip, S. et al. (2011). *X-ray phase contrast imaging of biological specimens with femtosecond*. Applied Physics Letters **99**. Melville, NY: AIP Publishing LLC.

Landau, L. (1946). *On the vibration of the electronic plasma*. JETP (**16**), p.574. Russia: International Academic Publishing Company (IAPC) "Nauka/Interperiodica".

Lee, S. (2012). *Accelerator Physics* (third ed.). New Jersey, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai: World Scientific Publishing Co. Pte. Ltd.

Leemans, W. et al. (2006). *GeV electron beams from a centimetre-scale accelerator*. Nature Physics **2**, pp.696–699. London: Macmillan Publishers Limited.

Leemans, W. et al. (2014). *Multi-GeV Electron Beams from Capillary-Discharge-Guided Subpetawatt Laser Pulses in the Self-Trapping Regime*. Phys. Rev. Lett. **113**, (245002). Ridge, NY: American Physical Society.

Macchi, A. et al. (2013). *Ion acceleration by superintense laser-plasma interaction*. Rev. Mod. Phys. **85**, pp.751–793. Ridge, NY: American Physical Society.

Maksimchuk, A. et al. (2000). *Forward Ion Acceleration in Thin Films Driven by a High-Intensity Laser*. Phys. Rev. Lett. **84**(18), pp.4108–4111. Ridge, NY: American Physical Society.

Mangles, S. et al. (2004). *Monoenergetic beams of relativistic electrons from intense laser-plasma interactions*. Nature **431**, pp.535–538. London: Macmillan Publishers Limited.

Mendonca, J. (2001). *Theory of Photon Acceleration*. Bristol, Philadelphia, Beijing, Tokyo: IOP Publishing Ltd.

Mourou, G. et al. (2013). *The future is fibre accelerators*. Nature Photonics (**7**), pp.258–261. London: Macmillan Publishers Limited.

Najmudin, Z. (2014). *Laser Driven Systems*. In The CERN Accelerator School 2014, Geneva, Switzerland. Geneva, Switzerland: CERN, European Organization for Nuclear Research.

Palmer, C.A. et al. (2011). *Monoenergetic Proton Beams Accelerated by a Radiation Pressure Driven* . Phys. Rev. Lett. **106**(1, 014801). Ridge, NY: American Physical Society.

Palmer, R.B. (1988). *Energy Scaling, Crab Crossing and the Pair Problem*. SLAC-PUB-4707. Menlo Park, CA: SLAC–PUB.

Parker, B. et al. (2007). *The Superconducting Magnets of the ILC Beam Delivery System*. SLAC-PUB-12832. Menlo Park, CA: SLAC–PUB.

Pashkin, A. and Leitenstorfer, A. (2014). *Particle physics in a superconductor*. Science **345**(6201), pp.1121–1122. Washington, DC: American Association for the Advancement of Science.

Pozimski, J. and Aslaninejad, M. (2013). *Gabor lenses for capture and energy selection of laser driven ion beams in cancer treatment*. Laser and Particle Beams **31**, pp.723–733. New York: Cambridge University Press Inc.

Qiao, B. et al. (2009). *Stable GeV Ion-Beam Acceleration from Thin Foils by Circularly Polarized Laser Pulses*. Phys. Rev. Lett. **102**(14, 145002). Ridge, NY: American Physical Society.

Qiao, B. et al. (2010). *Radiation-Pressure Acceleration of Ion Beams from Nanofoil Targets: The Leaky Light-Sail Regime*. Phys. Rev. Lett. **105**(15, 155002). Ridge, NY: American Physical Society.

Radovinsky, A. et al. (2013). *Variable Energy Acceleration in a Single Iron-Free Synchrocyclotron*. In Variable Energy Acceleration in a Single Iron-Free Synchrocyclotron. Cambridge, MA: Plasma Science and Fusion Center Massachusetts Institute of Technology PSFC/RR-13-9.

Raimondi, P. et al. (2008). *Suppression of beam-beam resonances in crab waist collisions*. In Proceedings of EPAC08, Genoa, Italy, pp.2620–2622. Proceedings of EPAC08.

