SLAC-R-1082

HIGH BRIGHTNESS ELECTRON BEAMS FOR FOURTH GENERATION LIGHT SOURCES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Bryant William Garcia
November 2017

This material is based upon work supported by the U.S. Department of Energy,
Office of Science, under Contract No. DE-AC02-76SF00515.

Abstract

In this dissertation, we examine the production and preservation of high brightness electron beams for fourth generation light sources. The relentless push toward brighter photon pulses from Free Electron Lasers (FELs) has been facilitated by an increase in the brightness of the driving electron beam. One method of increasing this brightness is to provide an electron beam which is pre-bunched at the FEL wavelength, thereby providing a fully coherent seed for the lasing process. We explore the technique of Echo-Enabled Harmonic Generation (EEHG) to seed the electron beam at a high harmonic wavelength of a conventional, fully coherent laser pulse. We build on the previous work at the Next Linear Collider Test Accelerator (NLCTA) to extend the harmonic up-conversion via EEHG to the 75th harmonic. This level of up-conversion provides a proof of principle for future FEL facilities to achieve in a single seeding stage fully coherent radiation in the soft X-rays. We additionally demonstrate the ability to create multi-color, tunable bunching spectra using a chirped electron beam and the EEHG technique.

As the electron beam brightness is increased to match the demands of the FEL facilities, however, collective effects work to destroy the meticulously prepared beam. We examine the interplay between the microbunching instability and EEHG, and show that the tunable laser and chicane parameters of EEHG offer the possibility of selectively exciting or suppressing the effects of the instability.

Finally, we develop a theoretical model for a novel source of emittance degradation due to a stochastic addition to the standard Coherent Synchrotron Radiation (CSR) field. This addition is found to grow with both electron beam energy and density, potentially limiting the ultimate brightness of some electron beams. Physically, the effect is found to be due to the stochastic scattering of electrons off the narrow-angle synchrotron radiation cones of other electrons as they traverse a bend magnet. The effect is therefore similar to intra-beam scattering, which involves the scattering off the Coulomb field rather than radiation field, and appropriate comparisons between these effects are performed.

Acknowledgements

I would like to acknowledge, first and foremost, the encouragement and support my advisor Tor Raubenheimer has given me over the years. Since the day I bumbled into his office in September of 2014, he has been a source of inspiration, ideas, climbing route suggestions, and advice as I have attempted to navigate the waters of Graduate School.

I would also like to thank everyone on the NLCTA team for showing me the ropes and assisting me whenever possible. In particular, I thank Carl Hudspeth, Doug McCormick, Keith Jobe, Juan Cruz, Igor Makasyuk, Joel England, Kent Wootton, and Carsten Hast. I am also indebted to my partners in crime from the Echo program: Erik Hemsing and Mike Dunning. To Erik, for mentoring me in all things physics related, and for the consistent reminders to take things one step at a time. To Mike, for providing comic relief and mentoring me in design, as well as for encouraging the true spirit of a graduate student.

Of course, the years of graduate school would not have been bearable without the friends I have made over the years. In particular, I am thankful for Andrei Kamalov's continuous indulgence of my own ludicrous behavior, his encouragement to 'get up there', and his companionship on many an expedition. Although there are too many to name here, I feel compelled to single out some of the other friends who have been especially transformative to me during this time: Greg McCracken, Leighton Watson, Andrew McLeod, Harry Desmond, and Simon Foreman. To all these friends: It has been an absolute pleasure to know and work with you.

I am grateful to Hannah Siemann, whose fellowship and support have helped to push me across the finish line.

I cannot overstate the importance of the support of my parents Judy and Luis through these twenty-some odd years of continuous schooling. Without their unwavering support, I doubt I could have found the courage to accomplish any of this.

And last but certainly not least, to my partner Lea Hirsch. You have been a consistent source of companionship, humor, support, and assistance with crossword puzzles throughout these years. Thank you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> Do not wait to strike till the iron is hot, but make it hot by striking.
>
> — William Butler Yeats

Particle accelerators, originally conceived to probe the fundamental constituents of matter, have become increasingly useful as light sources in their own right. This began with the discovery of synchrotron radiation [1], which at the time was considered more a parasitic energy loss mechanism than a useful tool. The potential of this new, high energy light source was soon realized and there was an explosion of storage ring based light sources, capable of operating in the X-ray region. The field was revolutionized, however, with the introduction of high-gain Free Electron Lasers (FELs). These accelerators allowed transversely coherent X-ray pulses to be obtained over a wide wavelength range, and with peak spectral brightnesses up to a billion times higher than pre-existing sources. There now exist several FELs operating in the X-ray regime, which are the LCLS [2], FLASH [3], SACLA [4], FERMI [5], and the more recent PAL [6] and XFEL [7]. These facilities have enabled fantastic advances in the fields of Atomic, Molecular, and Optical (AMO) physics, structural biology, chemistry, and nanoscale dynamics.

The basic principle behind a FEL involves the resonant interaction between a relativistic electron beam and an alternating transverse magnetic field, provided by

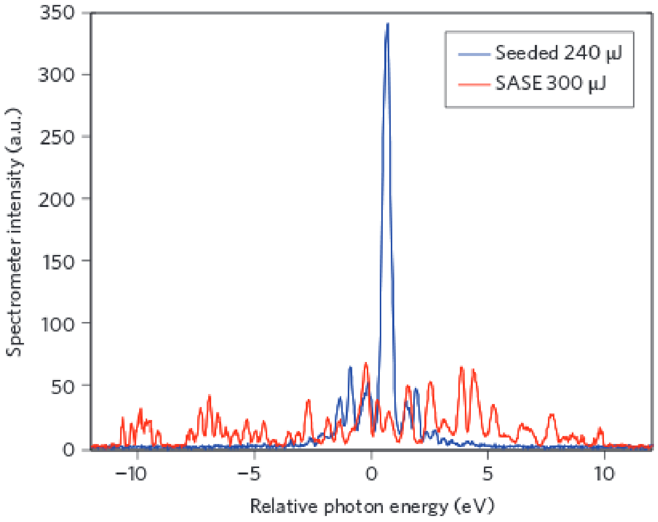


Figure 1.1: A typical SASE spectrum, shown in red, with many spikes and a relatively large bandwidth. By contrast, the blue curve shows FEL output using the so-called Hard X-Ray Self-Seeding scheme, which provides a temporally coherent seed. [8]

an undulator magnet. As the electrons radiate, the electromagnetic field acts back on the electron bunch to push the electrons into microbunches on the scale of the FEL wavelength. These microbunches can then radiate coherently, greatly amplifying the radiation power and thus causing more bunching, which is the essential physics of the FEL instability. This instability relies on the initial presence of either some small amount of microbunching, or a electromagnetic field at the resonant FEL wavelength.

Historically, X-ray FELs have initially operated in the so-called Self-Amplified Spontaneous Emission (SASE) mode, in which the seed for the instability is provided by spontaneous undulator radiation [9] [10]. While simple for operation, the fact that the seed to the instability is due to spontaneous emission, driven by random noise on the electron beam, means that the final output spectrum is not temporally coherent. The result is a temporal pulse that is non-uniform and contains many separate spikes, or in the frequency domain, a power spectrum which also contains these spikes, as shown in figure 1.1. It is clearly desirable, if possible, to seed the process in a coherent way and move towards a single-spike, Fourier-limited power spectrum.

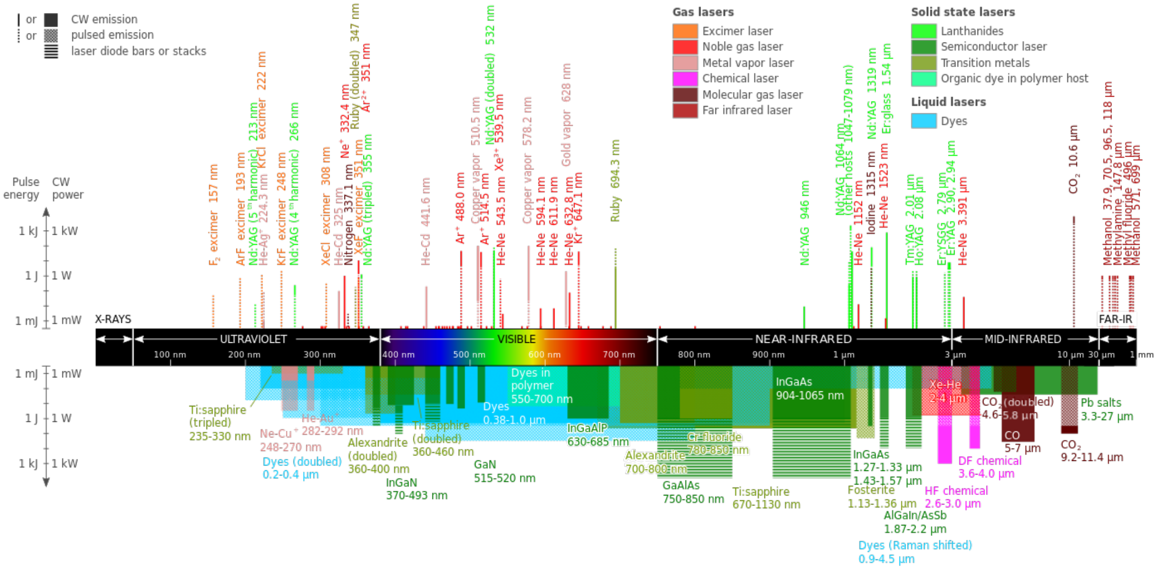


Figure 1.2: A diagram showing the wavelength range of commercially available conventional laser systems. Image from [11].

Such a seeding process is possible in the wavelength regimes accessible by conventional laser systems, which is illustrated in figure 1.2. We can readily observe a dearth of lasers in the UV, limited to perhaps 200 nm.

There exist exotic harmonic generation methods, such as the High-Harmonic Generation (HHG) scheme [12], which can assist in bridging the gap down to X-ray wavelengths. The HHG technique uses the non-linear interaction of a high intensity conventional laser pulse with a noble gas jet to produce very high wavelength harmonics of the initial seed laser. In this way, the production of wavelengths as short as several nanometers is possible, however the attainable pulse energy is rather low. In order to effectively seed the FEL process with a laser, the input power must be sufficient to overcome the noisy SASE startup process, and this is presently not possible for HHG sources below ∼20 nm. Therefore, it is necessary to contrive alternate means in order to coherently seed FELs in the Extreme Ultraviolet (EUV) and soft X-ray regimes.

The alternative to introducing a direct electromagnetic seed is to prepare the electron bunch to possess significant bunching on the radiation wavelength scale.

The bunching at wavevector k for a grouping of N_e electrons, indexed by i with longitudinal positions z_i, is defined as,

$$b(k) = \frac{1}{N_e} \sum_{i=1}^{N_e} e^{ikz_i}. \tag{1.1}$$

The bunching factor describes how piled up the electron distribution is at a particular wavevector, and is also directly related to how much coherent (Power$\sim N_e^2$) radiation the beam would produce at that wavelength. If one can produce bunching through arrangement of the electrons at the resonant wavenumber k of the FEL, the coherent radiation would quickly generate a seed which overwhelms the SASE startup, and the FEL would proceed fully seeded. In the context of FELs, this bunching at the FEL wavelength scale is often referred to as 'microbunching'.

This provides an essentially distinct, beam-based (as opposed to radiation-based), seeding method for the FEL, and offers the ability to reach into the soft X-ray wavelength regime. This microbunching can be produced from the interaction of a conventional laser with the electron beam. Conventional laser pulses are readily available down to $\lambda \approx 266$ nm, so one must provide significant harmonic up-conversion in order to seed in the soft X-rays. One particularly motivated wavelength range, known as the water window, lays between the K absorption edges of carbon (4.37 nm) and oxygen (2.33 nm). Radiation in this wavelength range allows for in-vitro imaging and analysis of biological samples. To provide a seed in the water window therefore requires a harmonic upconversion factor h from 266 nm of $60 \sim 115$. We choose a value somewhere in the middle, $h = 75$, as the target harmonic at which this type of seeding becomes viable in the soft X-rays. This harmonic upconversion factor is established through the Echo-Enabled Harmonic Generation technique in chapter 4 of this thesis.

Besides providing a seed from which the FEL amplification can take place, a relativistic electron beam must also be prepared to act as the lasing medium. The one dimensional theory of the FEL interaction furnishes a single parameter which influences most of the dynamics,

$$\rho \equiv \sqrt[3]{\frac{\mu_0 K^2 \mathcal{J}^2 e^2 n_e}{32\gamma^3 k_u^2 m}}, \tag{1.2}$$

where K, $\mathcal{J}$ and k_u are properties related to the undulator, while n_e is the electron beam volume density and γ its relativistic energy. The parameter ρ controls, among other things, the undulator length required for the amplification to saturate and saturation power of the FEL process. It is therefore generally advantageous to make it as large as possible. Since the electron beam energy and undulator parameters combine to yield the resonance wavelength of the FEL (see equation 2.42), in general the optimization of these parameters requires delicate trade-offs.

However, we can unambiguously state that increasing the electron beam density leads to a more efficient FEL process. This dependence has led to a push towards ever brighter (in the five-dimensional sense) electron beams, both in the transverse dimensions (through the horizontal and vertical emittances) and the longitudinal dimension through bunch compression. The creation of such high density beams, however, is hampered by collective effects in the electron beam which in general work to dilute it. There exist many separate effects which rise and fall in importance as the beam energy and density is changed. It is imperative to understand these collective effects if a high density electron beam is to be maintained so as to promote the most efficient FEL process possible. One major goal of this thesis is to therefore understand these deleterious effects and see how they might be avoided to facilitate the generation of high brightness radiation pulses.

The essential ingredients for an FEL are therefore, 1) the relativistic electron beam (and its associated self-destructive tendencies), 2) the resonant interaction, and instability, inside the long undulator magnet, and 3) the electromagnetic or microbunching seed.

The outline of the thesis is as follows. First, background material for each of the three essential FEL ingredients is reviewed in chapter 2. This thesis will detail novel contributions to both the understanding of limitations in relativistic electron beams and also the preparation of microbunching seeds for the FEL. The experimental portion of this thesis was conducted at the Next Linear Collider Test Accelerator (NLCTA) at the SLAC National Accelerator Laboratory. An overview of the facility and the previous experiments to generate microbunching structure using the Echo-Enabled Harmonic Generation (EEHG) is given in chapter 3 and section 3.1.2. Then, experimental upgrades made to the facility as enabling technologies to generate and probe higher harmonics of microbunching are presented in section 3.2, and a characterization of the NLCTA electron beam is given in section 3.3. The experimental results from the EEHG experiments are then presented and discussed in chapter 4.

I joined the echo program at NLCTA, which had been ongoing since 2009, in late 2014 as the experimental push towards the highest harmonics began. I was heavily involved in three main aspects of the experiment from that point onwards: The installation and commissioning of the new VISA undulator, the design, construction and calibration of a new EUV spectrometer, and the day-to-day electron beam tuning and operation of the NLCTA accelerator.

The results of numerical simulations of EEHG and other beam-based seeding schemes are presented in chapter 5. These simulations include benchmarking of new FEL simulation codes (5.1), new simulations of EEHG interacting with effects such as the microbunching instability (5.2), and start-to-end simulations of the NLCTA (5.3).

A theoretical model for a novel diffusive effect due to synchrotron radiation in bend magnets is developed in chapter 6. This theory is compared with full particle simulations (6.3), and the possible implications of this effect on current and future facilities is discussed (6.4).

Finally, chapter 7 provides a summary of this thesis and concluding remarks.

Chapter 2

Background Material

2.1 Electron Beam Dynamics

Free-electron lasers use relativistic electron beams as the gain medium for the amplification of radiation. Relativistic electron beams have been studied since the earliest days of electrostatic particle accelerators. Since the introduction of the RF accelerating cavity, the structure of an electron beam is generally individual bunches separated by subharmonics of the RF repetition rate. The modern formalism for the discussion of such bunches, and accelerator physics in general, proceeds largely from the pioneering work of Courant and Snyder in 1958 [13].

The coordinate system used in the study of accelerators generally considers particles moving on a curved trajectory, and defines the coordinate axes as shown in figure 2.1. In addition to the three spatial coordinates, we also define the momenta in the x and y direction as p_x and p_y. Finally, for the longitudinal dimension we define the fractional momentum offset with respect to the design particle $\delta = (p - p_0)/p_0$, where $p_0 = mc\sqrt{\gamma_0^2 - 1}$, m is the electron mass, c the speed of light, and $\gamma_0 = (1 - v^2/c^2)^{-1/2}$ is the reference particle relativistic Lorentz factor. These six coordinates form the phase space in which the particles reside, and are encapsulated in the vector $\vec{X} = (x, p_x, y, p_y, z, \delta)$. Occasionally, it is simpler to use the temporal coordinate $t = z/(\beta c)$ where $\beta = v/c$ is the normalized velocity. It is

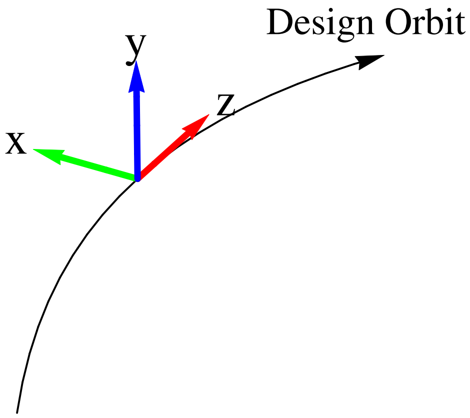


Figure 2.1: The conventional system of coordinates used in accelerator physics. The coordinate x points radially away from the local radius of curvature, z points along the direction of motion, and the vertical coordinate y is perpendicular to these two in order to form a right-handed system.

worth noting that all coordinates excepting z are defined with respect to the reference particle trajectory, such that a particle precisely on the design trajectory has $\vec{X} = (0, 0, 0, 0, z, 0)$.

In all situations we are concerned with, the electron beam has a bunched beam structure. The electrons are photo-emitted from a cathode at the repetition rate of a drive laser pulse, and the resulting bunch has temporal length roughly equivalent to the duration of the laser pulse. This assumes that the length of the drive laser is significantly shorter than the RF wavelength employed in the electron gun, which is true for cases considered in this thesis but in general can be violated high repetition rate accelerators. The individual laser pulse duration can range from tens of femtoseconds (fs) to hundreds of picoseconds (ps), which leads to the bunched beam structure. As they are emitted, the electrons are captured by a high-gradient electromagnetic field ($\sim$ 100 MV/m) and accelerated to relativistic velocities inside an electron gun. For cases considered in this thesis, the EM field is typically provided by an RF klystron in the S-band ($f_{\mathrm{RF}} \approx 2.85$ GHz) frequency range, corresponding to

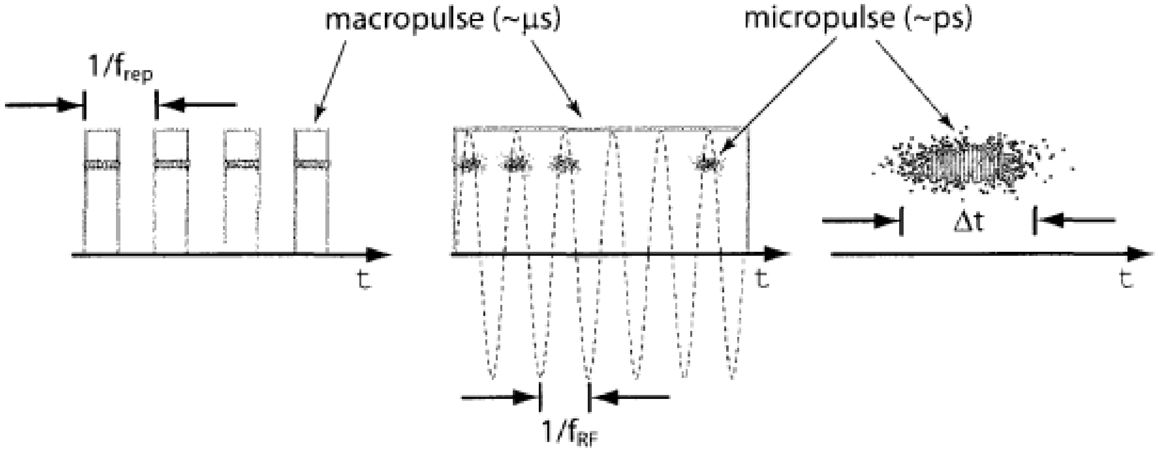


Figure 2.2: The relationship between the different timescales in a bunched beam. Figure from [14].

an RF cycle duration $1/f_{\mathrm{RF}} \approx 350$ps. Furthermore, the RF power from the klystron is also modulated and has typical duration of $\approx \mu$s, with repetition rate of perhaps 100Hz for a normal conducting RF gun. The relationship between these different timescales is shown schematically in figure 2.2.

We are primarily concerned with the dynamics within an individual electron bunch. Therefore, it is useful to define the intra-bunch coordinates,

$$s = z - \bar{\beta}ct, \tag{2.1}$$

where $\bar{\beta}$ is the average electron velocity and t is the time from some reference epoch. The coordinate s therefore naturally on the scale of the electron bunch, and we will frequently refer to it to describe the intra-bunch dynamics. Note that if we use s instead of z as our longitudinal coordinate, the particle on the design trajectory has precisely $\vec{X} = 0$ as its coordinates. From this point forward, our canonical phase space vector will be $\vec{X} = (x, p_x, y, p_y, s, \delta)$.

2.1.1 Multipole Magnetic Fields, Dipoles, and Quadrupoles

Modern accelerators must transport electrons over long distances, either in a ring configuration for many turns, or in a long linear accelerator (linac) through multiple accelerating cavities. The trajectory of the electrons can be controlled using electromagnetic fields, in which the electrons will respond via the Lorentz force,

$$\vec{F} = e\left(\vec{E} + \vec{v} \times \vec{B}\right), \tag{2.2}$$

where e is the electric charge, $\vec{E}$ the electric field, $\vec{B}$ the magnetic field, and $\vec{v}$ the electron's velocity. A relativistic electron moves at approximately the speed of light c, so $|\vec{v}| \approx c$, and we can compare the force due to some typical magnetic and electric fields. A typical magnetic field of 1 Tesla, easily achievable with permanent or electromagnets, is matched in force by an electric field of 3×10^8 V/m. This huge electric field far exceeds the electrostatic breakdown of air and most materials, rendering it impractical in comparison to the magnetic field for the manipulation of relativistic electrons.

Therefore, the dynamics of the electron beam is primarily controlled by the magnetic field, of which we can consider the first two multipole modes. A general magnetic field in cartesian coordinates can be expanded in terms of cylindrical multipole modes as [15],

$$B_y = B_0 \sum_{n=0}^{\infty} \left(\frac{r}{a}\right)^n \left(b_n \cos n\theta - a_n \sin n\theta\right), \tag{2.3}$$

$$B_x = B_0 \sum_{n=0}^{\infty} \left(\frac{r}{a}\right)^n \left(a_n \cos n\theta + b_n \sin n\theta\right), \tag{2.4}$$

where $x = r\cos\theta$, $y = r\sin\theta$, B_0 is the on-axis ($r = 0$) magnetic field, a is a reference radius, and the coefficients a_n and b_n are called the multipole coefficients. While a general magnetic field can have nonzero coefficients for all a_n,b_n, in the context of accelerator physics magnets are typically designed to have only a single nonzero coefficient. Furthermore, for a magnet, a_n and b_n are a related by a simple rotation about the direction of electron propagation (or equivalently, rotation in the coordinate system). The coefficients b_n are called normal multipoles, as they tend to bend the electrons in the normally defined propagation direction, while the coefficients a_n are skew multipoles. For our purposes, we consider only normal multipole magnets, for which $a_n = 0$ for all n.

The first possible configuration is given by $b_0 = 1$, and all other $b_n = 0$, in which case the magnetic field is given only by,

$$B_y = B_0, \tag{2.5}$$

while $B_x = 0$. This constant magnetic field is known as a dipole field, and via the Lorentz force, will attempt to curve the electron's trajectory along a circular path. In practice, this often takes the form of a series of dipole magnets used to bend the electron trajectories into a closed, circular arc, as in a synchrotron.

The next possible elementary magnetic configuration has only $b_1 = 1$. The field for this configuration is given by,

$$B_y = B_0 \frac{r}{a} \cos\theta = B_0 \frac{x}{a}, \tag{2.6}$$

$$B_x = B_0 \frac{r}{a} \sin\theta = B_0 \frac{y}{a}. \tag{2.7}$$

A plot of the field vectors is shown in figure 2.3. As can be seen by the emergence/convergence of field lines, there are north magnetic poles in the top left and bottom right, while there are south magnetic poles in the lower left and upper right. Due to these four poles, this configuration is called a quadrupole magnet. There is no field on-axis, as expected, but small displacements from the origin will result in linear restoring (or diverging) forces, as can be easily understood from the Lorentz force equation 2.2. In the case of the field configuration shown, for an electron propagating into the page, the quadrupole is restoring in the vertical direction and diverging in the horizontal. A 90° rotation about the direction of electron propagation would produce the opposite restoring/diverging directions. For a distribution of electrons, then, the quadrupole magnet focuses the distribution in one direction, while defocusing it in

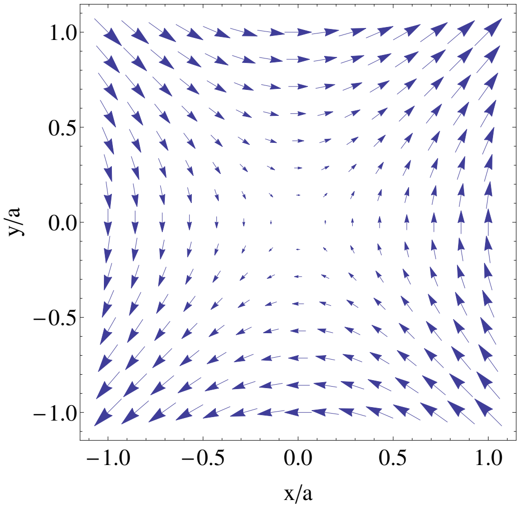


Figure 2.3: The magnetic field plot of a pure-quadrupole magnet.

the other. A combination of quadrupoles, however, can have the combined effect of focusing in both the horizontal and vertical planes, enabling the transport of electron beams over large distance without significant divergence.

There exist, of course, an infinite family of potential magnets such as the $b_2 = 1$ sextupoles, or $b_3 = 1$ octupoles. These higher-order magnets are generally used to reduce chromatic aberrations in the optical lattice which are introduced by pure quadrupole fields. We will, however, not concern ourselves with these higher-order magnetic field components, as the dipole and quadrupole are sufficient for all the beam manipulations we will consider.

2.1.2 The Beam Transport Matrix

Consider a particle offset from the reference trajectory by vector X_i where $i = \{1, \cdots, 6\}$ describe the 6 phase space coordinates. Then, the effect of any forces on the particle trajectory can be written in a power series expansion as,

$$X_i^{\text{final}} = R_{ij} X_j + T_{ijk} X_j X_k + U_{ijkl} X_j X_k X_l + \cdots , \tag{2.8}$$

where Einstein summation convention has been used in which repeated indices indicate summation: $R_{ij}X_j \equiv \sum_j R_{ij}X_j$. Since the coordinate deviations X_i are generally small, the series can usually be truncated after the first term.

The term 'generally small' requires some explanation, as the phase space vector X_i in general has dimensions, and thus must be small compared to some reference scales. The momenta p_x and p_y are small when compared to the total momentum p_0, usually referred to as the paraxial approximation, and is almost always a good approximation. The transverse coordinates x and y can be considered small compared to the aperture for typical magnetic components – for example, when compared to the good-field region in a quadrupole magnet. The longitudinal coordinate s should similarly be considered small compared to the longitudinal distance in the component – typically in comparison to the radio-frequency wavelength λ_{RF}.

In the case that the evolution of the electron beam phase space is linear, this is called *linear optics*. We can rewrite equation 2.8 in the vector and matrix form,

$$\vec{X}^{\text{final}} = \mathbf{R} \cdot \vec{X}. \tag{2.9}$$

The matrix $\mathbf{R}$ is then known as the transport matrix, and describes the linear evolution of the electron beam phase space through any given beamline element. The components of $\mathbf{R}$ are found by considering how each individual coordinate is transformed through the field of a given element in a straightforward procedure. All components of $\mathbf{R}$ are not independent, however, as the matrix is required to be symplectic [16]. In fact, for magnetic components with symmetry about $y = 0$ and no substantial radiation, a common situation in accelerator physics, the general from for the transport matrix is,

$$\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} & 0 & 0 & 0 & R_{16} \\ R_{21} & R_{22} & 0 & 0 & 0 & R_{26} \\ 0 & 0 & R_{33} & R_{34} & 0 & 0 \\ 0 & 0 & R_{43} & R_{44} & 0 & 0 \\ R_{51} & R_{52} & 0 & 0 & 1 & R_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.10}$$

Furthermore, by virtue of the linear theory, once the transport matrix is known for an individual beamline element, the total transport matrix can be constructed by consecutive multiplication of matrices,

$$\vec{X}^{\text{final}} = \mathbf{R}_j \mathbf{R}_{j-1} \cdots \mathbf{R}_1 \cdot \vec{X} = \mathbf{R}^{\text{total}} \cdot \vec{X}, \tag{2.11}$$

where the component 1 is the first encountered by the electron beam, all the way up to component j, and their successive multiplication defines the total transport matrix $\mathbf{R}^{\text{total}}$. With this in mind, we can list components for some common beamline elements and simply multiply matrices to construct any beamline composed of them. The transport elements for some common magnets are given in table 2.1 [17].

A word about each beamline component is in order. The drift space is simply a length L in the absence of any magnetic field. In this situation, transverse momenta are converted into transverse position (nonzero R_{12},R_{34}), and particles with different momenta move longitudinally with respect to one another simply by virtue of their differing velocities (nonzero R_{56}). This difference in velocities, being a non-relativistic effect, is proportional to γ^{-2}, which usually implies it can be ignored for high energy electron beams.

The sector dipole magnet bends the particle through an angle $\theta = L/\rho_0$ where ρ_0 is the bending radius and L the arc-length. The magnet is curved in the shape of the arc, such that the particle enters and exits the magnetic field perpendicular to the pole face. Note that this element too possesses non-zero R_{56}, although generally of opposite sign to the drift. Intuitively, this means that particles with higher momenta

R_{ij}	Drift	Sector Dipole	Rectangular Dipole	Thick Quadrupole	Thin Quadrupole
R_{11}	1	$\cos(\theta)$	1	$\cos(kL)$	1
R_{12}	L	$\sin(\theta)\rho_0$	$\sin(\theta)\rho_0$	$\sin(kL)/k$	0
R_{16}	0	$(1-\cos(\theta))\rho_0$	$(1-\cos(\theta))\rho_0$	0	0
R_{21}	0	$-\sin(\theta)/\rho_0$	0	$-k\sin(kL)$	$-1/f$
R_{22}	1	$\cos(\theta)$	1	$\cos(kL)$	1
R_{26}	0	$\sin(\theta)$	$2\tan(\theta/2)$	0	0
R_{33}	1	1	$1-\theta\tan(\theta/2)$	$\cosh(kL)$	1
R_{34}	L	L	L	$\sinh(kL)/k$	0
R_{43}	0	0	$\frac{\tan\left(\frac{\theta}{2}\right)\left(\theta\tan\left(\frac{\theta}{2}\right)-2\right)}{\rho_0}$	$k\sinh(kL)$	$1/f$
R_{44}	1	1	$1-\theta\tan(\theta/2)$	$\cosh(kL)$	1
R_{51}	0	$\sin(\theta)$	$-2\tan(\theta/2)$	0	0
R_{52}	0	$(1-\cos(\theta))\rho_0$	$(1-\cos(\theta))\rho_0$	0	0
R_{56}	L/γ_0^2	$(\theta-\sin(\theta))\rho_0+L/\gamma_0^2$	$(\theta-\sin(\theta))\rho_0+L/\gamma_0^2$	L/γ_0^2	L/γ_0^2

Table 2.1: First order transport elements for common beamline components.

arrive at the end of the bend *after* those with lower momenta. This is due to the curved trajectory, in which higher momentum particles are deflected less than those with lower momenta, and thus travel on longer paths. For most relativistic machines, this effect is much larger than the velocity difference of the L/γ_0^2 term.

The rectangular dipole, by contrast, is a simple rectangle through which the particle is bent through an angle θ. At the entry and exit pole faces, there is an additional contribution to the total transport. This pole face contribution can be calculated by treating the pole face as a thin quadrupole with focal length $\rho_0/\tan(\theta/2)$.

The quadrupole of length L is characterized by an inverse length scale k, which describes the strength of the focusing. A thick quadrupole is defined to be one in which the change in transverse position inside the quadrupole is appreciable, in which case the full matrix should be used. Many times, however, the effect of the quadrupole can be treated as an impulse and replaced by a thin lens optical element. This lens is characterized by a focal length f, which is formally derived by taking the limit $L \to 0$ as the product kL remains finite. It is clear to see that the quadrupole is focusing in one plane while defocusing in the other.

2.1.3 The FODO Lattice and Magnetic Chicane

With these simple components, we will generate two common magnetic structures that will be used throughout this work. The first is the four-dipole magnetic chicane, while the second is the focusing-defocusing (FODO) alternating quadrupole lattice.

The simplest four-dipole chicane contains two sections of two sector dipoles, of opposite bend angle, with a small drift section between them as shown in figure 2.4. The full linear transport matrix can be found by multiplying the successive matrices for the bends and drifts, and in general contains many terms even for this highly symmetric case. As is often the case, however, we can take the angle θ to be small, and the Lorentz factor γ to be large, and arrive at simple lowest order expressions for the transverse and longitudinal dispersion components in the four-dipole chicane,

$$R_{16} = -\theta^3 \frac{(L+L_M)(2L+L_2+4L_M)}{L_M}, \tag{2.12}$$

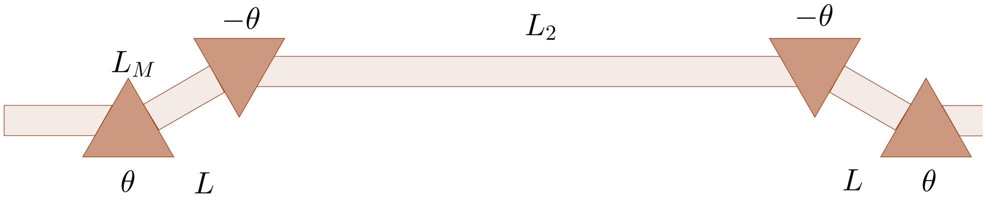


Figure 2.4: A four-dipole symmetric chicane with bends of angle θ, opposite bend magnet spacing L, inner drift distance L_2, and bend magnet length L_M.

$$R_{56} = \theta^2 \left(2L + \frac{4}{3} L_M \right). \tag{2.13}$$

This result, for a sector dipole chicane, reveals an asymmetry due to the presence of a nonzero R_{16} component. By contrast, when the chicane is composed of rectangular dipoles, R_{16} is identically zero, while the same equation holds for R_{56}. For this reason, rectangular dipole chicanes are often employed in linear accelerators to automatically zero this often unwanted horizontal dispersion component (see section 4.1.1).

For large deviations from the design energy, however, one may be forced to consider the effect of the non-linear transport through the four dipole chicane. In this case, we must use the second-order transport tensor T_{ijk}, which is nonzero for the bend magnets, to propagate the beam through the chicane. The relevant result is non-zero T_{166} and T_{566} quadratic transverse and longitudinal dispersion components, again given to lowest order in θ by,

$$T_{166} = \theta^3 \frac{2(L + L_M)\,(2L + L_2 + 4L_M)}{L_M} = -2R_{16}, \tag{2.14}$$

$$T_{566} = -\theta^2\,(3L + 2L_M) = -\frac{2}{3} R_{56}. \tag{2.15}$$

The second order transport components for the symmetric chicane, to lowest order in θ, are simply related to the first-order transport components. Note that again,

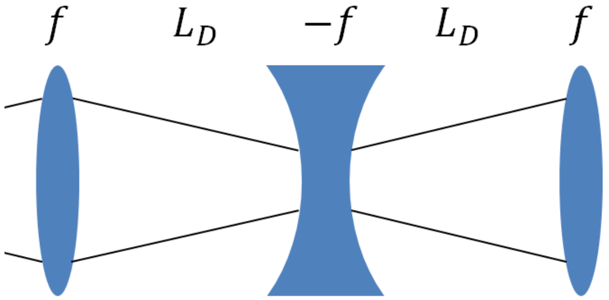


Figure 2.5: A FODO quadrupole lattice with quadrupoles of focal length f and a drift length L_D. The black rays indicate particle trajectories in one transverse plane, showing the focusing and defocusing effect of the quadrupole lenses.

when rectangular dipoles instead of sector dipoles are used, the nonlinear horizontal dispersion vanishes while the same relationship with T_{566} holds. Also of note is the sign difference between T_{566} and R_{56}, indicating that the total dispersion as second order effects begin to become relevant will be lower than the simple first-order (R_{56})estimate.

The second relevant magnetic structure is the FODO quadrupole lattice. In this configuration, quadrupoles of opposite sign (focusing first, then defocusing) are repeated with a drift length in between, shown schematically in Figure 2.5. The FODO cell is traditionally analyzed in a symmetric fashion, beginning the transport in a half focusing quad ($f \rightarrow 2f$) and ending in the other half of the focusing quad. In this case, the (two-dimensional) transport matrix, $R^{(2)}$, in one transverse plane in the thin lens approximation is given by,

$$R^{(2)}_{\mathrm{FODO}} = \begin{pmatrix} 1 - \frac{L_D^2}{2f^2} & 2L_D\left(1 + \frac{L_D}{2f}\right) \\ -\frac{L_D}{2f^2}\left(1 - \frac{L_D}{2f}\right) & 1 - \frac{L_D^2}{2f^2} \end{pmatrix}. \tag{2.16}$$

The fortunate result is that the combination of focusing and defocusing quadrupole lenses results in a net focusing in both planes, assuming that $f \gg L_D$. A more sophisticated analysis of the stability of such a lattice is commonly carried out in introductory texts, with the result that such stable, net focusing motion occurs when

$f > \frac{L_D}{2}$. This stable, net focusing arrangement can be used to transport beams over long distances, maintaining a relatively constant beam size.

2.1.4 Ensembles of Particles, Twiss Parameters, and Emittance

The discussion of the motion of electrons in the presence of magnetic fields has thus far concerned only a single electron. Of course, in most situations one has a rather large number of electrons contained within a single bunch. For example, the LCLS nominal operating charge per bunch is 250 pC [18], which is around 1.5×10^9 electrons per bunch.

To begin with, we take a slightly more analytical tack towards the motion of a particle in a periodic magnetic field lattice, following [15]. The most general linear differential equation for such a system is given by Hill's equation,

$$u'' + k_u(z)u = 0, \tag{2.17}$$

where u is a stand-in for either x or y, and the restoring force k_u is generally a function of the distance along the trajectory z. In the case of a punctuated magnetic lattice, for example, $k_u(z) = 0$ in drift spaces and is nonzero inside quadrupole magnets which exert a restoring (or diverging) force on the particle.

The most general solution is given by,

$$u(z) = Aw(z)\cos\Psi(z) + Bw(z)\sin\Psi(z), \tag{2.18}$$

where the amplitude function $w(z)$ is periodic, the constants A and B are determined by initial conditions, and there is a phase advance in the trigonometric functions $\Psi(z)$ with respect to an initial phase Ψ_0. The (normalized) phase advance $\Psi(z)$ is related to the amplitudes $w(z)$ by,

$$\Psi(z) = \int_0^z \frac{dz'}{w(z')^2} + \Psi_0. \tag{2.19}$$

The solution for $u(z)$ and its derivative can then be rewritten as,

$$u(z) = \sqrt{\mathcal{W}\beta(z)}\cos(\mu(z) + \mu_0), \tag{2.20}$$

$$u'(z) = -\sqrt{\frac{\mathcal{W}}{\beta(z)}}\left[\alpha(z)\cos(\mu(z) + \mu_0) + \sin(\mu(z) + \mu_0)\right], \tag{2.21}$$

where the phase advance has been replaced by the variable $\mu(z) = \Psi(z) - \Psi_0$, the amplitude is encoded in $\beta(z) \equiv w(z)^2$, and the function $\alpha(z) \equiv -w(z)w1'(z)$ encodes the correlation between the amplitude and derivative. The function $\beta(z)$ is called the *betatron function*, $\mu(z)$ as the *betatron phase*, and $\alpha(z)$ is sometimes known as the *correlation function.* We additionally define the related quantity $\gamma_T \equiv \frac{1+\alpha^2}{\beta}$. Together, the variables α, β, and γ_T are known as the *Twiss parameters.* The quantity $\mathcal{W}$ now encodes the initial (non-phase) conditions, and is found to be an invariant of motion given in terms of the initial conditions by,

$$\mathcal{W} = \gamma_{T0}u_0^2 + 2\alpha_0 u_0 u_0' + \beta_0 u_0'^2. \tag{2.22}$$

The quantity $\mathcal{W}$ is known as the *Courant-Snyder invariant* [13]. The Courant-Snyder invariant defines a skew-ellipse in $u - u'$ phase space along which the particle moves as it traverses the periodic lattice. The phase space ellipse, as well as some points of interest, are shown in figure 2.6.

Now, in the case of linear optics, we can also define functions α and β which depend only on the lattice elements in the system. These lattice functions are often taken to be periodic, as within a repeating FODO lattice. When an electron with equivalent α and β to this lattice solution is injected, its trajectory will follow the evolution of the lattice functions. In this case, the electron beam is said to be *matched*, and we need only study the lattice functions α and β rather than those Twiss parameters which pertain to the actual electrons.

For a matched beam, each particle only has a different value for $\mathcal{W}$, and starting position (u_0, u_0'), although its own phase space ellipse is aligned with the ellipse of every other particle. Therefore, for an ensemble of particles, we can define a large

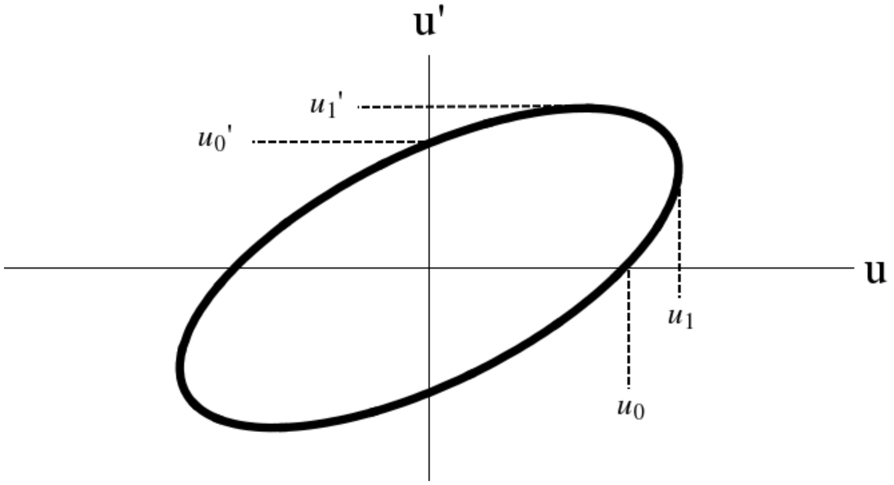


Figure 2.6: The ellipse in phase space defined by the Courant-Snyder invariant. The relevant points are given by $u_0 = \sqrt{\mathcal{W}/\gamma_T}$, $u_1 = \sqrt{\mathcal{W}\beta}$, $u_0' = \sqrt{\mathcal{W}/\beta}$, $u_1' = \sqrt{\mathcal{W}\gamma_T}$.

enough ellipse that contains some fraction of the total beam, defining the Courant-Snyder invariant for this ellipse as,

$$\epsilon = \gamma_T u^2 + 2\alpha u u' + \beta u'^2. \tag{2.23}$$

This quantity ϵ is known as the *beam emittance*, and is a property of the ensemble of particles, not the individual particles. Note that for an unmatched beam, the ellipses of the individual particles would generally be tilted with respect to one another, leading to a larger definition of this emittance than in a matched beam. The emittance, much like the Courant-Snyder invariant, is an invariant of motion as the particle ensemble traverses the magnetic lattice. This fact is guaranteed by the Liouville theorem, in which the emittance represents the phase space area occupied by the distribution of particles. This phase space area is strictly conserved under Hamiltonian systems, and therefore does not change as the beam propagates through magnetic elements.

Different conventions exist to relate the emittance to beam quantities, but we here define the emittance to be the RMS emittance, which defines the 1σ particle distribution, or in 2D space, contains about 39% of the particles. In this case, the emittance is related to the RMS quantities of the beam distribution by,

$$\sigma_u = \sqrt{\beta\epsilon}, \tag{2.24}$$

$$\sigma_u' = \sqrt{\gamma_T\epsilon}. \tag{2.25}$$

Where γ_T in equation 2.25 refers to the Twiss gamma function function, not the Lorentz gamma. It is important to note that this RMS beam emittance is not necessarily the same as the geometric emittance (area of phase space) referred to in the Liouville theorem. Therefore, although Liouville assures us that the true emittance is invariant, the RMS emittance may grow in the presence of Hamiltonian forces that are non-linear or a mismatch in input beam parameters when compared to the lattice [19]. The non-linear forces can be produced by magnetic imperfections, chromatic or geometric aberrations, [20], space charge forces [21], wakefields due to nearby structures [22], or a variety of other sources. Regardless of the source, it is the RMS emittance which is generally important to experimental procedures, and thus care must be taken to avoid or mitigate any potential source of unwanted RMS emittance growth.

The discussion so far has centered on a particle beam traversing only periodic magnetic components, which do not alter the beam energy. It turns out that the Courant-Synder invariant, or the emittance, is only invariant when the beam energy is constant. However, the normalized quantity,

$$\epsilon_N = \epsilon\beta\gamma, \tag{2.26}$$

is conserved even through changes in the total particle energy (under the mild assumption that changes in energy occur slowly compared to the betatron period), and is called the normalized emittance. Note that this implies that the true emittance $\epsilon = \epsilon_N/\beta\gamma$ decreases with beam energy, a process called adiabatic damping. It is important to note that the quantities β and γ in equation 2.26 are the relativistic quantities, not the Twiss parameters – an unfortunate overloading of variables which has confused many a student of accelerator physics. In situations where the particle energy changes over the region of interest, then, it is more convenient to use the quantity ϵ_N.

A final important remark regarding the emittance is that, due to the Liouville theorem, one cannot decrease the emittance of a particle beam using simple magnetic components. This is especially true of the RMS emittance, which as has been discussed, may actually grow through a number of means. One important exception to this is the process of synchrotron radiation, which has the effect of damping (or exciting) the emittance in the plane perpendicular to the bend and, for example, can lead to extremely small vertical emittances in electron storage rings [23]. Especially for linear machines, however, the ultimate limit to achievable emittance is set by the emittance of the beam at its moment of conception, either at a thermionic or photo-excited cathode. Once the particle distribution has been defined in this state, for linear machines this remains the limit to achievable emittance, and is a major driver in FEL performance. Therefore, there is just as much work on the origin of the electrons used in a modern FEL, to ensure a low normalized emittance, as there is in providing a careful beam transport to as best as possible preserve this emittance down the beamline to the radiating sections.

2.2 Free Electron Laser Physics

The fundamental process in the free-electron laser involves a bunch of electrons traversing a magnetic undulator. The core component of the magnetic undulator is a series of alternating North-South magnetic poles which cause the electron bunch to wiggle (undulate) back and forth in response to the fields. The magnetic poles can either be arranged in a linear relationship, causing the undulating motion to occur in a single plane, or in a fashion to cause a helical trajectory. We will begin with the analysis of the motion of an electron in the planar configuration.

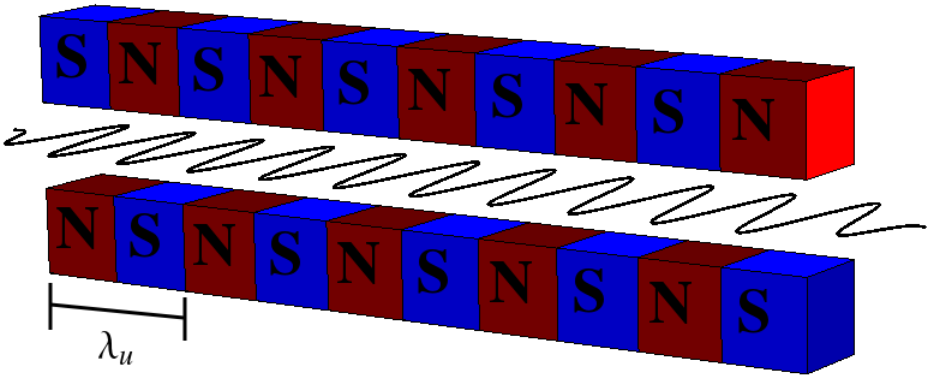


Figure 2.7: A schematic illustration of a planar magnetic undulator with period length λ_u. A representative trajectory traversing the undulator is shown by the black curve.