Raimondi, P. and Seryi, A. (2001). *Novel Final Focus Design for Future Linear Colliders*. Phys. Rev. Lett. **86**(17, 3779). Ridge, NY: American Physical Society.

Saldin, E. et al. (2010). *The Physics of Free Electron Lasers (Advanced Texts in Physics)*. Berlin and Heidelberg, Germany: Springer-Verlag.

Schlenvoigt, H.-P. et al. (2008). *A compact synchrotron radiation source driven by a laser-plasma wakefield accelerator*. Nature Physics **4**, pp.130–133. London: Macmillan Publishers Limited.

Seryi, A. and Napoly, O. (1996). *Influence of ground motion on the time evolution of beams in linear colliders*. Phys. Rev. E **53**(5, 5323). Ridge, NY: American Physical Society.

Sessler, A. and E. Wilson (2014). *Engines of Discovery. A Century of Particle Accelerators* (Revised and Expanded Edition ed.). New Jersey, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai: World Scientific Publishing Co. Pte. Ltd.

Shiltsev, V. et al. (1999). *Considerations on compensation of beam-beam effects in the Tevatron*. Phys. Rev. ST Accel. Beams **2**(071001). Ridge, NY: American Physical Society.

Shiltsev, V. (2010). *Observations of Random Walk of the Ground in Space and Time*. Phys. Rev. Lett. **104**(23, 238501). Ridge, NY: American Physical Society.

Silva, L. et al. (2004). *Proton Shock Acceleration in Laser-Plasma Interactions*. Phys. Rev. Lett. **92**(1, 015002). Ridge, NY: American Physical Society.

Snavely, R. et al. (2000). *Intense High-Energy Proton Beams from Petawatt-Laser Irradiation of Solids*. Phys. Rev. Lett. **85**(14), p.2945. Ridge, NY: American Physical Society.

Spence, D.J. and Hooker, S.M. (2001). *Investigation of a hydrogen plasma waveguide*. Phys. Rev. E **63**(1,015401(R)). Ridge, NY: American Physical Society.

Strickland, D. and Mourou, G. (1985). *Compression of amplified chirped optical pulses*. Optics Communications **56**(3), pp.219–221. Elsevier.

Stupakov, G. (2009). *Using the Beam-Echo Effect for Generation of Short-Wavelength Radiation*. Phys. Rev. Lett. **102**(7, 074801). Ridge, NY: American Physical Society.

Variola, A. et al. (2010). *ThomX - Conceptual Design Report*. In ThomX - Conceptual Design Report, pp.1–136. LAL RT 09/28, SOLEIL/SOU-RA-2678.

Wang, J.W. and Loew, G.A. (1997). *Field Emission and RF Breakdown in High-Gradient Room-Temperature Linac Structures*. SLAC-PUB-7684. Menlo Park, CA: SLAC–PUB.

Wiedemann, H. (2007). *Particle Accelerators Physics* (third ed.). Berlin and Heidelberg, Germany: Springer-Verlag.

Wilson, E.J.N. (2001). *An Introduction to Particle Accelerators*. New York: Oxford University Press Inc.

Wilson, E.J.N. (2011). *Sir John Adams: his legacy to the world of particle accelerators. John Adams Memorial Lecture, 2009*. Geneva, Switzerland: CERN, European Organization for Nuclear Research.

Yin, L. et al. (2011). *Three-Dimensional Dynamics of Breakout Afterburner Ion Acceleration Using High-Contrast Short-Pulse Laser and Nanoscale Targets*. Phys. Rev. Lett. **107**(4, 045003). Ridge, NY: American Physical Society.

Yu, T.-H. et al. (2010). *Stable Laser-Driven Proton Beam Acceleration from a Two-Ion-Species*. Phys. Rev. Lett. **105**(6, 065002). Ridge, NY: American Physical Society.

Zholents, A. and Zolotorev, M. (1996). *Femtosecond X-Ray Pulses of Synchrotron Radiation*. Phys. Rev. Lett. **76**(6), p.912. Ridge, NY: American Physical Society.

# Index