2.2.1 Spontaneous Undulator Emission

We consider a planar magnetic undulator with period length λ_u, shown schematically in figure 2.7. The magnetic field in the undulator is found by considering the magnetic Poisson equation for the scalar magnetic potential φ_M,

$$\nabla^2 \varphi_M = 0. \tag{2.27}$$

This is the magnetic field in the vacuum interior of the undulator, so no currents are present. We make the ansatz that the magnetic field will be periodic in z, and depend on the vertical coordinate y (leading to primarily horizontal wiggling in the x direction). Under this assumption, we solve for the potential and find the magnetic field in the y direction,

$$B_y(y, z) = B_0 \cosh(k_u y) \cos(k_u z), \tag{2.28}$$

while the requirement that $\vec{\nabla} \cdot \vec{B} = 0$ demands the presence of an z field,

$$B_z(y, z) = -B_0 \sinh(k_u y) \sin(k_u z). \tag{2.29}$$

Now, in typical cases, the displacement y is small compared to the undulator period, and in addition the transverse velocity which couples to this longitudinal field is much smaller than the longitudinal velocity. Therefore, in the following discussions, the longitudinal field is ignored, or equivalently, we consider only the on-axis particle with $y = 0$, and can use the idealized magnetic field, purely in the y direction,

$$\vec{B} = B_0 \cos(k_u z)\hat{y}. \tag{2.30}$$

The Lorentz force on a relativistic electron of mass m and Lorentz factor γ produces the two equations of motion,

$$m\gamma\ddot{x}(t) = -eB_0 \cos\left(k_u z(t)\right) \dot{z}(t), \tag{2.31}$$

$$m\gamma\ddot{z}(t) = eB_0 \cos\left(k_u z(t)\right) \dot{x}(t). \tag{2.32}$$

The first of these can be immediately integrated, together with the initial condition $\dot{x}(0) = 0$, to yield the transverse velocity equation,

$$\dot{x}(t) = \frac{-eB_0}{mk_u}\frac{1}{\gamma} \sin\left(k_u z\right). \tag{2.33}$$

Then, requiring that the electron velocity be βc, the Pythagorean relation gives the z velocity,

$$\dot{z}(t) = \beta c\sqrt{1 - \left(\frac{eB_0}{mck_u}\right)^2 \frac{1}{\beta^2\gamma^2} \sin^2\left(k_u z(t)\right)}. \tag{2.34}$$

The term under the square root represents the slower-than-β motion due to the wiggling trajectory, and the undulator-dependent prefactor is recognized as the undulator K parameter which characterizes the strength of the undulating motion:

$$K \equiv \frac{eB_0}{mck_u}. \tag{2.35}$$

The physical meaning of the K parameter can be seen in taking the ratio of x to z velocities, which coupled with the assumption that the x velocity is much smaller than that in z, or $\gamma \gg 1$, yields the maximum deflection angle ψ of the electron in the undulator,

$$\psi = \frac{1}{\sqrt{\left(\frac{\beta\gamma}{K}\right)^2 - 1}} \approx \frac{K}{\beta\gamma}. \tag{2.36}$$

Recalling that the natural opening angle of synchrotron radiation is $1/\beta\gamma$ [24], we observe that the electron deflection angle is equal to the natural synchrotron opening angle when $K = 1$. Magnetic devices with $K < 1$ therefore radiate into angles defined by their energy, whereas those with $K > 1$ radiate into angles defined by the wiggling motion. In the insertion device community, where the detailed spectral and angular radiation profile is of interest, the distinction is often made to call those devices with $K < 1$ *undulators*, while the more extreme motion in devices with $K > 1$ earns them the name *wigglers*. For the free electron laser, which is primarily concerned with on-axis, highly collimated emission, the distinction is not particularly relevant and all such devices are referred to as undulators. Therefore, we will allow K to be large compared to unity, but maintain the (almost always relevant) approximation that the product $K/\beta\gamma$ be small. Integrating again the equation for z yields the longitudinal trajectory,

$$z(t) = \bar{\beta}ct + \frac{K^2}{8\beta^2\gamma^2 k_u} \sin(2\omega_u t), \tag{2.37}$$

where the averaged electron velocity is defined by,

$$\bar{\beta} \equiv \beta \left(1 - \frac{K^2}{4\beta^2\gamma^2}\right), \tag{2.38}$$

and the undulator angular frequency has been defined as $\omega_u \equiv k_u c\bar{\beta}$. Similarly, we can integrate to obtain the transverse trajectory to lowest order,

$$x(t) = \frac{K}{k_u\beta\gamma} \cos(\omega_u t), \tag{2.39}$$

where we note that the distinction between $\bar{\beta}$ and β here is immaterial, since the correction term would be $\mathcal{O}(K/\beta\gamma)^3$. Before turning to the radiation produced by this motion, we first note that $\bar{\beta}$ can be expanded to lowest order in $1/\gamma$, assuming a fairly relativistic electron. From this averaged motion, we deduce that, due to this slower-than-light velocity and transverse motion, within one undulator period λ_u the electron slips behind a co-propagating photon by a distance,

$$\lambda_{\text{slip}} = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2}\right). \tag{2.40}$$

This slippage length, as it turns out, is important to maintaining a phase-resonant relationship with a co-propagating radiation pulse.

With the trajectory of the particle known, the spontaneous emission of radiation can be computed from the produced electromagnetic fields, for example, from the Liénard-Wiechert field [25]. This derivation is lengthy and not reproduced here, but we will heuristically derive the wavelength of this radiation. The first thing to note is that the averaged velocity $\bar{\beta}$ corresponds to an averaged γ, modified from its nominal value, of,

$$\bar{\gamma} = \frac{\gamma}{\sqrt{1 + K^2/2}}. \tag{2.41}$$

Now, in the rest frame of an electron traversing the undulator (primed frame), the undulator period is Lorentz-contracted to a shortened length $\lambda'_u = \lambda_u/\bar{\gamma}$. In this frame, the electron executes simple harmonic motion with this period, and thus, emits radiation in the primed frame with wavelength λ'_u. In the lab frame, this (forward-traveling) radiation is blueshifted by a factor of $1/2\bar{\gamma}$, yielding a lab-frame radiation wavelength of,

$$\lambda_{\text{rad}} = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2}\right). \tag{2.42}$$

This wavelength is identical to the slippage length from equation 2.40. Therefore, electrons co-propagating with light of wavelength λ_{rad} will be stationary in phase with respect to the light wave, allowing a resonant transfer of energy between the electrons and light. It is worth noting that this result was derived for on-axis forward emission of radiation. When observed off-axis, the Lorentz contraction and blueshift changes slightly, giving an angle-dependent wavelength for small angles θ from the

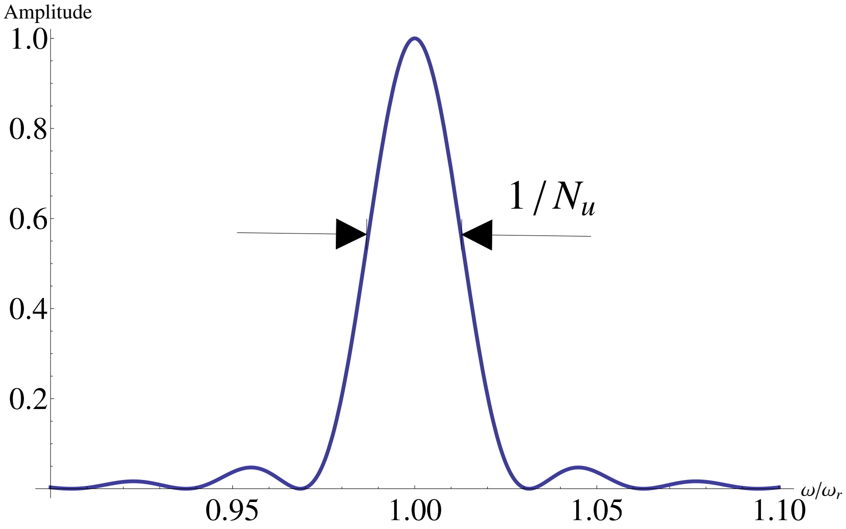


Figure 2.8: The undulator emission power spectrum for radiation of N_u periods centered about frequency ω_r.

axis of propagation,

$$\lambda_{\mathrm{rad}}(\theta) = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2} + \gamma^2\theta^2\right). \tag{2.43}$$

Radiation off-axis is therefore redshifted compared to its on-axis counterpart, and a detector with finite angular acceptance will register a larger bandwidth than the theoretical on-axis calculation would predict.

Finally, we note that an electron traversing an undulator with N_u periods will emit N_u cycles of radiation at wavelength λ_{rad}. Assuming an undulator which abruptly begins and ends, then, the frequency content of the radiation can be obtained by Fourier transforming this field,

$$P(\omega) = \left[\int_{-\infty}^{\infty} E(t) e^{i\omega t} dt\right]^2 = \left[\int_{-\frac{\pi N_u}{\omega_r}}^{\frac{\pi N_u}{\omega_r}} E_0 e^{i\omega_r t} e^{i\omega t} dt\right]^2 = \frac{4N_u^2\pi^2}{\omega_r^2}\mathrm{sinc}^2\left[\frac{N_u\pi(\omega_r - \omega)}{\omega_r}\right]. \tag{2.44}$$

A plot of the spectrum is shown in figure 2.8, with a central peak at frequency ω_r and central bandwidth inversely proportional to the number of undulator periods N_u. This is the characteristic spontaneous undulator emission spectrum, and determines the frequency range over which the electron beam will effectively couple to radiation.

2.2.2 Coherent Emission

The above discussion of spontaneous emission is valid for a single electron, which through radiation produces an electric field on-axis which has the form,

$$E(t) = E_0 e^{i(\omega_r t+\phi)}, \tag{2.45}$$

where E_0 is a constant describing the magnitude of the radiation, and ϕ is an arbitrary phase related to the precise time the electron enters the undulator. Generalizing now to a situation in which N_e electrons are present, we consider the effect of a spread in longitudinal position. The total electric field will be the sum of each individual contribution,

$$E_{\text{total}}(t) = \sum_{i=1}^{N_e} E_0 e^{i(\omega_r t+\phi_i)}, \tag{2.46}$$

where the ith electron has its own phase ϕ_i, and assuming the electron beam is monoenergetic, all the resonant frequencies ω_r and amplitudes E_0 are the same. The total radiation power is proportional to the square of the electric field,

$$P(t) \sim |E_{\text{total}}(t)|^2 = E_0^2 \sum_{i,j} e^{i(\omega_r t+\phi_i)} e^{i(\omega_r t+\phi_j)}. \tag{2.47}$$

The double summation can be split into those terms with $i = j$ and those with $i \neq j$, and we will denote the single electron power $P_e \sim E_0^2 e^{i(\omega_r t+\phi)}$,

$$\begin{array}{rcccc} P(t) & \sim & E_0^2 \sum_{i=j} e^{2i(\omega_r t+\phi_i)} & + & E_0^2 \sum_{i\neq j} e^{i(\omega_r t+\phi_i)} e^{i(\omega_r t+\phi_j)} \\ & = & N_e P_e & + & N_e(N_e-1)P_e \langle e^{i(\phi_i+\phi_j)}\rangle_{i\neq j}^2 . \end{array} \tag{2.48}$$

The second term involves the correlation between the individual phases of emission for each electron, averaged over the entire ensemble of electrons. We can equivalently think about the power in the frequency domain,

$$P(\omega) = N_e P_e(\omega) + N_e(N_e - 1)P_e(\omega)\langle e^{i\omega(t_i+t_j)}\rangle^2_{i\neq j}, \tag{2.49}$$

where the phases can be construed as relative to the frequency ω in that $\phi_i = \omega t_i$ for a timing offset t_i. When the number of particles is large, the discrete averaging implied by $\langle\cdots\rangle$ can be computed as an integral over the normalized charge distribution $\rho(t)$ in time as,

$$F(\omega) \equiv \langle e^{i\omega(t_i+t_j)}\rangle_{i\neq j} \approx \int_{-\infty}^{\infty} \rho(t)e^{i\omega t}dt, \tag{2.50}$$

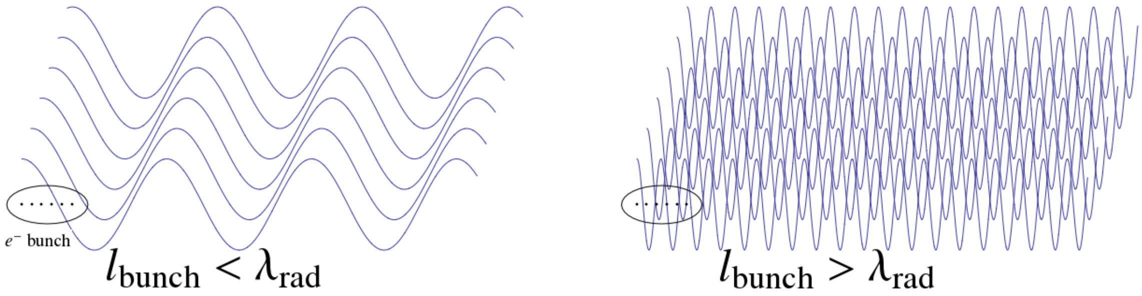
where the frequency dependent function $F(\omega)$ is known as the *form factor*, which is intimately related to the bunching factor of equation 1.1. Typically, the term bunching factor is used when structure exists in the longitudinal phase space at scales shorter than the beam length, while the form factor typically refers to the effects of a uniform, but potentially short beam. We can immediately see that for frequencies where $F(\omega)$ is nonzero, the radiated power will be proportional to N_e^2, which, for $N_e \gg 1$, can often be a factor of 10^6 or greater. In this case, the emission is generally referred to as *coherent*, whereas when the form factor is negligible, the emission is said to be *incoherent*. For example, consider a Gaussian electron bunch with temporal density distribution,

$$\rho(t) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-t^2/2\sigma_t^2}, \tag{2.51}$$

in which case the form factor is computed to be,

$$F(\omega) = e^{-\frac{1}{2}\sigma_t^2\omega^2}. \tag{2.52}$$

We see that for bunches with temporal duration significantly longer than the radiation period (or equivalently, with bunch lengths longer than the radiation wavelength), the form factor is exponentially small. Therefore, Gaussian bunches will radiate coherently at wavelengths comparable to and larger than their lengths, and incoherently at shorter wavelengths. A diagrammatic understanding of this principle is shown in figure 2.9. The fact that electron bunches are readily available on the mm scale led



(a) A short electron bunch emitting coherent radiation. The phase of all the individual electromagnetic waves are roughly equal.

(b) A long electron bunch emitting incoherent radiation.

Figure 2.9: A schematic illustration of the difference between coherent and incoherent emission

to the early production of coherent synchrotron radiation in the THz regime [26] [27], coherent Smith-Purcell radiation in the mm region[28], and coherent transition radiation in the far infrared [29]. State of the art bunch compression techniques are able to produce electron bunches as short as several fs, reaching down into the optical wavelength range [30]. Exotic proposals based on plasma accelerator concepts [31] [32] [33] or laser acceleration [34] exist which may be able to push electron bunch lengths into the attosecond regime. While advances on this front may push coherence with short bunches down into the EUV, high quality electron beams with lengths approaching the X-ray do not seem on the horizon. Therefore, at present special measures to modulate the bunch density must be taken to obtain coherent synchrotron radiation in the deep UV through to hard X-rays.

In addition to the intensity of coherent radiation scaling with the number of electrons squared, the growth rate of this radiation within an undulator is also distinct [35]. To see this, consider that the total energy U in the undulator radiation can be written as an integral over the spectral energy density,

$$U_{\text{tot}} = \int d\omega U_{1e}(\omega) N_e \left(1 + (N_e - 1) F(\omega)\right), \tag{2.53}$$

where $U_{1e}(\omega)$ is the spectral energy density emitted by a single electron. In the case of an un-bunched beam, $F(\omega) = 0$, and $U_{1e}(\omega)$ is given by a form similar to the power density of equation 2.44. In this case, $U_{1e}(\omega) \sim N_u^2$, but the sinc function limits the integration to a $1/N_u$ bandwidth, so the resultant energy has the dependence,

$$U_{\text{tot}}^{\text{un-bunched}} \sim N_u N_e, \tag{2.54}$$

which is the expected and intuitive result. By contrast, for a strongly bunched beam, the bunching from $F(\omega)$ can be much narrower than the natural $1/N_u$ spectral line width from the undulator length. This situation commonly occurs if the electron beam is relatively long, in terms of modulation periods, compared to the number of undulator periods. By modulation period, we mean that there is a density modulation on the electron beam with harmonic content at the undulator resonant frequency ω_r, but which may be periodic with a smaller frequency ω_r/h. In this case, the limit of frequency integration is determined by the electron bunch length, and we have,

$$U_{\text{tot}}^{\text{bunched}} \sim N_u^2 N_e^2. \tag{2.55}$$

As previously derived we see the total energy is proportional to N_e^2, but we now also realize that the energy grows quadratically along the length of the undulator. This unintuitive result is only correct until the number of undulator periods traversed rivals the electron beam length (in modulation periods), or the radiation field becomes strong enough to significantly affect the longitudinal phase space of the electron beam (as in a high-gain FEL). Owing to the fast nature of the radiation growth, this is referred to as the *superradiant* regime.

We note that the above discussion has been in reference to the degree of longitudinal coherence, with which this work is chiefly concerned. However the radiation may also possess coherence in the transverse plane. Transverse coherence is understood as the degree of far-field interference of the radiation from electrons at different transverse positions at the time of emission, shown in figure 2.10. From the classical theory of synchrotron radiation, it is demonstrated that transverse coherence relies

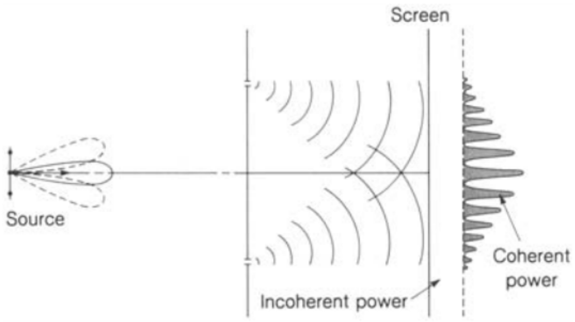


Figure 2.10: A schematic illustration of transverse coherence. The degree of interference in the far field is a measure of the transverse coherence of the produced light. Figure taken from [36].

on the electron beam emittance satisfying,

$$\epsilon < \frac{\lambda_r}{4\pi}, \tag{2.56}$$

where the quantity $\frac{\lambda_r}{4\pi}$ is recognized as the emittance of a Gaussian radiation beam at wavelength λ_r. In practice, this criterion is typically fulfilled for single pass free electron lasers, and even when it is not, optical mode guiding can still yield full transverse coherence [37]. Therefore, in what follows we will use coherence to exclusively refer to the longitudinal coherence of the radiation pulse.

2.2.3 Laser-Electron beam interaction

The principal dynamics in a free electron laser is the resonant interaction between the relativistic electrons and electromagnetic wave. Before detailing this full case, we describe the interaction between the relativistic electron beam and a co-propagating laser beam. In this limited case, the laser beam power is assumed to remain constant with propagation, and we study only its effects on the electron beam phase space.

The laser beam is assumed to be produced by a conventional laser system, operating at the lowest-order Gaussian mode of the laser at wavelength $\lambda_L = 2\pi/k_L$ in which the electric field can be written as [38],

$$E(r,t,z) = E_0 \frac{w_0}{w(z)} e^{-\frac{r^2}{w(z)^2}} e^{-i\left(k_L z - \omega t + k_L \frac{r^2}{2R(z)} - \phi(z) + \phi_0\right)}, \tag{2.57}$$

with amplitude E_0 and $w(z)$ is the evolving beam waist size,

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R}\right)^2}, \tag{2.58}$$

where w_0 is the beam size at the waist, and $z_R = \pi w_0^2/\lambda_L$ is the Rayleigh range (analogous to the Twiss β function). The wavefront curvature is given by $R(z) = z\left(1 + \left(\frac{z_R}{z}\right)^2\right)$, the Guoy phase $\phi(z) = \arctan\left(\frac{z}{z_R}\right)$, and ϕ_0 is an arbitrary phase. This electric field can be considered in an arbitrary direction since polarization has not be specified, but we will typically take the polarization to be perpendicular to the direction of electron propagation.

While this is the general from of the Gaussian laser pulse, we consider the useful case in which both the transverse and longitudinal dimensions of the laser pulse are large compared to the electron beam, and equation 2.57 reduces to a plane wave,

$$E(t,z) = E_0 e^{-i(k_L z - \omega_L t + \phi_0)}. \tag{2.59}$$

We want to consider the effect of an electron co-propagating with this electric field inside a magnetic undulator. The trajectory will be primarily determined by the undulator field in the limit where the dimensionless laser parameter $a_0 = eA_0/m_e c \ll 1$, where A_0 is the vector potential of the laser field. This limit essentially corresponds to the limit in which an electron at rest will not be significantly accelerated by the passing laser field, and instead will simply execute harmonic motion. Further, in the lab frame this harmonic motion component is small and rapid compared to the undulator motion (since $\omega_L \gg \omega_u$), and will be neglected in what follows. The electron trajectory is therefore simply given by equation 2.39 and is not influenced by the laser field.

The energy γ transferred to the electrons from the laser field up to a time t is computed as,

$$\frac{d\gamma}{dt} = \frac{e}{mc^2}\left(\vec{v}\cdot\vec{E}\right). \tag{2.60}$$

We can swap the t derivatives for z derivatives since $\frac{d\gamma}{dz} = \frac{d\gamma}{dt}\frac{dt}{dz} \approx \frac{d\gamma}{dt}\frac{1}{\bar{\beta}c}$. Further, considering the real part of the electric field with dispersion relation $\omega_L = ck_L$, we have the energy modulation equation,

$$\frac{d\gamma}{dz} = -\frac{eKE_0}{\gamma mc^2}\sin(k_u z)\sin(k_L(z-ct)). \tag{2.61}$$

We want to consider this interaction in the co-moving frame with the averaged electron velocity described by the intra-bunch coordinate $s = z - \bar{\beta}ct$. Invoking equations 2.37 and 2.38, we can rewrite $(z-ct)$ in terms of the intra-bunch coordinate to lowest order in γ^{-2},

$$z - ct = \frac{s}{\bar{\beta}} - \frac{z}{2\gamma^2}\left(1+\frac{K^2}{2}\right) + \frac{K^2}{8\gamma^2 k_u}\sin(2k_u z). \tag{2.62}$$

We now specify to the case where the laser frequency is tuned to the undulator resonance, defined by the slippage criterion 2.40, but allow the laser to be an integer harmonic h of the fundamental undulator wavenumber,

$$k_L = hk_u\frac{2\gamma^2}{1+\frac{K^2}{2}}, \tag{2.63}$$

The nested sine functions implicit in equation 2.61 are dealt with through the following Bessel expansion,

$$\sin(a - b\sin x) = \sin a\left[2\sum_{n=0}^{\infty}J_{2n}(b)\cos(2nx)\right] - \cos a\left[2\sum_{n=0}^{\infty}J_{2n+1}(b)\sin(2nx)\right]. \tag{2.64}$$

Finally, we are not concerned with the dynamics on a sub-period scale, so we choose to average equation 2.61 over an undulator period. The result is that only those modulations with odd harmonic average to a nonzero value. The first summation selects out only the Bessel modes with $n = \lfloor\frac{h+1}{4}\rfloor$, while the second picks out $n =$

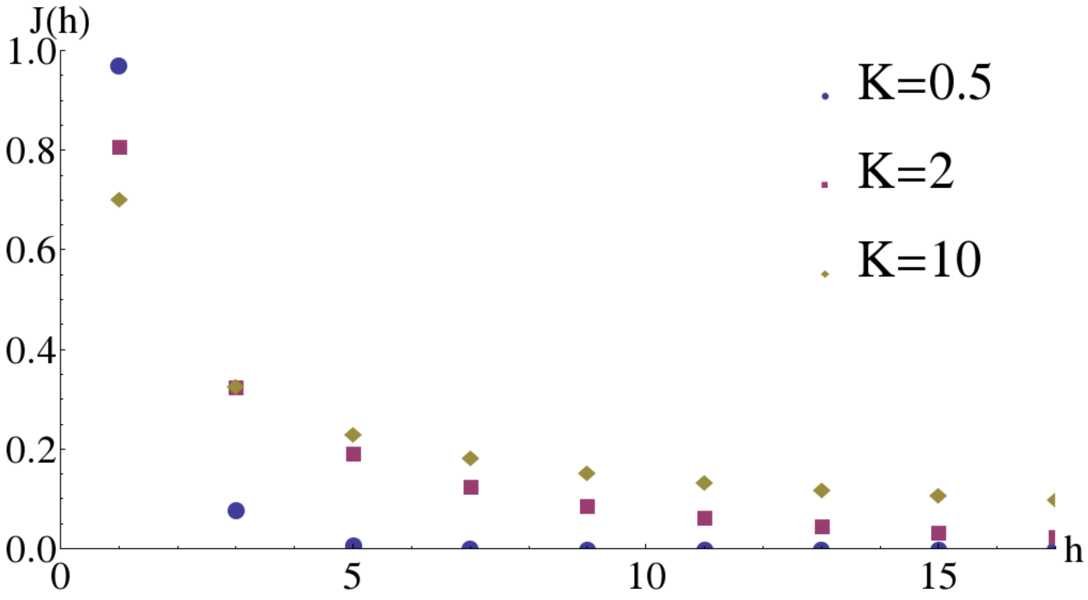


Figure 2.11: The undulator coupling factor $\mathcal{J}(h)$ for several values of K. For large K such that $\xi \approx 1$, the coupling decays only slowly with harmonic number.

$\lfloor \frac{h-1}{4} \rfloor$, with overall sign $(-1)^{(h+1)/2}$. The period averaged result is then given by,

$$\frac{d\gamma}{dz} = \frac{eKE_0\mathcal{J}(h)}{2\gamma mc^2} \cos\left(\frac{k_L s}{\bar{\beta}}\right), \tag{2.65}$$

where we have defined the harmonic dependent $\mathcal{J}$ factor as,

$$\mathcal{J}(h) \equiv (-1)^{(h+1)/2} \left[J_{\left(\frac{h+1}{2}\right)}\left(h\frac{\xi}{2}\right) - J_{\left(\frac{h-1}{2}\right)}\left(h\frac{\xi}{2}\right) \right], \tag{2.66}$$

where $\xi \equiv \frac{K^2}{2+K^2}$. First, we note that since $\bar{\beta}$ is roughly constant for energy changes small compared to the energy in the electron beam, the electron beam energy modulation will acquire the sinusoidal structure of the laser field. Second, through the factor $\mathcal{J}$, we see that it is possible in a planar undulator to couple only to odd harmonics of the undulator harmonic frequency, and this coupling can be non-negligible for large K. A plot of $\mathcal{J}(h)$ is shown in figure 2.11. We note that not only will a laser at a harmonic of the undulator frequency couple to the electron beam, but the beam itself will also radiate at harmonics of the undulator frequency. This is a particular case of the more general statement (known as the general acceleration theorem [39]) that the possibility of a resonant interaction at a given frequency is intimately related to the production of spontaneous radiation at that frequency.

There are several corrections to equation 2.65 to account for the finite transverse and longitudinal extent of the laser beam [40]. It is necessary to account for these effects to optimize the modulation amplitude, however the fundamental fact of a modulation which mirrors the sinusoidal structure of the laser pulse is preserved. By contrast, if the modulation is strong enough, the longitudinal motion of electrons within the modulating undulator (modulator) will be appreciable with respect to the laser wavelength and this sinusoidal structure is altered. The modulator has a longutidinal dispersion element $R_{56} = 2N_u\lambda_r$, where $\lambda_r = \lambda_L/h$ for harmonic h. Then, the requirement that the differential drift between the accelerated and decelerated electrons be much less than a laser wavelength yields the condition [41],

$$\frac{\Delta\gamma}{\gamma} \ll \frac{1}{4N_u h}. \tag{2.67}$$

We quantify this requirement by defining the parameter $\kappa \equiv 4N_u h\Delta\gamma/\gamma$, and show how the modulation waveform depends on κ in figure 2.12. However, most schemes which seek to use laser modulations to generate coherent bunching also rely on the FEL process, which is sensitive to the induced energy spread. Therefore, in the methods that follow, $\frac{\Delta\gamma}{\gamma}$ is kept small enough that a sinusoidal structure can be assumed.

In the next section, we begin by discussing the one-dimensional FEL, in which the transverse emittance of the electron beam as well as any diffraction effects are negligible. The one-dimensional FEL model captures most of the physics of the FEL amplification, and refinements will be discussed afterwards.

2.2.4 The FEL Mechanism

The FEL is fundamentally based on the interaction between a co-propagating electromagnetic wave and a beam of electrons inside an undulator. As we have just shown,

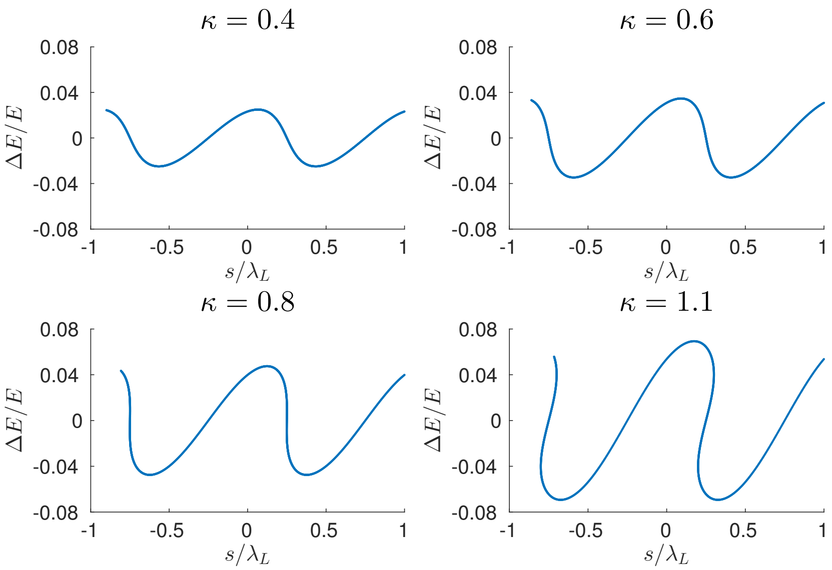


Figure 2.12: The dependence of the modulation waveform on the parameter κ. As κ approaches and exceeds unity, the waveform becomes increasingly non-sinusoidal. This is associated with electrons which slip an appreciable distance in laser phase during the modulation.

in the limit where the electrons do not radiate appreciably, a sinusoidal energy modulation develops on the electron beam. We also demonstrated that this process begins to break down as the induced energy modulation becomes “large”. We now generalize to the case where the electrons contribute to the electromagnetic field, and, following the excellent review of reference [37] and the text of [42], study the coupled system that is the free-electron laser.

We begin by assuming that the motion of an electron in a planar undulator is defined primarily through the the magnetic fields of the undulator, not the co-propagating electromagnetic wave. In this case, the transverse electron velocity is given by equation 2.33. We model the co-proopagating radiation of wavelength λ_L as a plane wave described by $E_x = E_0 \sin(k_L z - \omega_L t + \psi_0)$, with amplitude E_0 and phase ψ_0. The energy change of this electron is then given by,

$$mc^2\frac{d\gamma}{dt} = ev_xE_x = \frac{-eE_0K}{\gamma}\left\{\cos\left[(k_L - k_u)z + \psi_0 - \omega_L t\right] + \cos\left[(k_L + k_u)z + \psi_0 - \omega_L t\right]\right\}. \tag{2.68}$$

We now introduce the phase of the electron relative to the EM wave,

$$\theta = (k_L + k_u)\, z - \omega_L \bar{t}, \tag{2.69}$$

where $\bar{t} = \int dz/\bar{v}_z$ is the electron arrival time averaged over the undulator period, and $\bar{v}_z$ is given by equation 2.38. The change in this phase is then simply found as,

$$\frac{d\theta}{dz} = k_u - k_L\left(\frac{1 + K^2/2}{2\gamma^2}\right), \tag{2.70}$$

which clearly implies that a stationary phase is obtained only when $\gamma = \gamma_0 \equiv \sqrt{\frac{k_L}{2k_u}\left(1 + \frac{K^2}{2}\right)}$, which we then call the resonant energy. Defining the relative energy deviation then to be $\eta \equiv (\gamma - \gamma_0)/\gamma_0$, to lowest order the change in phase is,

$$\frac{d\theta}{dz} = 2k_u\eta. \tag{2.71}$$

That is, the phase drift is linearly proportional to the fractional energy deviation. Returning to equation 2.68, in terms of the variable η and keeping only slowly varying terms, we have,

$$\frac{d\eta}{dz} = \frac{eK\mathcal{J}}{2\gamma_0^2 mc^2}E_0\cos\left(\theta + \psi_0\right), \tag{2.72}$$

where $\mathcal{J} \equiv \mathcal{J}(h = 1)$ as defined by equation 2.66. Equations 2.71 and 2.72 are collectively referred to as the *pendulum equations* due to their shared form with the equations of motion for a simple pendulum mass. The curves described by this motion are shown in figure 2.13. There is a clearly defined bucket of stable, bound motion centered on the points $(2\pi n, 0)$, for integer n. These buckets are separated from the unbound motion by separatrices. The influence of the electromagnetic wave is called the ponderomotive potential, and thus these stable buckets in phase space are known as ponderomotive buckets. We see that inside the buckets, an initial phase offset is

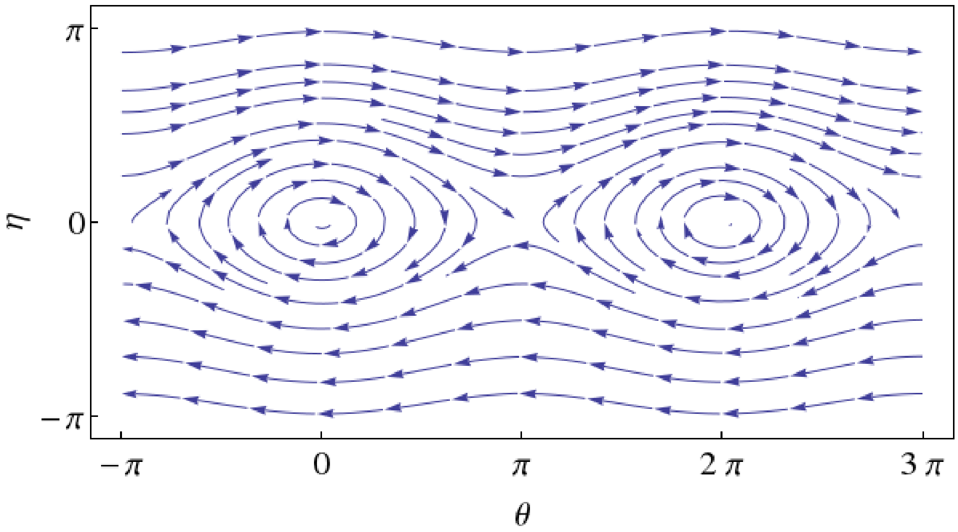


Figure 2.13: Trajectories in (θ,η) space described by the pendulum equations of FEL motion.

transformed into an energy offset, which then converts back into a changing phase. This is the essential phase-energy interplay of the FEL interaction. However, as can be easily seen, the net energy loss of a long electron bunch (many wavelengths in θ) will be zero. It is therefore desirable to *microbunch* the beam at the locations of θ favorable for energy loss (energy transfer to the EM wave).

Turning now to the evolution of the radiation field, its amplitude will obey the Maxwell equation,

$$\left[\left(\frac{1}{c}\frac{\partial}{\partial t}\right)^2 - \left(\frac{\partial}{\partial z}\right)^2 - \nabla_\perp^2\right] E_x(\vec{x},t;z) = -\frac{1}{\epsilon_0 c^2}\left[\frac{\partial j_x}{\partial t} + c^2\frac{\partial(en_e)}{\partial x}\right], \tag{2.73}$$

where $\nabla_\perp^2$ is the transverse Laplacian, n_e is the electron volume density, and $j_x = ev_x n_e$ is the transverse current. Examining the relative magnitude of the two terms on the right hand side of equation 2.73, we can translate the relationship between the derivatives as,

$$c^2\frac{\partial(n_e e)}{\partial x} \overset{?}{\gg} \frac{\partial j_x}{\partial t} \Rightarrow \sigma_x \overset{?}{\gg} \lambda\gamma. \tag{2.74}$$

A priori we do not know if this inequality will be satisfied, so we employ our knowledge of how high-gain FELs operate to assist the analysis and only later justify these assumptions. In particular, σ_x should be comparable to the radiation beam size, which should itself be related to its Rayleigh length L_R and wavelength λ. Furthermore, we know the relationship between the FEL radiation wavelength λ and beam energy from equation 2.42, so we can convert this condition to,

$$L_R \overset{?}{\gg} \lambda_u. \tag{2.75}$$

The leap in reasoning now occurs and we assume that the Rayleigh length L_R is comparable to the gain length of the FEL. The gain length in an X-ray facility is of order meters, while typical undulator periods are in the cm range. Qualitatively, we may also find it difficult to believe that effective coupling between the electron beam and radiation could take place if the radiation diverged significantly after only one undulator period.

Having achieved this sleight of hand, we switch now to a one-dimensional picture and consider this equation in the frequency domain for the Fourier conjugate to E_x such that $\tilde{E}_x(z,t) = \tilde{E}_x(z)e^{i(k_L z - \omega_L t + \psi_0)}$,

$$\left[2ik_L \frac{d\tilde{E}_x(z)}{dz} + \frac{d^2\tilde{E}_x(z)}{dz^2}\right] e^{i(k_L z - \omega_L t + \psi_0)} = \mu_0 \frac{\partial \tilde{j}_x}{\partial t}, \tag{2.76}$$

where we now have the Fourier current component $\tilde{j}_x$, and the dispersion relation for light in vacuum has been used. Now, we assume that the function $\tilde{E}_x(z)$ changes slowly on the length scale of a single undulator period, an assumption known as the *slowly varying envelope approximation* (SVEA). This allows us to drop the second derivative term to arrive at,

$$\frac{d\tilde{E}_x(z)}{dz} = -\frac{i\mu_0}{2k_L} \frac{\partial \tilde{j}_x}{\partial t} e^{-i(k_L z - \omega_L t + \psi_0)}. \tag{2.77}$$

From the electron equations of motion 2.33-2.34, it follows that $\tilde{j}_x = \tilde{j}_z \frac{K\mathcal{J}}{\gamma} \cos(k_u z)$. We describe this current fluctuation as a constant portion which acquires a z-dependent contribution of the form of the phase (eq: 2.71) θ: $\tilde{j}_z = \tilde{j}_0 + \tilde{j}_1(z) e^{i[(k_L + k_u)z - \omega_L t]}$. Such a choice is motivated by the pendulum equations, although in effect we anticipate the microbunching effect by choosing this basis to expand into. Ultimately, this yields the field evolution equation,

$$\frac{d\tilde{E}_x}{dz} = -\frac{\mu_0 c K \mathcal{J}}{4\gamma}\tilde{j}_1. \tag{2.78}$$

The longitudinal space charge within the density modulated electron bunch also gives rise to an associated variation in longitudinal electric field via Maxwell's equations. Under the SVEA, the complex amplitude for the longitudinal electric field can thus be related to the current modulations as,

$$\tilde{E}_z(z) = -\frac{i}{\epsilon_0 k_L}\tilde{j}_1. \tag{2.79}$$

With some of this knowledge of the dynamic nature of the field, we can rewrite our pendulum equation for η to include not just the static field, but one which evolves in time as,

$$\frac{d\eta}{dz} = \frac{-e}{mc^2\gamma_r}\Re\left\{\left(\frac{K\mathcal{J}\tilde{E}_x}{2\gamma_r} - \frac{i}{\epsilon_0 k_L}\tilde{j}_1\right)e^{i[(k_L+k_u)z-\omega_L t]}\right\}, \tag{2.80}$$

where $\Re\{\cdots\}$ denotes the real part.

Up until this point we have discussed only a single electron, and we now move into the discrete case in which a beam is composed of many electrons. To obtain the longitudinal distribution coefficient $\tilde{j}_1$, we assume that the bunch is much longer than the radiation wavelength, and that it can be assumed periodic on this scale. Then, the longitudinal distribution function $S(\theta)$ can be written in terms of Fourier modes as,

$$S(\theta) = \frac{c_0}{2} + \Re\left\{\sum_{k=1}^{\infty} c_k e^{ik\theta}\right\}, \tag{2.81}$$

with complex coefficients c_i. We briefly note here that this description is inadequate to treat the shot noise which is essential to the SASE startup process. Using the Fourier trick, we can work out the first c_1 coefficient, $c_1 = \frac{1}{\pi}\sum_{n=1}^{N_e} e^{-i\theta_n}$, where N_e is the number of electrons in the bunch. This leads to the solution for $\tilde{j}_1$,

$$\tilde{j}_1 = j_0 \frac{2}{N_e} \sum_{n=1}^{N_e} e^{-i\theta_n}, \tag{2.82}$$

where j_0 is the (real) current density, and we see $\tilde{j}_1$ is clearly proportional to the bunching factor already discussed in equation 1.1. Thus, when considered as a system for N_e electrons, the equations 2.72, 2.80, 2.78, and 2.82 form a closed system of differential equations for the electron positions and field amplitude as a function of distance along the undulator z. Of course, as the number of electrons is quite large, such a system cannot be solved analytically and we must resort to smoothing the distribution of electrons to obtain elementary analytical results.

We describe the smoothed distribution by the function $F(\theta, \eta, z)$, such that $dn_e = n_e F(\theta, \eta, z) d\theta d\eta$. As before with the modulation j, we split this into an unmodulated and modulated portion,

$$F(\theta, \eta, z) = F_0(\eta) + \Re\left\{\tilde{F}_1(\eta, z) e^{i\theta}\right\}. \tag{2.83}$$

Assuming the range of η to be small, as is typical in a modern linear accelerator, we assume the distribution in η ranges from $-\delta$ to δ, then $\tilde{j}_1 = j_0 \int_{-\delta}^{\delta} \tilde{F}_1(\eta, z) d\eta$, allowing us to connect to the analytical description developed above.

The smoothed distribution function F, as it describes particles in a Hamiltonian system, must obey the Vlasov equation,

$$\frac{dF}{dz} + \frac{\partial F}{\partial \eta}\frac{d\eta}{dz} + \frac{\partial F}{\partial \theta}\frac{d\theta}{dz} = 0. \tag{2.84}$$

Combining this statement with equation 2.71 and 2.80 results in the (sufficient) requirement on F that,

$$\frac{\partial \tilde{F}_1}{\partial z} + i2k_u \eta \tilde{F}_1 - \frac{e}{mc^2\gamma_r}\frac{dF_0}{d\eta}\left[\frac{K\mathcal{J}}{2\gamma_r}\tilde{E}_x + \tilde{E}_z\right] = 0. \tag{2.85}$$

Assuming zero modulation at the entrance to the undulator, this equation has a general solution in integral form. Combining this solution with the relation between $\tilde{j}_1$ and $\tilde{E}_x$ (eq. 2.78) then yields the following integro-differential equation for the evolution of the electric field along the undulator,

$$\frac{d\tilde{E}_x}{dz} = ik_u \frac{\mu_0 K \mathcal{J} n_e e^2}{2m\gamma_r^2} \int_0^z dz' \left[\frac{K\mathcal{J}}{2\gamma_r} \tilde{E}_x + i \frac{4\gamma_r c}{\omega_L K \mathcal{J}} \frac{d\tilde{E}_x}{dz} \right] \left[\int_{-\delta}^{\delta} (z - z') e^{-2ik_u \eta (z-z')} F_0(\eta) \right] . \tag{2.86}$$

This equation, in principle, completely specifies the evolution of the transverse field $\tilde{E}_x$ along the undulator. In practice, however, we can simplify further to obtain analytical solutions. An obvious simplification is to assume a monoenergetic beam at $\eta = \eta_0$, which collapses the second integral. By applying two consecutive derivatives, one ultimately arrives at a pure differential equation for $\tilde{E}_x$,

$$\frac{d^3\tilde{E}_x}{dz^3} + 4ik_u\eta_0 \frac{d^2\tilde{E}_x}{dz^2} + \left(k_p^2 - 4k_u^2\eta_0^2\right) \frac{d\tilde{E}_x}{dz} - i\Gamma^3 \tilde{E}_x = 0, \tag{2.87}$$

with the definitions,

$$k_p = \sqrt{\frac{2k_u \mu_0 n_e e^2 c}{\gamma_r m \omega_L}}, \tag{2.88}$$

which is the plasma parameter, and is related to the relativistic plasma frequency, and,

$$\Gamma = \sqrt[3]{\frac{\mu_0 K^2 \mathcal{J}^2 e^2 k_u n_e}{4\gamma_r^3 m}}. \tag{2.89}$$

As a typical physicist, we attempt a solution by ansatz of the form $\tilde{E}_x(z) = Ae^{\mathfrak{K}z}$, which, in the $k_p \to 0$ limit, results in the restriction on $\mathfrak{K}$ of $\mathfrak{K}^3 = i\Gamma^3$.

This cubic equation yields three general solutions: one with negative real part which is exponentially damped, one pure imaginary which is oscillatory, and one with positive real part which grows exponentially. It is this third, exponentially growing mode which is the mode responsible for the remarkable amplification power of the FEL, and is given by $\mathfrak{K} = \left(i + \sqrt{3}\right)\Gamma/2$. This exponential growth implies a gain length, which we define as the *power* gain length (as opposed to the field gain length),

$$L_{g0} = \frac{1}{\sqrt{3}\Gamma} \equiv \frac{1}{2\sqrt{3}k_u\rho}, \tag{2.90}$$

where we have finally defined the quantity $\rho = \Gamma/2k_u$. This ρ (introduced in equation 1.2) is known as the Pierce parameter [10], or FEL parameter, and is the main quantity which defines the performance of the one-dimensional, high-gain FEL. As we have seen, it is clearly related to the exponential gain-length of the FEL. Were the parameter ρ to appear only in this place it would perhaps not receive a special name, however it controls various other parameters of the FEL performance as well.

The saturation power of the FEL is given approximately by [37],

$$P_{sat} \approx 1.6\rho P_{beam}, \tag{2.91}$$

while the undulator distance to this saturation is given by,

$$z_{sat} \approx \lambda_u/\rho, \tag{2.92}$$

where $P_{beam} = N_e\gamma mc^2$ is the total electron beam power. The fractional bandwidth of the radiation at saturation is also proportional to ρ. These are but some of the places in which the parameter ρ appears, and suffice it to say that in most cases it is advantageous to increase it (shorter gain length, more power).

The situation in which the FEL process is not seeded by monochromatic radiation, but instead by electron beam shot noise (SASE) is of particular interest. The assumptions that led to equation 2.87 are clearly incompatible with this, but without derivation we state some results from this noisy start-up process. The rms SASE bandwidth under some assumptions is found to be [43] [44],

$$\sigma_\omega = \sqrt{\frac{3\sqrt{3}\rho}{k_u z}}\omega_L, \tag{2.93}$$

and since at saturation $z_{sat} \sim 1/\rho$, the saturated FEL bandwidth $\sigma_\omega \sim \rho$. This

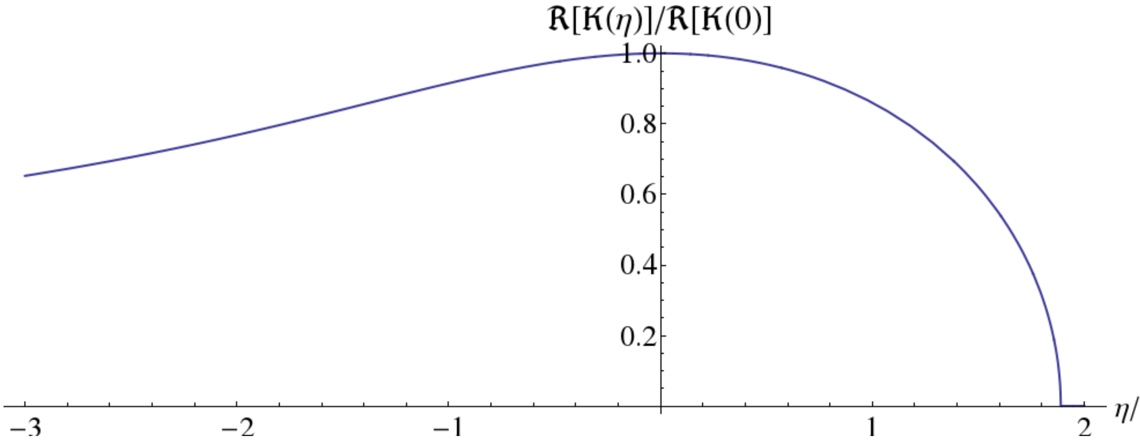


Figure 2.14: The FEL detuning curve as a function of off-resonance energy η. The real part of the root vanishes for $\eta > 1.88\rho$.

bandwidth equivalently describes a coherence time,

$$t_c \approx \frac{\sqrt{\pi}}{\sigma_\omega}. \tag{2.94}$$

In general, however, the electron beam length T will be much longer than this coherence time for X-ray FELs, and therefore the resulting radiation pulse will consist of many such uncorrelated pulses. The resulting spectrum, as shown originally in figure 1.1, is a chaotic spectrum made up $\approx T/t_c$ distinct modes, each with bandwidth $\sim 1/T$, all within an overall spectral envelope defined by ρ. It is this broad SASE spectrum that beam based seeding methods, such as Echo-Enabled Harmonic Generation demonstrated in chapter 4, are designed to improve on.

One final bit of FEL theory which will become relevant is the effect of off-resonance particles. In the 1D picture, the effect of particles with $\eta \neq \eta_0$ can easily be handled and affects the roots of the cubic polynomial which define the growing, damped, and oscillatory FEL modes. The root with positive real part which leads to exponential growth becomes a function of the detuning off-resonance parameter η as,

$$\frac{\mathfrak{K}}{\Gamma} = \frac{1}{6}\left(u\left(\frac{\eta}{\rho}\right) - 4\frac{\left(\frac{\eta}{\rho}\right)}{u\left(\frac{\eta}{\rho}\right)} - 4i\left(\frac{\eta}{\rho}\right)\right), \tag{2.95}$$

where the function $u(x)$ is given by,

$$u(x) = \sqrt[3]{108i - 8ix^3 + 12\sqrt{12x^3 - 81}}. \tag{2.96}$$

A plot of the real part of this root $\mathfrak{K}$ as a function of η is shown in figure 2.14. We observe that the so-called detuning curve is asymmetric – the FEL gain drops precipitously for $\eta > 0$, while it falls off slowly for $\eta < 0$. The implication is that higher energy particles (which emit in the blue) will not radiate strongly, while there may be a long tail of lower energy particles (which emit in the red) which remain. In this sense, the FEL can be said to act as a low-pass filter, an observation which will become important with regards to the FEL response to the microbunching instability (see section 5.2).

Finally, we note that the derivation provided above has been the simplest possible with the goal of arriving at the quantity ρ. A great many other physical effects need be included to accurately model existing FELs. Many of these can be analytically treated to some extent, but we note here that an impressive number can be accurately modeled by the fitting formulas of Ming Xie [45]. In these formulas, 1D parameters such as the gain length become altered as,

$$L_{g,3D} \approx L_{g0} \left(1 + \Lambda\right), \tag{2.97}$$

where Λ is a polynomial function which incorporates the transverse size of the beam, its angular spread (and emittance), and non-zero energy deviation. Due to the success of this fitting formula, the 1D theory discussed above remains relevant even in the presence of 3D effects, with only minor modifications.

Returning to the 1D theory, the value of ρ depends on various aspects of the electron beam and undulator system. It is proportional to the peak current of the electron beam and electron beam size (through the factor n_e), and also the coupling to the undulator through the factor $K\mathcal{J}$.

Much of the optimization in the construction of an FEL involves optimizing the parameter ρ. Since the resonance wavelength 2.42 is determined by both the undulator parameters and beam energy, there is immediately a trade-off between these parameters for a given target radiation wavelength. Besides this tradeoff, it is advantageous to obtain the highest density beam possible in order to maximize ρ. This push towards high density electron beams causes the appearance of otherwise unimportant physical effects, the understanding and mitigation of which is a major topic of this work.

2.3 FEL Limiting Effects

The essential physics of the X-ray FEL just discussed can be thought of as embodied in the parameter ρ and its maximization. Through both the peak current and transverse beam size, ρ is clearly proportional to the brightness of the of the electron bunch, and thus it is advantageous to create as bright a beam as possible. However, there are several physical effects which become important when such bright beams are considered. These effects must be considered when designing various parts of the X-ray FEL facility, and if ignored can ruin an otherwise well-prepared, bright electron beam.

2.3.1 Coherent Synchrotron Radiation

The question of how tightly electrons could be bunched in a ring has been around almost as long as electromagnetism. This question chiefly concerns the synchrotron radiation produced by the electrons in the bending sections interacting with other electrons to heat the beam. In the classical regime, this process can be described as a wakefield effect and is known as Coherent Synchrotron Radiation (CSR). In the quantum regime, by contrast, the response of individual electrons to the quantum kick $\hbar\omega$ is known as Incoherent Synchrotron Radiation (ISR) – a qualitatively different effect (see equation 4.1). We are concerned here with the classical CSR effect. Many of the main, one dimensional results were derived already in the early 20th century by Schott [46], refined by Schwinger in the middle of the century [47], and only revisited

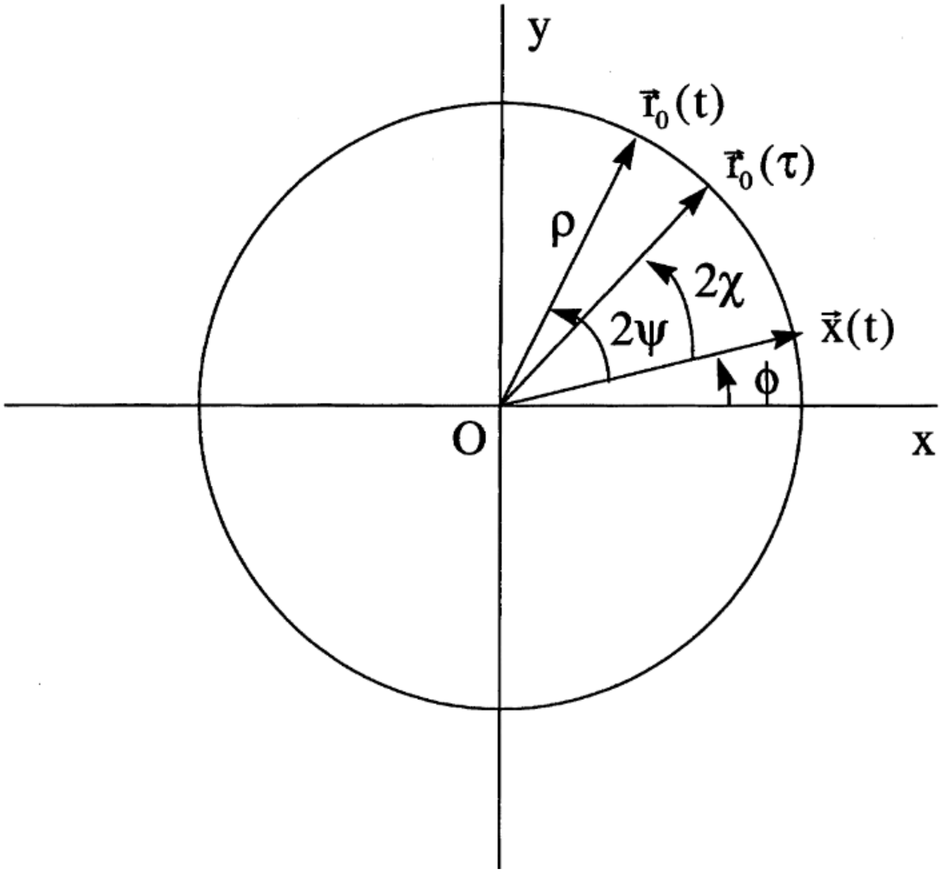


Figure 2.15: The 1D CSR Coordinate system. Radiation emitted the retarded position $\vec{r}_0(\tau)$ at retarded time τ is observed at the point $\vec{x}(t)$ at later time t. Note that $\vec{x}$ is *ahead* of the emission point. Figure from [48].

much later when their importance to the construction of high brightness electron machines finally became relevant [48] [49].

In the standard one-dimensional picture, we have a zero-emittance, monochromatic electron beam at speed β moving on an arc of radius ρ. The associated coordinate system is shown in figure 2.15. In this coordinate system, radiation emitted at the retarded position $\vec{r}_0(\tau)$ at retarded time τ is received at a later time and position $\vec{x}(t)$. In this picture, $\vec{x}(t)$ is *ahead* of the emitting particle.

The physical picture is that electrons in the tail of the bunch produce forward-beamed radiation which catches up with the head electrons and interacts with them.

This is possible because while the electrons are confined to their circular arc trajectory, the radiation travels in a straight line between the emission and observation points. Of course, this implicitly assumes such a straight-line path is always available, which may not be true when a finite-size beam pipe is considered although we ignore this complication. We further assume that the electrons have been on this trajectory for long enough that, during the time it takes for this catch-up, both particles remain on the circular trajectories. This is known as the 'steady-state' approximation, and neglects the field variations at the beginning and ends of the bend magnet.

Given this geometry, we can compute the electromagnetic field from the Liénard-Wiechert (L-W) expressions. The complication comes in treating the retarded time condition, which can be expressed in terms of the angular variables from figure 2.15 as,

$$\psi = \chi + \beta \sin \chi. \tag{2.98}$$

Due to the circular geometry, this provides a transcendental relationship between the various coordinates which cannot be solved in general. Nevertheless, we can treat some angles as small and introduce the auxillary variables $\xi \equiv \pi - \psi$, $\alpha \equiv \pi - \chi$. For short-ranged interactions, the angles ξ and α are small, and thus the sine function in equation 2.98 can be expanded. Furthermore, in this same approximation, we typically assume a relativistic electron beam and trade β for $\gamma \gg 1$, which produces the approximate retarded equality,

$$\alpha^3 + \frac{3}{\gamma^2}\alpha - 6\xi = 0. \tag{2.99}$$

This expanded equation produces analytic solutions in the parameter $\mu = 3\gamma^3\xi = 3\gamma^3 s/(2\rho)$ for an arclength separation s as,

$$\alpha = \frac{1}{\gamma}\left(\Omega^{1/3} - \Omega^{-1/3}\right), \tag{2.100}$$

with,

$$\Omega = \mu + \sqrt{\mu^2 + 1}. \tag{2.101}$$

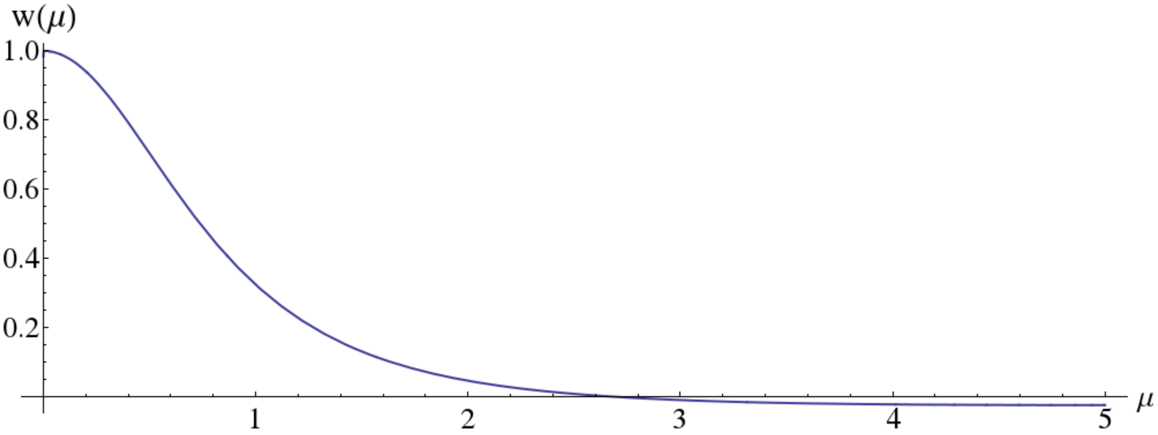


Figure 2.16: The 1D CSR wake function w as a function of the scaled arc distance μ.

Higher order expansions are possible to capture longer-range effects, but the standard near-field result is well captured by the third order equation. When inserting this result into the L-W equations, care must be taken with the singular Coulomb part of this expression. Fortunately, this can be clearly separated from the non-singular term due to synchrotron radiation. The tangential electric field (along the trajectory) can then be expressed in the form of a wakefield as,

$$E_\psi = -\frac{4}{3}\frac{e\gamma^4}{\rho^2}\frac{d\nu(\mu)}{d\mu}, \tag{2.102}$$

where,

$$\nu(\mu) = \frac{9}{16}\left[\frac{-2}{\mu} + \frac{1}{\mu\sqrt{\mu^2+1}}\left(\Omega^{1/3} + \Omega^{-1/3}\right) + \frac{2}{\sqrt{\mu^2+1}}\left(\Omega^{2/3} - \Omega^{-2/3}\right)\right]. \tag{2.103}$$

The function $w(\mu) \equiv \frac{d\nu(\mu)}{d\mu}$ is therefore called the free-space CSR wake function and is plotted in figure 2.16. Note that the 1D CSR wake is a factor of $1/\gamma^4$ smaller for the region $\mu < 0$ than for the $\mu > 0$ region, and is therefore identically zero to the order we have computed equation 2.99. This means that in the 1D ultrarelativistic theory, the wakefield extends only in front of the emitting particle. By expanding the complicated solution of equation 2.103, we can explore simple limiting cases,

$$w(\mu) = \begin{cases} 1 - \frac{14}{9}\mu^2 & , \mu^2 \ll 1 \\ -\frac{3}{4\sqrt[3]{2}\mu^{4/3}} & , \mu^2 \gg 1 \end{cases}. \tag{2.104}$$

The wakefield therefore has a long tail for large μ which slowly regresses to zero. Note that we can immediately state that since w is a total derivative and vanishes at both extremes,

$$\int_{-\infty}^{\infty} w(\mu) d\mu = 0. \tag{2.105}$$

This is actually not quite correct, as the function $w(\mu)$ is discontinuous at $\mu = 0$, where $w(\mu = 0^+) = 1$ while $w(\mu = 0^-) = 0$. The reasonable definition of $w(0) = 1/2$ then gives the total energy loss of the bunch in agreement with the total power radiated which can be computed by e.g. the Larmor formula.

Up until this point, the discussion concerns only a single electron emitting radiation, and therefore despite equation 2.102 being referred to as the coherent synchrotron radiation wake, it it is really a single particle effect. The coherence comes from the fact that the wakefield shape is a well-defined function of distance along the bunch, as opposed to a random radiation process such as ISR. Note that if the electron bunch has a length which is short when compared to the scale length μ, indeed there will be significant coherent radiation. In the more likely case, however, that the bunch is long compared to μ, it will be the long portion with $w \sim \mu^{-4/3}$ which produces the bulk of the interaction.

The result for a Gaussian bunch with rms length σ can be found by simply integrating this profile with the wakefunction. The result is the wakefield shown in figure 2.17. This is the standard 1D wakefield from CSR theory, and is in most cases an excellent approximation to the true contribution due to radiation in bend magnets. Even within the context of this 1D picture, however, there is room for significant improvements over the results quoted here. These improvements generally treat the retarded condition in a more sensible way, and can be focused on forms more efficient for computation [50] or including other physical effects such as radiation shielding and bunch compression in the magnet [51]. Novel corrections to this simplistic 1D

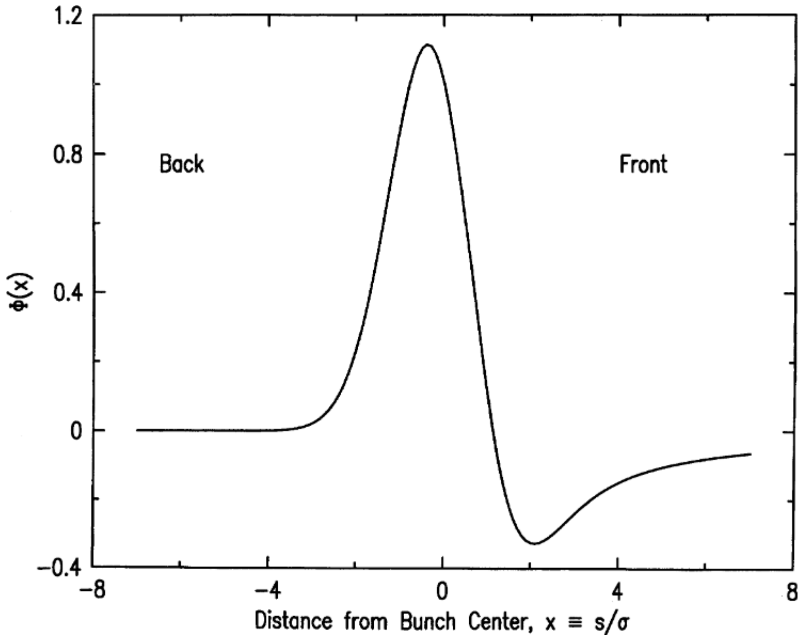


Figure 2.17: The Gaussian bunch wakefield ϕ as a function of $x = s/\sigma$. Figure from [48].

picture will also be explored in chapter 6.

Having described the mechanism of CSR energy spread generation, the next question is how it can degrade electron beam parameters. As the electron bunch traverses a 4-bend chicane and loses energy due to CSR, this energy loss couples to the transverse dispersion functions through terms like R_{16}. Since this occurs inside the chicane, it is not enough to simply ensure that the chicane as a whole has zero R_{16}. Furthermore, since in the case of a bunch compressing chicane the bunch length changes inside the chicane, in general the CSR contribution from each individual magnet will not be identical. Therefore, CSR-based emittance growth strongly constrains the design of bunch compressing chicanes [52] [53] [54] [55].

The emittance growth contribution due to this energy spread growth in a chicane can be computed in a simple manner if we assume it is small compared to the initial emittance as,

$$\Delta\epsilon \approx \frac{1}{2}\left(\frac{\sigma_E}{E}\right)^2 \langle\mathcal{H}\rangle, \tag{2.106}$$

where the (imaginatively called) curly-H function $\mathcal{H}$ is defined in terms of the Twiss parameters as,

$$\mathcal{H} = \gamma\eta^2 + 2\alpha\eta\eta' + \beta\eta'^2. \tag{2.107}$$

This formula, while only a small-growth approximation, provides a useful guideline to designing bend systems when appreciable energy growth is expected. The magnitude of emittance growth is proportional to $\mathcal{H}$ averaged over the dispersive section and proportional to the square of the induced energy spread.

Note however, that however bad these mismatches and dispersion related emittance growths may be, they are always in principle reversible. Since the wakefunction is a well defined function of the longitudinal position within the bunch, as in figure 2.16, no 'information' is ultimately lost in this process and a careful phase space manipulation should enable a full recovery of the initial emittance. This stands in stark contrast to the quantum ISR effect, or the particle granularity correction we discuss in chapter 6 which are stochastic in nature and cannot be removed.

2.3.2 Microbunching Instability

The so-called microbunching instability (UBI) was first discovered during the development of the LCLS in the early 2000s [56] [57] [58] [59]. At its core, the microbunching instability is a self-reinforcing energy and density modulation on the electron beam. Initially a small density modulation is present on the electron beam, which is typically assumed to be due to simple shot noise. This density modulation, through the beam's own self-fields, lead to a beam energy modulation. As this energy modulation passes through a dispersive region with nonzero R_{56}, the energy modulation is (partially) converted into a density modulation, which reinforces the instability. The dispersive section of R_{56} is often the large bunch compressors employed to increase the peak current of the bunch, but can also be a long drift section. The overall amplification in this cycle is referred to as the gain of the instability.

The mechanism of the UBI is fundamentally quite similar to the HGHG (or klystron) mechanism discussed in section 2.4.1. However, rather than generating a well-defined bunching factor at a desirable wavelength, the UBI generally creates

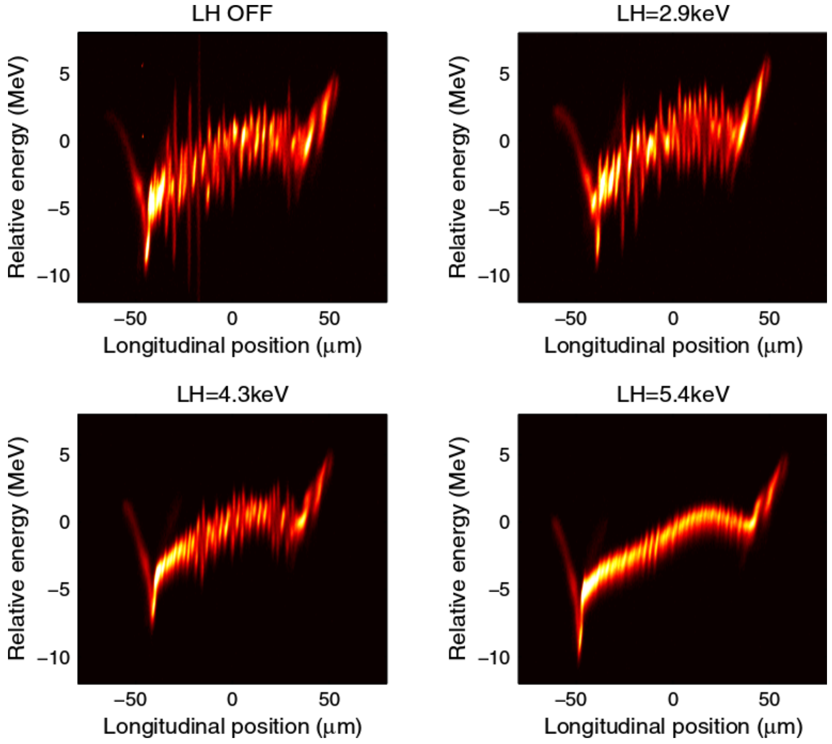


Figure 2.18: The microbunching instability directly imaged at LCLS. The LH values correspond to the mean heating from the laser heater, which is seen to suppress the UBI. Figure from [60].

a broad-band modulation which effectively destroys the slice properties of the beam. The effect of the UBI on the electron beam at LCLS (and its partial control) is shown in figure 2.18. Therefore, understanding the origins of this instability and how to mitigate it is essential to obtaining a high quality FEL beam.

The self-fields responsible for the amplification of the UBI can come in several forms. One possibility is that the CSR fields in the bunch compressors themselves can provide the necessary gain to amplify the instability. Another important effect is the wakefields present in the linac structures. Since FEL facilities typically consist of many accelerating structures periodically separated by bunch compression chicanes, the wakefields in the structures can provide sufficient impedence to generate the instability. Finally, the simple free-space longitudinal space charge provides another impedence source [61]. Ideally, all three should be considered when taking the UBI

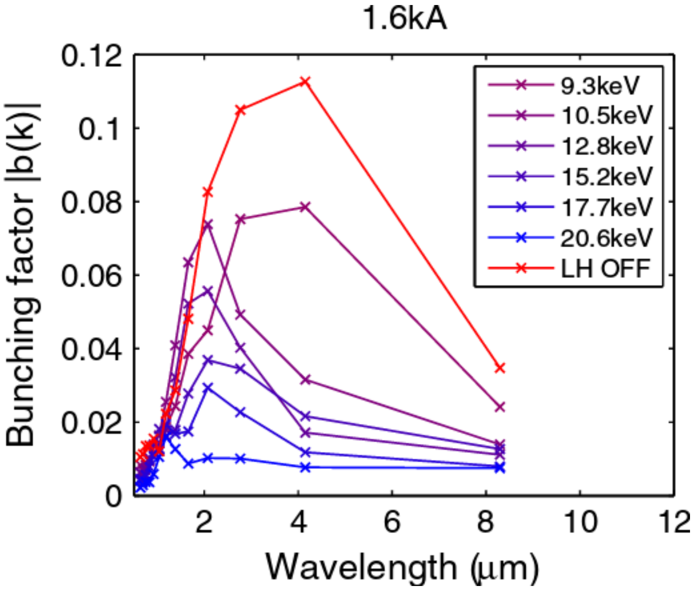


Figure 2.19: The microbunching gain (bunching) as a function of wavelength at LCLS for a variety of laser heater settings. Figure from [60].

into account.

A single stage of the microbunching instability produces a gain (defined as the ratio of final to initial bunching) as a function of wavenumber with the approximate form [62],

$$G(k) \approx \frac{I_0 k}{C\gamma I_A}\left| R_{56}\int_0^l ds \frac{4\pi Z(k/C,s)}{Z_0}\right| \exp\left(\frac{-k^2 R_{56}^2 \delta^2}{2}\right), \tag{2.108}$$

where $I_A \approx 17$ kA is the Alfvén current, C is the compression factor of the stage, l its length, δ the slice energy spread, Z is the impedance of the stage, and $Z_0 \approx 377\Omega$ is the free space impedance. For low frequencies, the impedance can be written approximately as,

$$Z(k) \approx \frac{iZ_0 k}{4\pi\gamma^2 C}, \tag{2.109}$$

where we have dropped a term logarithmic in $1/k$. The general form of the gain function is therefore $G(k) \sim k^2 e^{-k^2}$, where the gain grows quadratically for small k

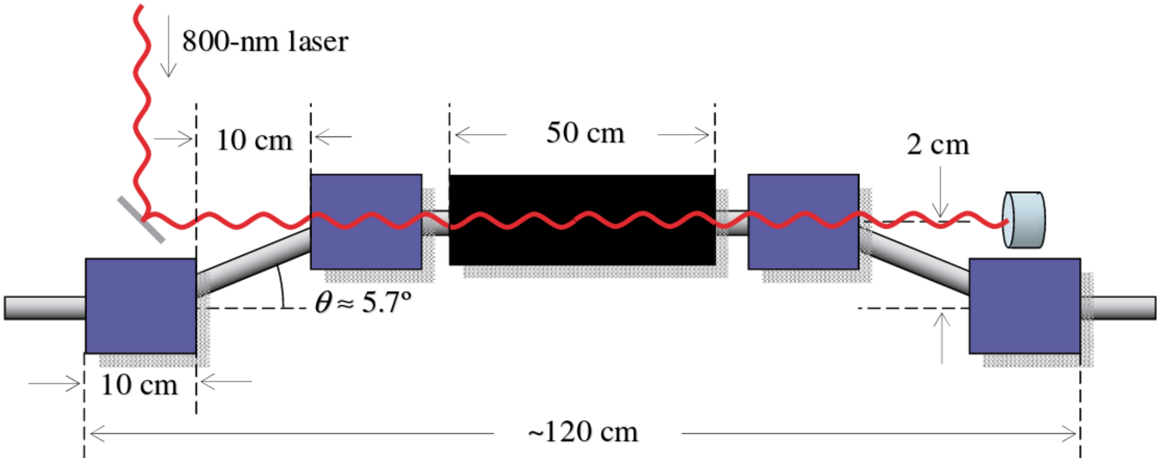


Figure 2.20: A diagramatic representation of the LCLS laser heater. Figure from [62].

until $k_c \approx 1/(R_{56}\delta)$, after which the gain is exponentially suppressed. This approximate form has been experimentally measured, and is shown in figure 2.19. When considering the response of seeding systems (e.g. section 5.2)to the UBI, we will use this general form to model its modulations.

The approximate expression 2.108 also provides a hint for how to mitigate the UBI: Increase the uncorrelated energy spread δ. This is precisely what is done in order to suppress the UBI, and it is accomplished by means of a laser heater (LH) [62]. A small modulator is placed in the center of a 4-dipole magnetic chicane, and a resonant laser beam is allowed to modulate the electron beam in this high dispersion region. As the electron beam then enters the latter half of the chicane, this energy modulation is smeared out, effectively increasing the slice energy spread. A diagram of this laser heater is shown in figure 2.20.

The effectiveness of this procedure is shown in figures 2.18 and 2.19, wherein it is seen that larger energy modulations in the laser heater suppress the UBI. In practice, it is a balancing act between setting the LH high enough to suppress the UBI to the level required for a given experiment, but also setting it low enough such that the induced energy spread does not make the FEL gain too small. This is only one of the myriad of tradeoffs necessary in designing an entire FEL facility.

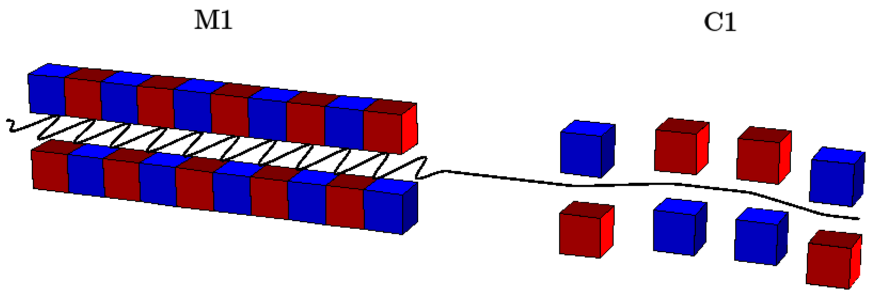


Figure 2.21: A diagramatic illustration of the HGHG setup. The electrons first enter a modulating undulator (M1), followed by a four dipole chicane (C1).

2.4 Beam-Based FEL Seeding Techniques

There are at present a large number of beam-based seeding techniques [40], but this work is concerned with only two of the most basic: High-Gain Harmonic Generation (HGHG) and Echo-Enabled Harmonic Generation (EEHG).

2.4.1 High-Gain Harmonic Generation

HGHG is perhaps the simplest beam-based seeding technique, and also has the longest history. The fundamental dynamics of the system were first understood as the optical klystron in the 1980s [63] [64]. In such a scheme, the beam is used merely for the generation of coherent radiation, not as a seed for an FEL. In this guise, it is often referred to as Coherent Harmonic Generation (CHG). However, it was soon realized that an identical setup could be used to seed the FEL process [65] [66]. Since the fundamental physics and setup is identical in either case, we will consistently refer to this setup as “HGHG”, even if there is no significant gain in the system. This system relies on the two bread-and-butter building blocks that we have previously introduced: the modulating undulator (modulator) and the 4-dipole chicane and is shown in figure 2.21. The electrons first traverse a modulating undulator, copropagating with a resonant laser beam, and is then dispersed by a four dipole chicane. Note, it is not strictly necessary to have a chicane, as any R_{56} will do.

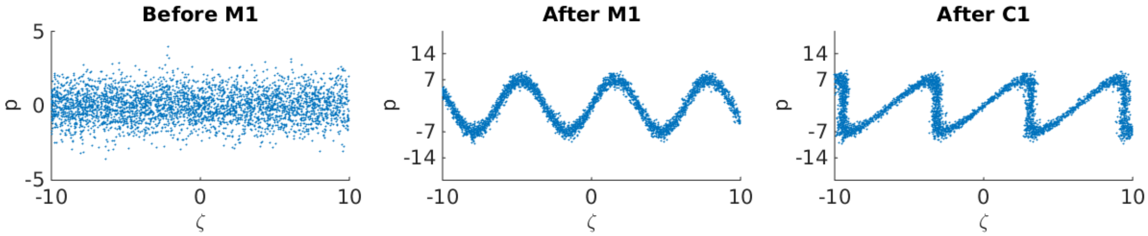


Figure 2.22: The electron phase space during HGHG. This simulation has $A = 7$ and B optimized for bunching at $n = 7$.

To analyze this system, consider an electron beam of total energy E_0 and a slice energy spread of σ_E. Then, define the energy deviation $p = (E - E_0)/\sigma_E$. We assume that the modulation ΔE attained from the modulator is perfectly sinusoidal, and uniform for all electrons. In practice, this corresponds to a modulating laser pulse which is both transversely and longitudinally much larger than the electron beam. We can therefore define the phase relative to the laser of wavenumber k_L as $\zeta = k_L s = k_L(z - \bar{\beta}ct)$. Therefore, after M1, we have,

$$p_1 = p_0 + A\sin(\zeta), \tag{2.110}$$

where $A = \Delta E/\sigma_E$ and the modulation amplitude ΔE could be found for a given undulator from e.g. eq. 2.65. In the chicane, no changes to p occur (assuming no radiation), but the particles are dispersed according to their momenta, changing ζ. In this coordinate, we have,

$$\zeta_1 = \zeta_0 + Bp_1, \tag{2.111}$$

where $B = R_{56}k_L\sigma_E/E_0$ is the scaled dispersive strength of the chicane. The HGHG setup is therefore characterized by only two dimensionless parameters: A and B. The phase space dynamics from these two equations are illustrated in figure 2.22 for the individual electrons in a bunch. The fundamental dynamics are clear: The modulator M1 creates a sinusoidal modulation in energy which is turned into a density modulation by means of the chicane C1. Half of the beam is decompressed, and half of it is compressed into density spikes. Since the bunching factor is a fourier transform, it is the sharpness of the density peaks after C1 which determine the high harmonic content for the HGHG scheme. The sinusoid does not fold over perfectly under dispersion, and it is natural to ask how one might perform better. The essential task then is to create as linear as possible a chirped beam section, which will optimize compression. This can be accomplished through cascaded modulator chicane sections with normal laser pulses [67], or through direct use of sawtooth laser forms [68]. Nevertheless, we do not pursue these options and continue with the analysis of the vanilla HGHG configuration.

The bunching spectrum generated by this configuration is found by computing [69],

$$b(h) = \left| \int f(p_0, \zeta_0) e^{-ih\zeta_1(\zeta_0, p_0)} d\zeta_0 dp_0 \right|, \tag{2.112}$$

where $h = k/k_L$ is the harmonic number, and $f(p, \zeta)$ is the original distribution prior to the modulator. Typically, the initial distribution is Gaussian in energy and ζ, so that we have $f(p,\zeta) = \frac{1}{2\pi\sigma_\zeta} e^{-p^2/2} e^{-\zeta^2/(2\sigma_\zeta^2)}$.

For the HGHG scheme, we have,

$$\zeta_1(\zeta_0, p_0) = \zeta_0 + Bp_1 = \zeta_0 + B\left(p_0 + A\sin(\zeta)\right). \tag{2.113}$$

In order to facilitate the computation of equation 2.112, the exponentiated sine can be expanded into Bessel functions (similar to 2.64),

$$e^{ia\sin\zeta} = \sum_{n=-\infty}^{\infty} e^{in\zeta} J_n(a). \tag{2.114}$$

We can now compute for the HGHG setup,

$$b(h) = \left| \sum_{n=-\infty}^{\infty} e^{-\frac{1}{2}\left(B^2h^2 + (h-n)^2\sigma_\zeta^2\right)} J_n(-hAB) \right|. \tag{2.115}$$

We see from this formula that we have peaks in the bunching spectrum centered around the integer values $h = n$, with width determined by the bunch length σ_ζ. Compared to the laser wavelength σ_ζ is typically quite large, and in this limit the exponential collapses to $\delta(h-n)$, and we have the bunching only at integer values n,

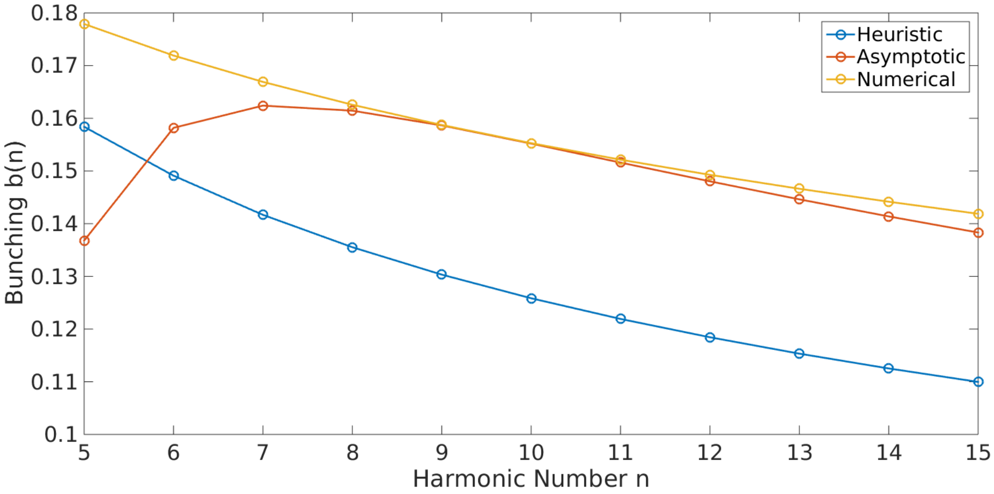


Figure 2.23: The optimized bunching factor for various HGHG harmonics at fixed $A = n$. The optimal B chosen is either found numerically, through the asymptotic expression 2.117, or from the heuristic $B = 1/A$.

$$b_n = \left| e^{-\frac{1}{2}(B^2 n^2)} J_n(-nAB) \right| . \tag{2.116}$$

We immediately notice the exponential suppression in front, which requires $\frac{1}{B} \gtrsim n$ to generate significant bunching. However, the Bessel function also has its peak when its argument is approximately equal to its order, suggesting $B \sim 1/A$. This provides a heuristic optimization to generate harmonics at the harmonic n: A should be around n, and $B \sim 1/A$. The exact optimization is transcendental due to the Bessel function. We will fix $A = n$, for the purpose of optimization at a given harmonic n. Then, we can analyze in the limit that $A = n \gg 1$, that the optimal B is,

$$B_{\text{opt,asmptotic}} = \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{n}}\right) . \tag{2.117}$$

A comparison between these optimization procedures and the true numerical optimum is shown in figure 2.23. We see that the asymptotic optimum agrees quite well above $n \approx 8$, while the heuristic guess $B = 1/A$ consistently underperforms.

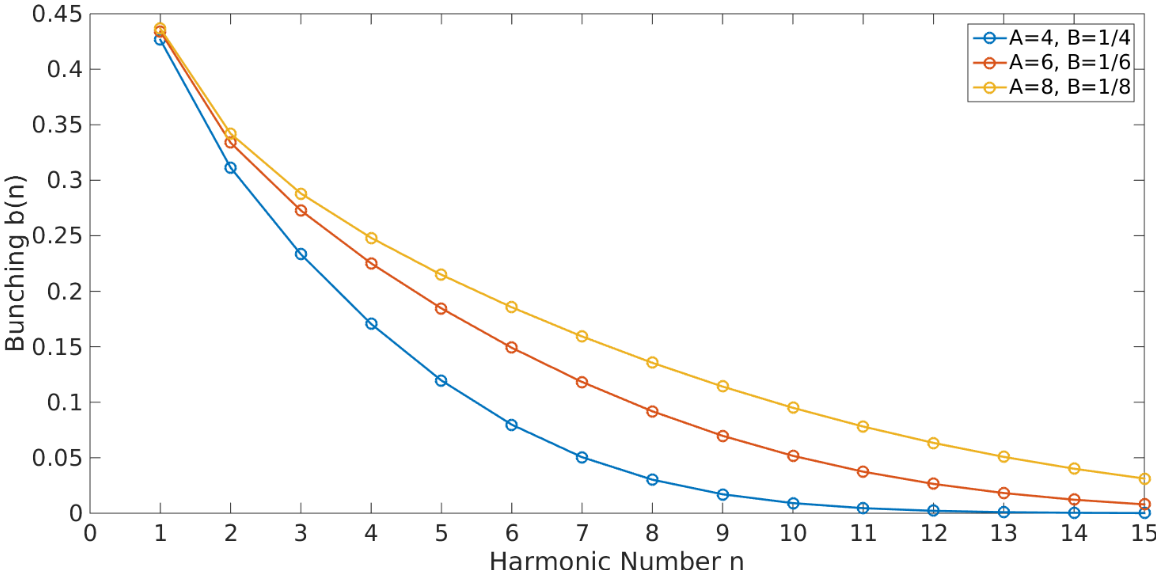


Figure 2.24: The HGHG bunching factor as across many harmonics for three separate configurations. The settings with higher A have uniformally higher bunching, and the bunching factor drops off rapidly at high harmonics.

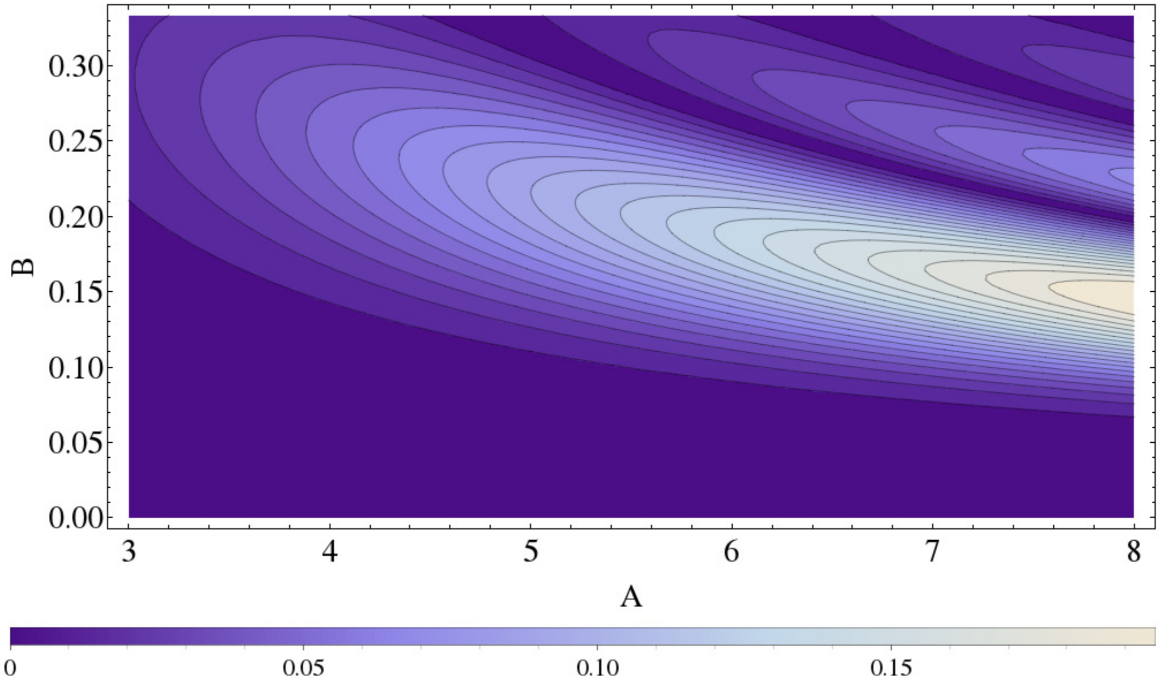


Figure 2.25: The HGHG bunching factor at $n = 7$ as a function of A,B.

Exploring the full A, B parameter space produces contours of bunching, shown in figure 2.25 for $n = 7$. We see that indeed it is optimal to push A as high as possible to obtain the greatest harmonic bunching. Furthermore we can see, for a given A, there are auxillary local maxima at higher values of B due to the Bessel function behavior. These outer islands are suppressed by the exponential factor, but nevertheless these provide an important clue for things to come with the EEHG scheme. It is also of note that one is able to slightly compensate for a lower value of A by increasing B slightly, indicated by the slope of the bunching region. This is of course the statement that in order to 'stand up' the density spikes in phase space at lower A, one must use a stronger chicane.

Moving away from the optimal bunching at a single target harmonic, we ask what is the bunching across many harmonics for a single HGHG setting? A plot for three different configurations is shown in figure 2.24. Of note is the fact that the configuration with higher A has uniformally higher bunching than that with lower A. Furthermore, the bunching decays rapidly at high harmonic number. This is the essential scaling of the HGHG upconversion process: In order to reach high harmonics, one must provide correspondingly large energy modulation relative to σ_E. As we have seen, the energy spread is a critical factor for FEL performance, so in practice the harmonic upconversion is typically limited to $n \lesssim 10$ [40]. This limitation may be extended through the use of cold electron beams, but the longitudinal compression required to achieve high peak currents, and thus an appreciable ρ, necessarily drives the slice energy spread to larger values.

Before moving on to EEHG, it is perhaps important to note that the HGHG seeding scheme has been in successful use for some time at the FERMI facility. This HGHG setup is able to reach down to the 15th harmonic in a single stage [70], or the 60th harmonic in two stages [71] from a 266nm laser pulse. Therefore, as HGHG is a proven technique for beam based seeding down into the soft-X ray, the challenge of a new seeding mode such as EEHG is to either extend down to lower harmonics or improve performance over the same spectral range.

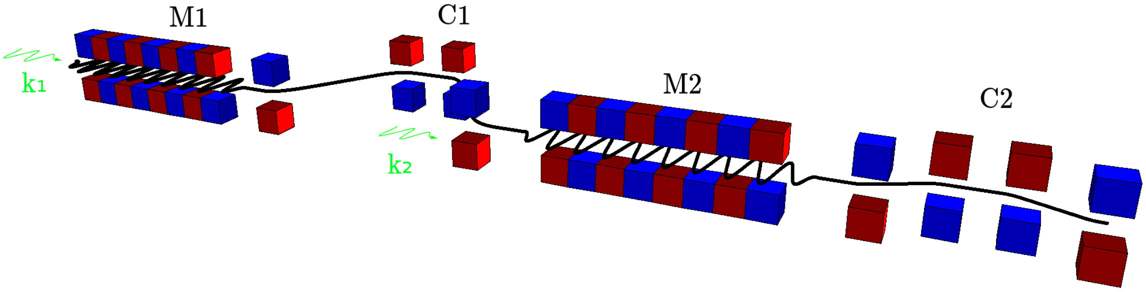


Figure 2.26: A diagrammatic illustration of the EEHG setup. The electrons first enter a modulating undulator (M1) with resonant wavenumber k_1, followed by a four dipole chicane (C1), then another modulator (M2) with resonant wavenumber k_2, and a final chicane (C2).

2.4.2 Echo-Enabled Harmonic Generation

The Echo-Enabled Harmonic Generation scheme was only relatively recently discovered [72] [69], although the more general echo effect has been known in many physical contexts for a long time [73] [74] [75]. The general principle involves a system which is perturbed, and then non-dissipatively damped or mixed up. After some time, a second perturbation is incident on the system and the same non-dissipative mechanism is again engaged. Then, at some future time the initial perturbation signal appears, seemingly out of nowhere, as an 'echo' of the past. The essential theme of the echo effect is that the non-dissipative mixing of the system preserves phase information, which is allowed to re-cohere at a later time as an echo.

Regarding EEHG, this effect is obtained by using a two modulator and two chicane system, as shown in figure 2.26. At first glance, it looks similar to back-to-back HGHG setups. However, the first chicane is significantly larger (in terms of dispersive strength R_{56}) than the second chicane. Furthermore, the two modulators M1 and M2 may be resonant at different wavenumbers k_1 and k_2, potentially necessitating different input lasers. As with the HGHG case, we analyze the electron bunch in the scaled coordinates p and ζ, where $\zeta = k_1 s$, since $k_1 \neq k_2$ in general, and define the ratio $\kappa = k_2/k_1$. The series of transformations in terms of normalized modulation

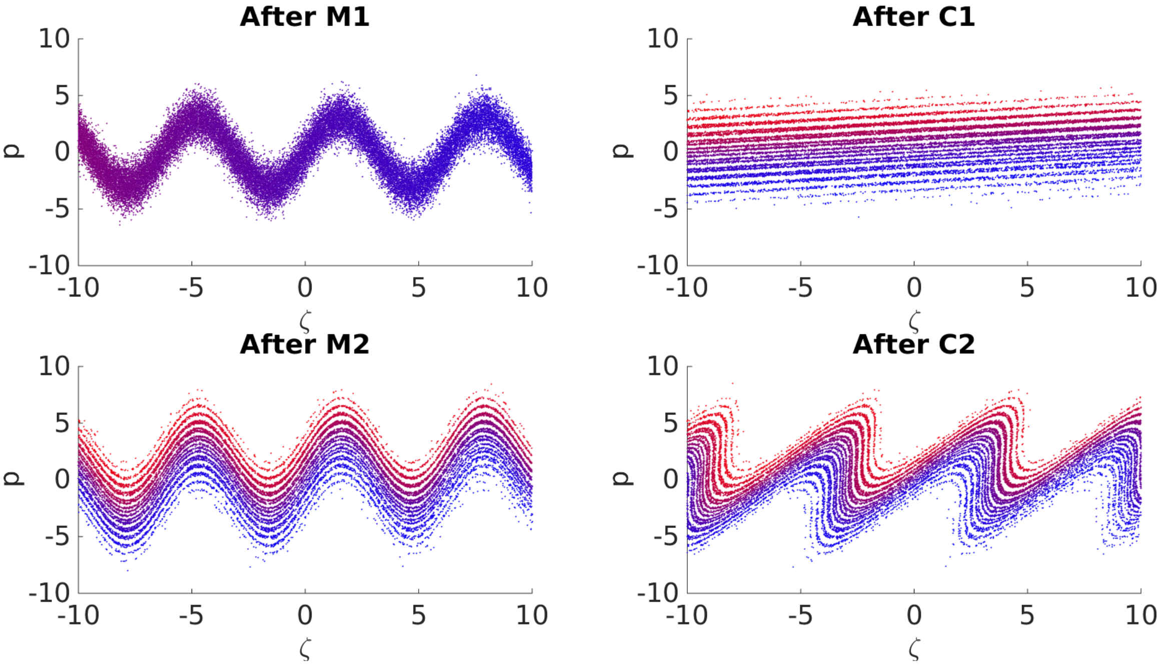


Figure 2.27: The electron phase space during EEHG. This simulation has $A_1 = A_2 = 3$ and $B_1 = 8.94, B_2 = 0.37$, optimized for bunching at $n = -1, m = 26$. Electrons initially from $-\zeta$ are colored red, while those from $+\zeta$ are colored blue.

amplitudes $A_{1,2} = \Delta E_{1,2}/\sigma_E$ is then,

$$p_1 = p_0 + A_1 \sin(\zeta_0), \tag{2.118}$$

$$\zeta_1 = \zeta_0 + B_1 p_1, \tag{2.119}$$

$$p_2 = p_1 + A_2 \sin(\kappa\zeta_1), \tag{2.120}$$

$$\zeta_2 = \zeta_1 + B_2 p_2. \tag{2.121}$$

One subtlety is that the definition of the B factors have changed slightly, and now $B_1 = R_{56,1} k_1 \sigma_E / E_0$ and $B_2 = R_{56,2} k_1 \sigma_E / E_0$. Formally, the B quantities could be positive, negative, or have differing signs according to their dispersive strengths.

However, as we are interested in a practical setup realized with two chicanes, they will have the same (positive) sign in all that follows.

The phase space evolution of individual electrons through these transformations is shown in figure 2.27. The effect of the large first chicane is to striate the modulated beam into a series of very thin (in terms of slice energy spread) beamlets. We see that during this process the electrons migrate over many laser periods of modulation, indicated by the initially far separated blue and red electrons ending up at the same ζ position after the first dispersion. Once these beamlets are set up, the second stage is essentially HGHG in which an energy modulation is again added and then converted into a density modulation. In contrast to HGHG however, with the multitude of beamlets the possibility for harmonic content is much greater. One can view this heuristically as due to the beamlets having a smaller slice energy spread than the total beam.

To compute the bunching from such a configuration, we proceed as before, by equation 2.112, except now we must compute $\zeta_2(\zeta_0, p_0)$,

$$\zeta_2 = \zeta_0 + B_1\left[p_0 + A_1\sin(\zeta_0)\right] + B_2\left[p_0 + A_1\sin(\zeta_0) + A_2\sin\left\{\kappa\zeta_0 + \kappa B_1\left[p_0 + A_1\sin(\zeta_0)\right]\right\}\right]. \tag{2.122}$$

As before, while computing the bunching factor we employ the Bessel identity, however twice this time. We assume the same initial Gaussian beam, and find,

$$b(h) = \left|\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} e^{-\frac{1}{2}(h\mathfrak{B}-m\kappa B_1)^2} e^{-\frac{1}{2}(h-n-m\kappa)^2\sigma_\zeta^2} J_n(-h\mathfrak{B}A_1 + m\kappa B_1 A_1) J_m(-hA_2B_2)\right|, \tag{2.123}$$

where $\mathfrak{B} \equiv B_1 + B_2$ and n and m are integers. We see the spectrum is characteristically broadened due to the finite bunch length σ_ζ. As we let $\sigma_\zeta \to \infty$, the bunching spectrum collapses around those points with $h = n + m\kappa$,

$$b_{n,m} = \left|\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} e^{-\frac{1}{2}(nB_1+(n+\kappa m)B_2)^2} J_n\left(nB_1 + (n+\kappa m)B_2\right) J_m\left(-(n+m\kappa)A_2B_2\right)\right|. \tag{2.124}$$

Therefore, the EEHG process produces bunching harmonics at the wavenumbers

$k = nk_1 + mk_2$. Prelimintary analyses of the EEHG phenomenon revealed that the maximum bunching factor is obtained from the $n = -1$ mode. Furthermore, recent analysis also suggests that the sensitivity of the system is proportional to $|n|$, providing more motivation for using the $n = -1$ configuration [76].

As the parameter space is four-dimensional, and the function extremely complicated, analytical optimization is not possible and numerical optimization must be performed. However, there are some rules of thumb for the optimization of an echo setup.

To begin with, as the final chicane is used merely to stand up the modulation of M2 in a manner similar to HGHG, we can borrow from that theory the (rough) relationship,

$$B_2 \sim 1/A_2. \tag{2.125}$$

Furthermore, it is well known that the ratio of the dispersions is approximately the harmonic number,

$$h \approx \left|\frac{nB_1}{\kappa B_2}\right|. \tag{2.126}$$

These relations, with a fixed set of A parameters, allow one a reasonable starting point for the echo optimization. The actual procedure is however done numerically, and one such optimization contour is shown in figure 2.28. The EEHG optimization space is markedly different from that of HGHG. The first notable thing is that for each optimal bunching, there is actually a twin optimum point with identical bunching and B_2, but different value of B_1. This comes about from the degeneracy in the $(nB_1 + (n + \kappa m)B_2)$ terms in the bunching formula. Although the bunching generated at precisely the n, m harmonic is identical, the performance of these two configurations in other respects is not. A second feature of note is that there are many isolated island pairs of bunching across the B_1, B_2 space. Indeed, one can identify several places where merely increasing B_2 would take the configuration between three or four of these islands.

For a given echo configuration, however, multiple harmonics besides the desired one may be excited. We show a plot of this effect in figure 2.29. We see that not only is bunching excited in a small island around the target harmonic of 25, but that there

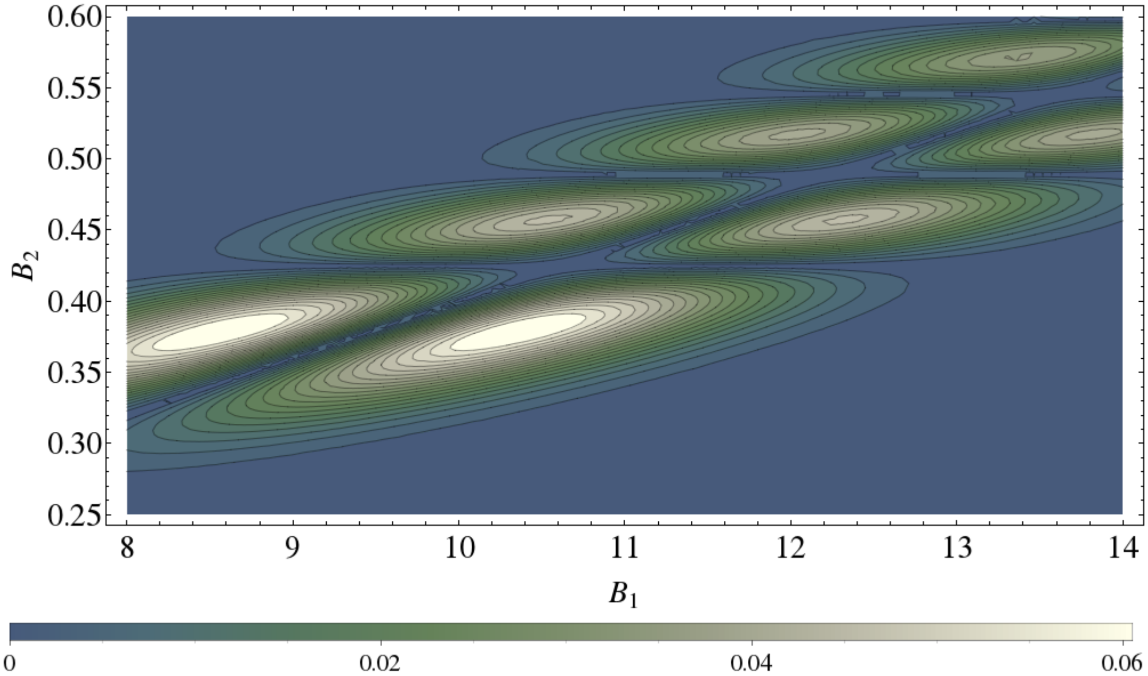


Figure 2.28: The EEHG bunching optimization space in terms of B_1,B_2. This setup has $A_1 = A_2 = 3$, $\kappa = 1$ and is for $n = -1$ $m = 26$.

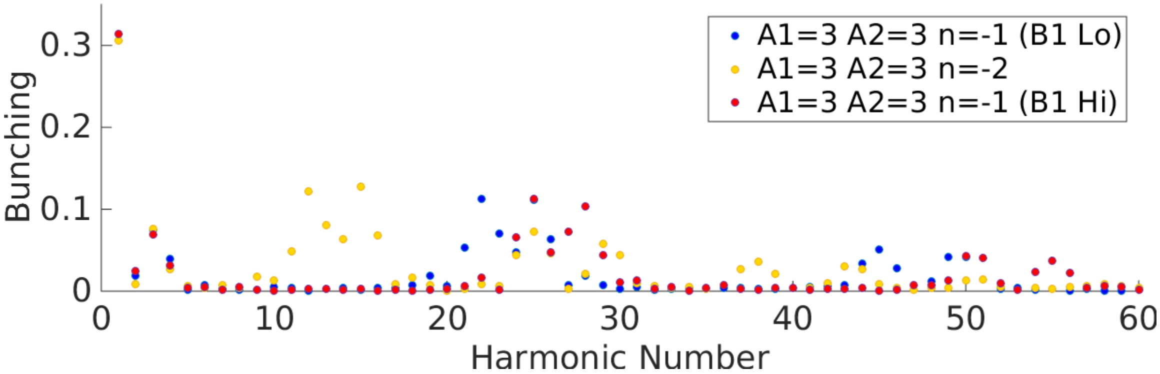


Figure 2.29: The bunching generated by several EEHG configurations. κ=1 for all these simulations, and the target harmonic for optimization was 25.

is also a small region of bunching at approximately twice this value, near a harmonic number of 50.

We investigate how these auxillary bunching regions change for various changes in the EEHG configuration. First, we switch to the $n = -2$ configuration and optimize for the same target harmonic. As a result, we see more islands of bunching outside the optimized location, but all generally of lower amplitude. This provides a strong motivation to operate at the $n = -1$ configuration whenever possible. In practice, this turns out to be a requirement on the dispersive strength of the first chicane, since we have the relation $h \approx |n|B_1/B_2$. If one cannot make B_1 large enough to reach the target harmonic h, one is forced to increase $|n|$ and decrease the bunching and performance of the EEHG configuration.

For the second set, we compared the B_1 high and B_1 low optimal configurations seen in figure 2.28. Both configurations have identical bunching at the target harmonic, however the B_1 low configuration excites harmonics below the target, while the B_1 high excites higher harmonics. Therefore, if one is concerned with exciting nearby harmonics around the target, the distinction between these configurations may become important.

To further investigate how EEHG treats nearby harmonics and side frequencies, define the following scaling parameter,

$$\xi(k) = \frac{k}{k_1}\mathfrak{B} - m\kappa B_1, \tag{2.127}$$

where this parameter is defined not only for the echo harmonics $k = nk_1 + mk_2$, but for all wavenumbers for a given set of n,m values. With this definition, we can rewrite the bunching factor at a general wavenumber k as,

$$b_{n,m}(k) = e^{-\xi^2/2} J_n\left(-\xi A_1\right) J_m\left(-\frac{k}{k_1}A_2B_2\right). \tag{2.128}$$

This expression reduces to the bunching factor at the echo harmonics as in equation 2.124 for $k = k_E = nk_1 + mk_2$. Since the expression 2.128 is valid for all k, for a given EEHG configuration it is referred to as the *excitation bandwidth* (EB).

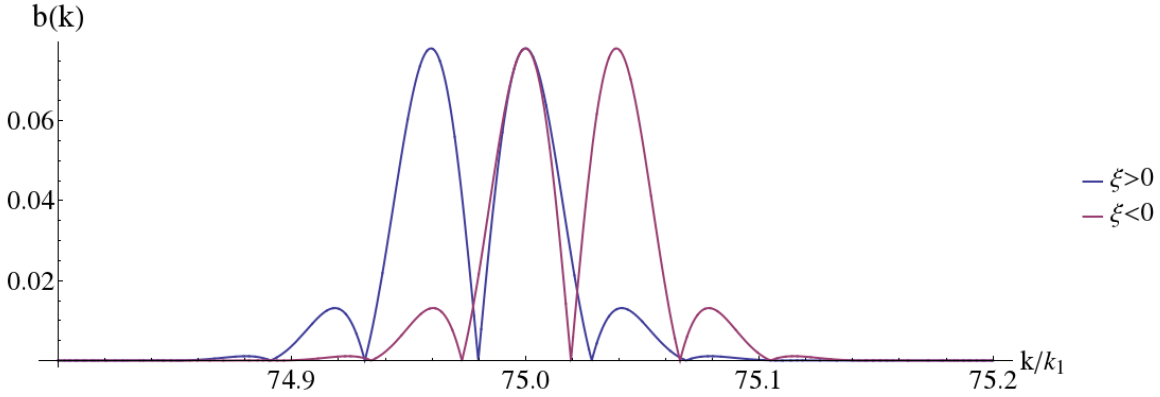


Figure 2.30: The excitation bandwidth for EEHG configurations with $\xi < 0$ and $\xi > 0$. The bunching is optimized for the 75th harmonic of the seed laser with $A_1 = A_2 = 3$ in the $n = -1$, $m = 76$ mode with $\kappa = 1$.

It defines how frequency components at wavenumber k are amplified or suppressed by the EEHG manipulation process, i.e. if it is capable of exciting those frequencies. Note the configurations referred to above as B_1 Hi or B_1 Lo are understood in this notation as either $\xi < 0$ or $\xi > 0$.

These positive or negative ξ configurations are equivalent in the bunching produced at the target harmonic, but have different EBs, as is shown in figure 2.30. While the bunching at the target harmonic (the 75th) is identical, we see that the $\xi > 0$ configuration preferentially excites nearby low frequency content while the $\xi < 0$ excites high frequency content. Furthermore, we note that the excitation side-bands are significant only quite close to the target harmonic. These sidebands therefore provide a means of controlling sub-harmonically spaced frequency content in the EEHG setup. This is further explored in the context of the microbunching instability in sections 5.1.2 and 5.2.

Perhaps the most substantial value in EEHG, as visible in the examples of 2.28 and 2.29, is the ability to attain high harmonics with relatively small energy modulations. It can be shown by analysis of equation 2.124 that for the $n = -1$ and m positive choice, the maximal bunching factor scales as,

$$b_{-1,m} \approx \frac{0.39}{m^{1/3}}. \tag{2.129}$$

This scaling with $m^{-1/3}$ is in stark contrast to the exponential scaling of the HGHG bunching factor. Further, the analysis leading to 2.129 is valid for $A_1 \approx 3$, which suggests that with this relatively low modulation amplitude, it should be possible to reach bunching values of several percent with harmonics over 100. For example, the optimized bunching factor with $A_1 = A_2 = 3$ at $n = -1, m = 101$ has $b_{-1,101} \approx 7\%$. This favorable scaling is what allows EEHG the promise to reach into the soft x-ray regime with a single stage of seeding.

However, in contrast to HGHG, EEHG has not yet been put in place at a major operational user facility. EEHG was demonstrated to be a suitable seed for lasing in the UV [77], and various proof of principle experiments have been performed which will be discussed in subsection 3.1.2. There are two key components to elevating EEHG to the status of a user facility ready technique.

The first is a demonstration of reliability and tunability sufficient to meet the demanding needs of users. Some of this experience has been gained with the similar HGHG setups at FERMI [70] [71], and also at the higher wavelength lasing demonstration [77]. The second component is the demonstration of harmonic up-conversion interesting enough to be competitive with other techniques. Given the success of HGHG providing seeds at FERMI well into the EUV, in practice this means the ability to reach the soft X-ray.

While it would have greatly pleased the author to have achieved both these criterion and demonstrated a fully operational, user-ready EEHG seeded beam, this was not possible. Rather, by demonstrating a harmonic up-conversion factor of 75 in the scaled experiments at NLCTA, the second key to the puzzle appears to be well in hand. What remains, therefore, is to combine the knowledge and experience in producing high harmonics acquired at NLCTA with the operational experience of other seeded FEL facilities. This merger would provide an excellent shot at producing a user ready EEHG-seeded FEL beam in the soft X-ray.

Chapter 3

Characterization and Upgrades of the NLCTA

> If I have not seen as far as
> others, it is because there were
> giants standing on my shoulders.
>
> — Hal Abelson

In this chapter, I will outline the Next Linear Collider Test Acclerator (NLCTA) facility wherein the experimental work of this thesis was performed. I will first review the history of the NLCTA in section 3.1, including the previous experiments related to Echo-Enabled Harmonic Generation performed prior to my arrival. After this review, I will discuss in section 3.2 the upgrades to the facility I performed with the intention of enabling the production of higher harmonics. These upgrades include an EUV spectrometer which I designed, constructed, and calibrated to facilitate the measurement of high harmonic EUV light produced through the Echo process. Finally, I will describe the characterization of the whole NLCTA facility I performed through simulations and electron beam measurements in section 3.3.

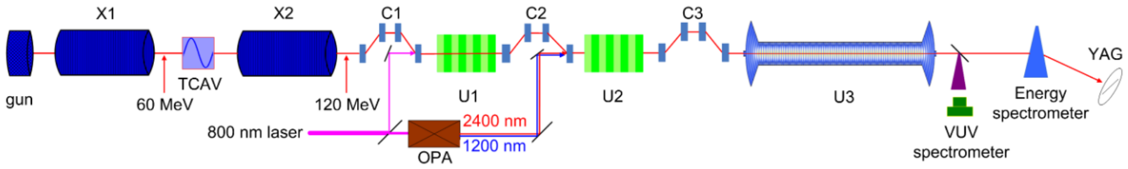


Figure 3.1: The NLCTA main beamline as of Fall 2014. Figure from [83]. An updated version, which includes the large first chicane and E163 dogleg is given in figure 3.10.

3.1 NLCTA Overview

3.1.1 The NLCTA Facility (2009-2014)

The Next Linear Collider Test Accelerator (NLCTA) was originally conceived as a testbed for the ill-fated Next Linear Collider (NLC) of the 1990s [78]. In particular, high gradient X-band accelerating structures were needed for the NLC and it was necessary to test them and demonstrate the compensation for the beam loading effect [79]. Therefore, the primary goal of the NLCTA was to study the generation of high powered X-band RF, its transport into high gradient structures, and the dynamics of the electron beam in this environment. After this initial period, the NLCTA was rebranded as a user-oriented facility named ORION [80] where the focused shifted away from RF to advanced accelerator concepts. In particular, the E163 experiment utilized the electron beam and constructed many laser subsystems to test laser acceleration both in vacuo [81] and inside a dielectric structure [82]. Concurrent with these laser acceleration experiments was a secondary focus towards phase space manipulation of the electron beam using beam-laser interactions.

The status of the facility as of Fall 2014 is shown in figure 3.1. The overall facility layout has not changed significantly between 2014 and this work, so we will first describe the beamline before later detailing the modifications that were made as part of this work.

The electron beam for the NLCTA facility is generated from a copper photocathode embedded in a 1.6 cell S-band copper photoinjector of the BNL/ANL/UCLA/SLAC type [84]. Emittance compensation in this low energy injector section is provided by a solenoid magnet located approximately 18cm from the cathode in a manner described in reference [85]. The beam has a nominal energy of 5 MeV as it enters the 180cm long X-band accelerating cavity X1, after which its energy is nominally 60 MeV. At this point, there exists a large four-dipole chicane which occupies 6.5m of floor space (shown in figure 3.10). This chicane is not commonly in use, however, and a bypass line instead takes the beam straight for this drift distance. An 11-cell X-band transverse deflecting cavity was then present, but was removed prior to this work. Finally, the beam reaches the second X-band accelerating structure which increases the energy to a nominal 120 MeV.

This point is the beginning of the nominal ECHO beamline, and begins with a chicane (C0) used to generate an orbital bump for the injection of an 800nm laser. The beam then traverses the first undulator (U1, 10 periods, $\lambda_u = 3.3\text{cm}$, $K = 1.82$) before arriving at a second chicane (C1). This second chicane again provides an orbital bump for the injection of the second echo laser, while also providing the dispersive R_{56} strength needed for the echo effect. The second echo undulator (U2, 10 periods, $\lambda_u = 5.5\text{cm}$, $K = 2.76$) is then encountered before a third and final chicane (C2) completes the echo process. The beam then enters a final radiating undulator (U3) for the generation of radiation and diagnosis of the echo-induced bunching. In the studies of 2014, and shown in figure 3.1, this final undulator was an electromagnetic RF undulator with period 1.39cm and adjustable K ($0.2 \lesssim K \lesssim 0.7$) [86]. Photon diagnostics were performed using a commercial McPherson model 234/302 scanning monochromator [87], equipped with a 2400G/mm or 1200G/mm corrected concave grating. The detection of photons was ultimately performed by an Andor CCD camera. Finally, the electron beam enters a dipole magnet spectrometer to allow for electron energy and longitudinal phase space diagnostics.

The chicanes used for the echo manipulations are composed of four edge dipole magnets to automatically eliminate horizontal dispersion components. However, they also contain two quadrupole magnets between the first and second and third and fourth dipole magnets. These quadrupoles are referred to as trim quadrupoles. A drawing of the second echo chicane (C2) is shown in figure 3.2 showing the location of the trim quadrupoles. The quadrupoles are not only allowed to freely vary in strength

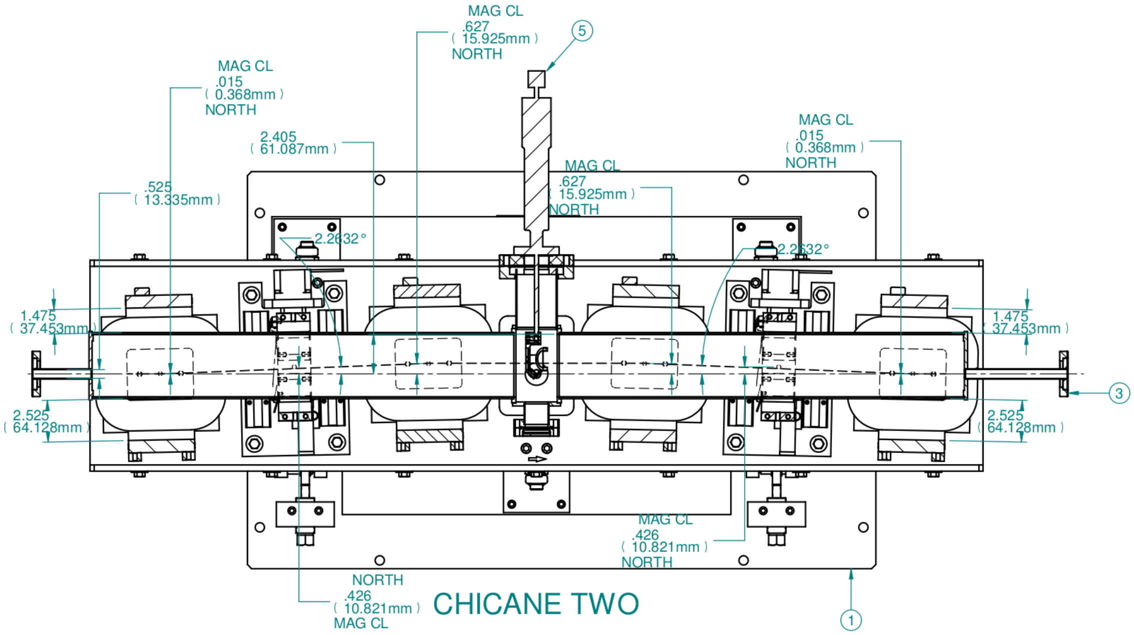


Figure 3.2: A computer drawing of the second echo chicane. The magnetic center line of the chicane is shown as the dashed line. The objects between the first and second and third and fourth dipoles are the trim quadrupoles. In the center is a pop-in screen, which ejects laser 2 for this chicane, but is otherwise used as an insertion point for the echo lasers in the other two chicanes.

(positive and negative field values), but are also mounted on linear motion stages which allows them to be moved to center on the true magnetic and orbital center line. These quadrupoles are used to remove any remnant linear horizontal dispersion R_{16} in the chicane, which can arise from imperfect magnetic fields and fringe fields. Since even small dispersions can lead to large subsequent deviations compared to the radiation harmonic, it is imperative to remove this dispersion. A solution to lowest order can be found analytically for the two trim quadrupoles assuming ideal chicane optics. In practice however, the quadrupoles must be fine tuned to account for magnetic imperfections and misalignments, which accounts for a significant portion of the echo tuning for high harmonics.

These chicanes have a minimum usable R_{56} component of ≈ 500 µm to avoid scraping of the electron beam against the inserted mirror. Magnetic hysteresis and other imperfections lead to a known ≈ 100 µm uncertainty in the dispersion at such small values. The maximum R_{56} attainable also limits the utility of the Echo technique, and for the large first chicane is limited to a maximum of about 13 mm.

Throughout the beamline, and at the final dipole spectrometer diagnostic, Optical Transition Radiation (OTR) screens are employed for direct imaging of the electron beam.

The NLCTA is equipped with an array of lasers, both for the generation of the electron beam and the scientific programs. The laser begins in a mode locked Ti:Sa oscillator at 800 nm, which is then split into two separate pulses. One traverses a regenerative amplifier before transport to the electron gun, where a nonlinear BBO tripling crystal transforms it into the UV at 266 nm, after which it is used to photoemit electrons from the copper cathode. The other branch traverses a separate regenerative amplifier before being split again in paths to either the first or second undulator (injection at either C1 or C2 in figure 3.1). The injection point at C2 is further equipped with an optical parametric amplifier (OPA), which allows the wavelength to be tuned, notably, between 1600 $\sim$ 2400 nm. By the time these pulses reach the undulator sections for interaction, they have approximate duration of 1 ps.

3.1.2 The Echo Programme (2009-2014)

The first foray into EEHG was achieved at NLCTA in 2010, when the 3rd and 4th harmonic of the second seed laser were generated [88]. This established the underlying physics of the EEHG phase space manipulation, and importantly, distinguished this effect from the already-known HGHG signals. The relative insensitivity of the central echo wavelength to electron beam chirp was also demonstrated in this early experiment.

Focus then turned towards higher harmonics, as well as the ability of EEHG to generate high harmonics with only modest energy spread increases. Next, the 7th harmonic of the 2nd seed laser was reached, and clear evidence of bunching was observed when the energy modulation was only 2-3 times the slice energy spread. This demonstration established the key result that holds the promise for EEHG to reach high harmonics for beam based FEL seeding in a single stage [89].

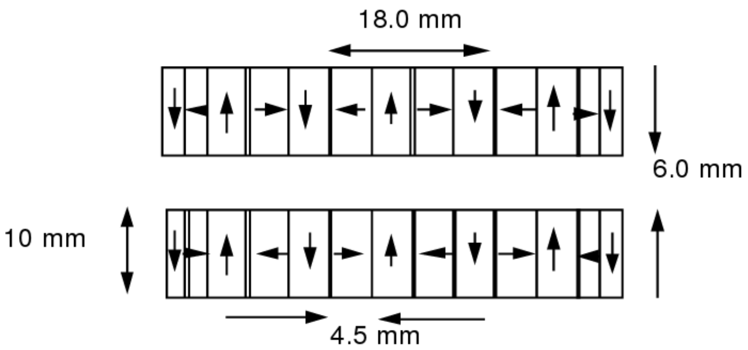


Figure 3.3: The Halbach mangetic array setup of the VISA undulator.

Finally, in 2014 the harmonic number was pushed well into the VUV with the 15th harmonic being generated [83]. Importantly, at this stage of the experimental programme, the 2nd seed laser was altered to the 2400nm wavelength. This change allowed higher harmonics to be accessed without decreasing the target wavelength, which necessitates more difficult diagnostics. During this study, several other fundamental properties of the echo signal were established, such as the insensitivity to electron beam energy curvature, and the robust central wavelength with respect to electron beam energy chirp. At this point, analysis suggested that considerable bunching ought to have been in place on the electron bunch up to harmonics 30 $\sim$ 45. However, due to the relatively low (120 MeV) electron beam energy, the properties of the radiating undulator (at this time, the RF undulator [86]), and the photon diagnostics in place, it was not possible to generate or measure enough signal to confirm the echo effect up to these harmonics. Bearing in mind that at a full scale FEL facility, a target wavelength in the region 3 $\sim$ 5nm meant upconversion of a 266nm seed laser to the 50 $\sim$ 75th harmonics, the next experimental push was devoted to demonstrating the highest harmonic upconversion.

3.2 Experimental Upgrades towards Higher Harmonics

The first upgrade to enable the investigation of higher harmonics was the installation of the Visible to Infrared SASE Amplifier (VISA) undulator. The need for a strong radiation signal in order to perform diagnostics on the high harmonic bunching necessitated the push towards a longer, more efficient radiator than the pre-existing RF undulator. The VISA undulator was originally designed as a high efficiency undulator to investigate the properties of SASE radiation in anticipation of the LCLS experiment [90]. While the original VISA undulator is a full 4m in length, for the experiments at NLCTA only a 2m section was used. The magnetic array is of a Halbach design [91] with permanent NdFeB magnets in the configuration illustrated in figure 3.3. The magnetic gap is fixed at 6mm, with period 1.8cm, and a peak on-axis B-field of 1.25T. The VISA undulator thus has strength $K = 1.26$.

Furthermore, the VISA is equipped with a distributed focusing array which yields a FODO cell length of 24.75 cm, with a focusing gradient of 33.3 T/m and an effective quadrupole length of 6cm. For the operating points of $E = \{120, 160, 192\}$ MeV, this yields an average matched beta function of $\langle\beta\rangle = \{42, 55, 65\}$ cm. This small beam size presents the tantalizing possibility of high gain performance in the EUV that was unfortunately unable to be realized due to other limitations in the electron beam parameters.

The alignment of the VISA undulator was originally carried out using a pulse-wire setup, which provided an accuracy of 50 µm for the magnetic axis [92]. Once installed on the NLCTA beamline, the straight line trajectory through the undulator was established via a small guide laser injected into the beamline immediately upstream of the undulator. This ballistic trajectory was used as the guide for the electron beam trajectory through the undulator as measured by the intra-undulator OTR screens.

The VISA undulator is additionally equipped with four picture-frame dipole corrector magnets to allow beam steering within the undulator. These dipole magnets are capable of independently providing a maximum of 2 mrad kick in the x or y direction. For intra-undulator diagnostics, the undulator is also equipped with four evenly

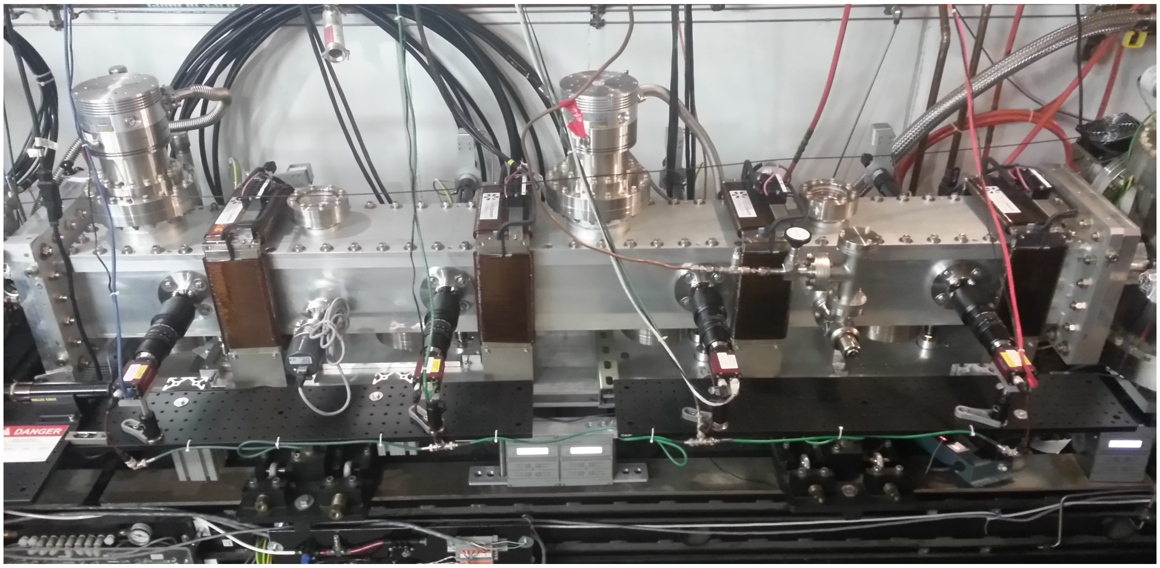


Figure 3.4: The VISA undulator as installed on the NLCTA beamline. The electron beam enters from the left. Visible are the four dipole corrector magnets, as well as the four OTR imaging cameras.

spaced OTR screens mounted on pneumatic actuators. The transition radiation can be imaged on cameras, which allow a positioning of the beam and intra-undulator monitoring of visible spectrum radiation. The VISA undulator as installed on the NLCTA beamline is shown in figure 3.4.

The push towards higher harmonics, and naturally shorter wavelengths, created a strong drive towards higher electron beam energies. At the nominal operation energy (in 2014) of 120 MeV, the VISA undulator is resonant at $\lambda_r \sim 290$ nm. To diagnose EEHG bunching at the 75th harmonic of the 2400 nm seed laser would therefore require radiation on the 9th undulator harmonic. For the modest $K = 1.26$ of the VISA, it is clear to see from figure 2.11 that the radiation coupling is quite low at this high harmonic. In fact, the coupling to the radiation at the 9th harmonic is only $\approx 3\%$ of the coupling to the fundamental. Therefore, in order for the radiator to provide an efficient diagnostic of the electron bunching it was necessary to significantly increase the electron beam energy.

To this end a third, one meter long, X-band accelerating structure (X3) was installed following the final EEHG chicane C2. As the EEHG modulator and laser

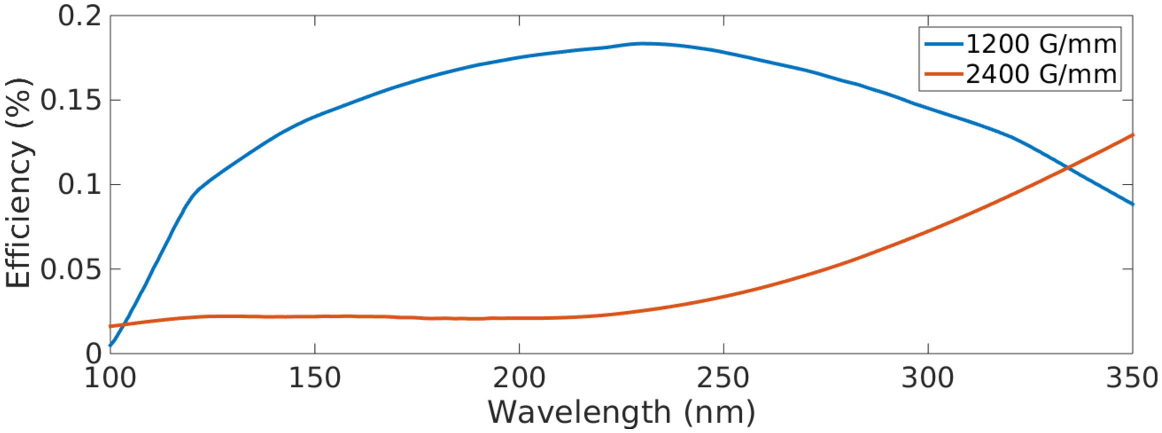


Figure 3.5: The theoretical efficiency of the commercial VUV spectrometer as a function of wavelength for each of the available gratings.

system were designed for resonance at 120 MeV, it was not possible to place the final accelerating structure earlier in the beamline. Initially driven by a normal X-band klystron, this structure provided a maximum of ≈ 45 MeV energy boost, bringing the total beam energy to ≈ 165 MeV. Some time after the initial installation, the X-band klystron was modified to include a SLED cavity [93], which extended the range of the electron beam energy to a maximum of 196 MeV, although the more stable 192 MeV became the highest used operational point. At this 192 MeV working point, the VISA is resonant at 114 nm, allowing the 75th harmonic to be radiated near the 3rd or 4th undulator harmonic.

Because X3 was installed downstream of the EEHG manipulation section, the delicate EEHG phase space needed to be transported through the structure. Experimentally, the bunching was observed to survive the acceleration section, although it is not clear how much degradation (if any) was caused by either the acceleration or wakefields due to the structure. Regardless, the configuration suggests the possibility that future EEHG setups could perform beam manipulations at a relatively low energy and only later boost to the final energy for the radiation stage. Such a setup would allow a significantly smaller (in terms of both laser power and chicane design) EEHG section, and therefore may be an attractive economical possibility.

Finally, the need to diagnose high harmonics with radiation wavelengths that extended well into the EUV necessitated the construction of new photon diagnostics. The previously installed McPherson monochromator [87] used a 45° angle of incidence (AOI) UV enhanced aluminum mirror to deflect light into the spectrometer placed 90° to the beamline. The light then scatters off a grating (either 1200 or 2400 grooves/mm) at a relatively high (32°) AOI. The photon detection was performed by an Andor CCD with typical QE of $\sim 30\%$ over the spectral range of interest. The theoretical efficiency of the setup up to the Andor CCD is plotted in figure 3.5. The efficiency of the gratings is taken directly from the manufacturer, while reflectance data is not provided for the ejection mirror below ≈ 200 nm. Data in this range is therefore interpolated between the high wavelength data and known low-wavelength data in the X-ray regime for Al from a wavelength of 41 nm [94]. The overall low transmission efficiency is due to the low efficiency of the gratings ($\mathcal{O}(0.1)\%$), while the dropoff in efficiency near 150 nm is due to the poor performance of the ejection mirror at the relatively high AOI.

Due to the long lead time required for the construction of the new EUV spectrometer, an intermediate solution was required to allow the program to progress. As the geometry and efficiency shown in figure 3.5 were essentially fixed, the only option to extend the range of the setup was to increase the gain of the photodetector. To this end, the CCD camera was replaced with an 18 mm microchannel plate (MCP) (BOS-18, Beam Imaging Solutions) to provide a large gain and extend the range of the device. The MCP detector consisted of two back-to-back MCP devices, with the first plate coated in CsI to enhance the detection of EUV photons. A P-43 phosphor screen was placed behind the MCP to convert the MCP electrons back into photons for detection on an imaging camera focused to the plane of the phosphor. The MCP is powered by a 2 kV high voltage power supply, while the phosphor is biased by a 5 kV power supply. The gain of this setup depends exponentially on the MCP supplied voltage, but can conservatively be estimated to be 10^4. The spatial resolution is determined by the pitch size of the MCP (12 µm), the gain of the device, and the phosphor screen. For large gains in double MCP configurations, the cascade electrons spread to multiple channels and degrade the resolution. The resolution was

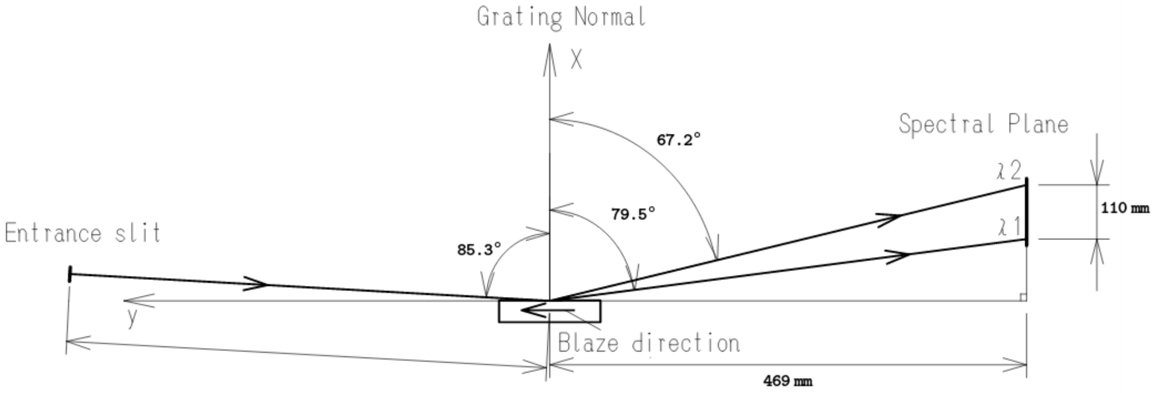


Figure 3.6: The geometry of the Hitachi EUV diffraction gratings. The wavelengths of the spectral plane λ_1,λ_2 depend on the groove density of the grating. Figure adapted from [95].

optimistically estimated, based on similar devices, to be 50 µm, although it appeared to be significantly higher in the case of our devices. Regardless, the installation of the MCP effectively increased the functional range of the VUV spectrometer down to ≈ 45 nm, although detection efficiency was lost above ≈ 150 nm due to the CsI coating.

3.2.1 Constructing the EUV Spectrometer

There were several important criterion for a new EUV spectrometer to enable the EEHG experiment to reach its goal. The first, and most obvious, was a useful wavelength range which extended at least to the lowest wavelengths expected: in our case, we aimed for the 75th harmonic of the 2400 nm seed laser, i.e. 32 nm. However, the process of tuning and diagnosing EEHG harmonics generally starts with lower harmonics which are bright and stable, and allow for a relatively easy tuning of the seed lasers and electron orbit. After efficient bunching is established at these low harmonics, the harmonic number can be gradually raised by adjusting the seed laser powers and chicane settings. Therefore, an ideal spectrometer will also have a large wavelength range within a single grating to allow for a large range of tuning without the need to physically modify the spectrometer setup. The existing VUV spectrometer, after the upgrade with the MCP, allowed diagnostics at low efficiency in the $45 \sim 150$ nm range, so the upper desired wavelength range for the EUV spectrometer was chosen to be around 100 nm for substantial overlap. It is important to note here that EEHG machine parameters had already been demonstrated at 160 nm in the previous experiment [83], so this became the approximate starting point for the upgraded experimental setup.

The second requirement for the spectrometer was that it be relatively high efficiency. Although the increase in beam energy significantly enhanced the coupling of the high echo harmonics to the VISA undulator, the radiation for 32 nm light still took place between the 3rd and 4th undulator harmonic. This fact, coupled with other deleterious effects which will be discussed later, necessitated a high efficiency spectrometer that could capture as much of the light as possible. In the EUV, this translates to a spectrometer constructed with grazing incidence optics in order to maximize the reflectance over the wavelengths of interest.

The need for a grazing incidence instrument conflicted with the more practical requirement of space constraints in the accelerator housing downstream of the VISA undulator. The X-band Test Area (XTA) is a small, independent accelerator within the same housing as the main NLCTA beamline which happened to reside directly adjacent to the terminus of the VISA undulator. The need for seamless transition between the operation of either the XTA or NLCTA beamlines required that a) the XTA setup could not be disturbed, and b) there needed to be sufficient clearance between the two to allow for routine work and access to the final segment of either beamline. Therefore, the overall design of the EUV spectrometer also needed to be compact enough to fit these space requirements.

Finally, it was necessary to have an entrance slit to provide the highest achievable resolution and limit the impact of any stray reflections. Due to the relatively long electron bunch, the bandwidth of any individual EEHG harmonic is quite narrow ($\mathcal{O}(0.01)$ nm). With the envisioned MCP setup, it was not considered achievable to attain such a high resolution due to the need for a high gain. The theoretical resolution of the spectrometer setup was limited only by the smallest attainable slit gap, the pore size and pitch of the MCP, and the phosphor grain, and was actually quite small ($\mathcal{O}$(0.01 nm). However, experience with similar setups which required considerable gain from the MCP showed that such a high resolution was not attainable in practice [96] [97], and a more reasonable estimate for the resolution was $\sim$ 0.1 nm. The requirements on the smallest attainable slit gap were therefore relaxed compared to an ideal spectrometer.

The heart of the new EUV spectrometer was a pair of Hitachi aberration-corrected concave gratings (models 001-0639, 001-0640) with 600 and 1200 G/mm respectively. These gratings have an angle of incidence (with respect to the tangent) of 4.7°, a blaze angle of 3.7°, and focus to a flat field 469 mm along their tangent away from the center. The blaze angle is a slight sawtooth tilt at the level of the grating line spacing designed to optimize diffraction at a certain wavelength and diffraction order (usually the first). In this case, the blaze for gratings 001-0639 and 001-0640 are optimized for 31 and 16 nm respectively. The flat field, aberration-correction, and blaze angle create an effective wavelength range of 22 $\sim$ 124 (11 $\sim$ 62) nm over a 110 mm flat focal plane. The existence of a flat focal plane allowed the simultaneous observation of a large wavelength range without the need for spectrometer adjustment. In practice, the simultaneously observable wavelength range was limited not by the grating but by the size of the detector at the focal plane. The overall geometry of the gratings is illustrated in figure 3.6. The light was ejected from the electron beamline by a gold-coated mirror at 15° angle of incidence into the spectrometer arm. The 15° AOI offered a total 30° angular kick out of the beam path, which was critical to satisfying the geometrical requirements of figure 3.6. As the light coming from the undulator was primarily horizontally polarized, the reflectivity of the gold mirror over the wavelengths of interest is generally $\sim$ 40% [94].

The net efficiency of the combined mirror-grating system is shown in figure 3.7. Compared with the VUV spectrometer (shown in figure 3.5), efficiency was increased by almost two orders of magnitude. This is primarily due to the grazing incidence angle of the setup, both in the diffraction grating and gold-coated ejection mirror. Further, due to the blazing process on the diffraction gratings, the efficiency was more peaked than the VUV spectrometer around this blaze wavelength. The upper and

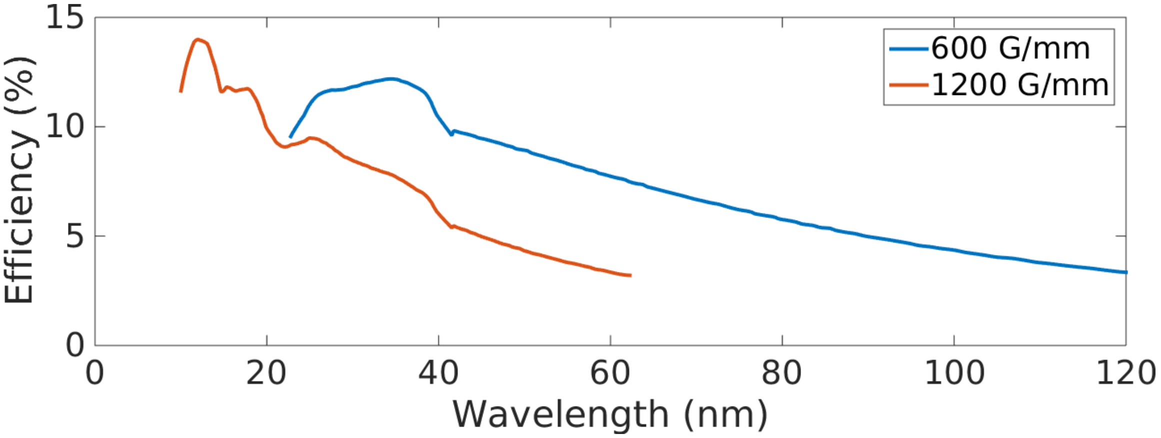


Figure 3.7: The theoretical efficiency of the constructed EUV spectrometer (Au ejection mirror + grating) for each grating. The reflectance of the gold ejection mirror is assumed constant in the wavelength range 41 $\sim$ 120 nm.

lower wavelength limits of each grating are due to the lack of achievable focus and aberration correction in the focal plane for wavelengths outside the design region.

Before reaching the grating, the light was baffled by a custom slit assembly. During the design of the spectrometer, several commercial designs were considered and ultimately rejected due to the space constraints discussed, the need for high vacuum, and the desire for a wide and controllable aperture range. A custom assembly was therefore constructed based upon two 2 3/4” vacuum cubes mounted vertically atop one another. Into each was inserted a linear actuator with a stepper motor attached to a straight (or cantilevered) slit blade. The stepper motors allowed each side of the slit to be moved independently to center the slit on off-axis or misaligned radiation while maintaining a constant aperture. The vertical orientation of the vacuum cubes and associated linear actuators required only a minimal footprint on the spectrometer beamline, which fit within the space constraints described. A computer aided drawing of the designed slit assembly is shown in figure 3.8.

The grating itself was mounted in a cylindrical vacuum chamber on a fixed mount. The linear motion stage at the focal plane eliminated the need for a rotation mount to view the entire spectral range, as was required in the VUV spectrometer setup. Furthermore, the fixed mount allowed greater confidence in the fixed geometry of

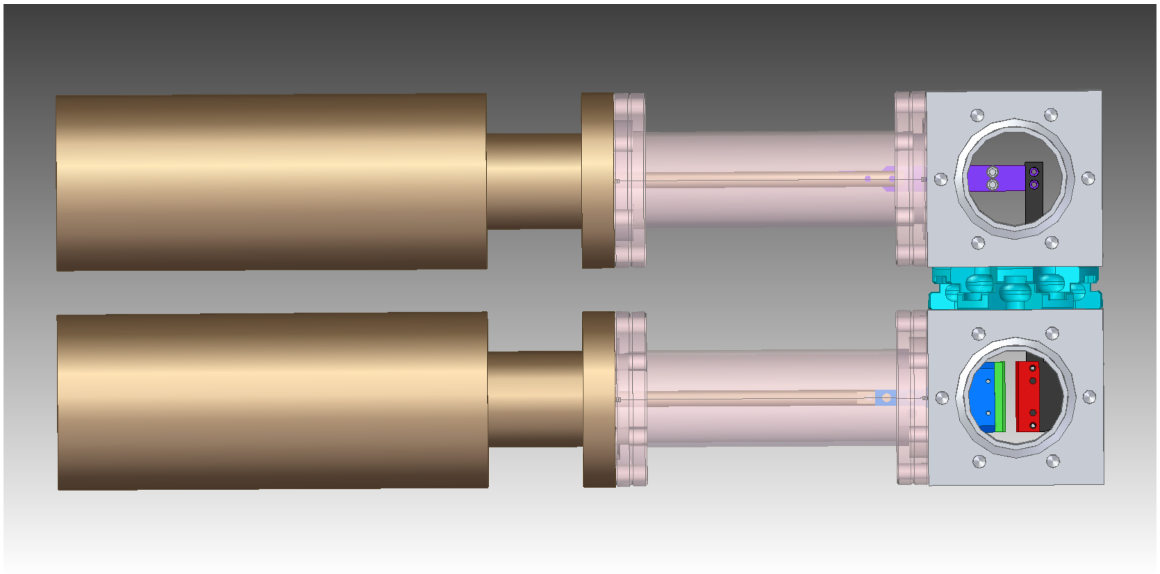


Figure 3.8: A computer drawing of the custom slit assembly showing the independent linear actuators and physical slit blades.

the setup, which was critical to the calibration of the spectrometer. A large turbomolecular pump was mounted atop this grating chamber to create a high vacuum environment necessary for the MCP high voltage operation, grating cleanliness, and EUV propagation.

After being diffracted off the grating and leaving the holding chamber, the light traversed a 6” diameter, approximately 14” long corrugated high vacuum bellows. The large diameter of this bellows was chosen so as to allow light (including the zeroth order reflection) to travel all the way to the focal plane of the grating without secondary reflections. Furthermore, the end of the bellows was attached to a linear translation stage which allowed it to traverse approximately 120 mm parallel to the focal plane of the grating. This 120 mm displacement over the length of the 14” bellows created a strong shear which was handled by the corrugated structure. Ultimately, however, the buckling of the corrugation structures at extreme shears limited the translation to this 120 mm range of motion. Since the bellows also needed to traverse to the zeroth order radiation, the motion did not quite cover the entire 110 mm focal plane of the grating.

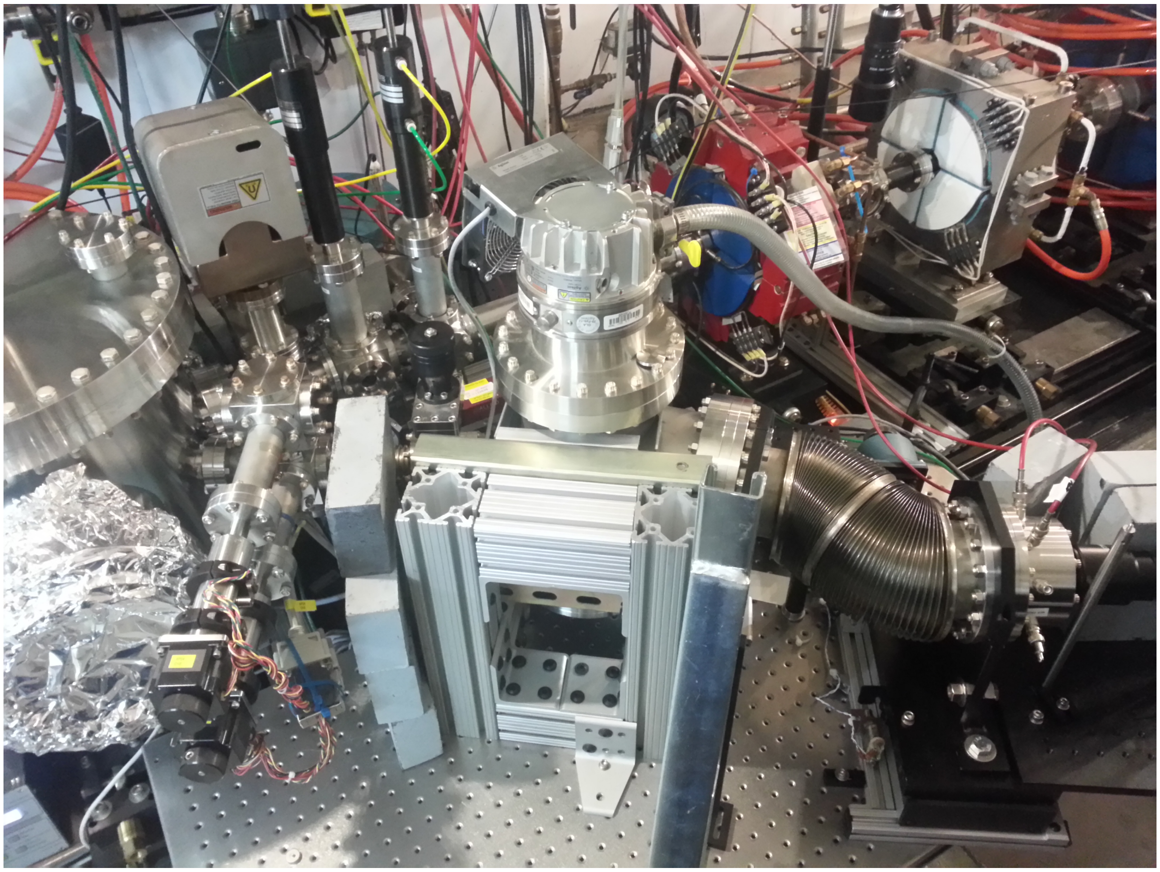


Figure 3.9: The EUV spectrometer as installed on the NLCTA beamline.

The light finally reached a 40 mm MCP (BOS-40, Beam Imaging Solutions) mounted to the end of the bellows section. This MCP was in a dual chevron configuration, had a CsI coating, and 5 µm pore size. The MCP was operated off the same 5 kV power supply as the smaller MCP which was installed on the VUV spectrometer. Behind this 40 mm MCP was another P-43 phosphor screen which was then directly imaged with a camera to observe the spectrum.

The full EUV spectrometer as installed on the beamline is shown in figure 3.9. The light was ejected by the gold mirror in the large vacuum chamber to the left of the image (electrons continue on towards the top-right in the main beamline). The two-cube slit assembly is visible, while the grating chamber (center) is partially obscured due to its support. To the right, the sheared bellows is seen near its position

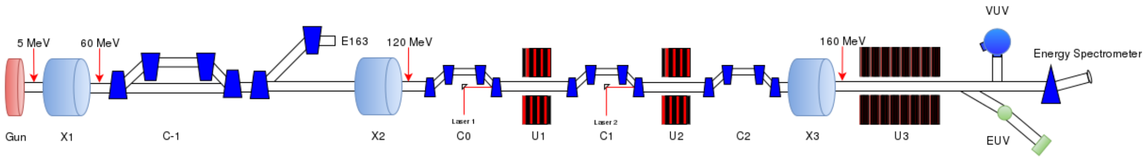


Figure 3.10: The final (2015) NLCTA beamline layout showing the accelerating structures, modulators and chicanes, the VISA undulator (U3), and photon and electron beam diagnostics. The first large, by passed chicane and E163 dogleg are also shown in this updated diagram. Shown is the energy of 160 MeV, but with the SLED upgrade this could be increase to 192 MeV.

of maximal shear, representing the limit of travel for the stage. Finally, the circular device attached to the end of the bellows is the 40 mm MCP, the phosphor screen of which was imaged by a camera off image right. As the light was ejected by the gold mirror, the electron beam also strikes the mirror and generates secondary particles which create a noisy image on the high-gain MCP. Therefore, an effort was made to shield the MCP from these particles by employing lead bricks placed roughly between the mirror and the MCP position. Furthermore, the large torque generated by the high vacuum on the grating chamber and bellows necessitated the construction of a strong back support to avoid inadvertent motion during either pumpdown of the vacuum line or the experiment.

The final layout for the main NLCTA beamline during the echo experiments is shown in figure 3.10.

3.2.2 Calibrating the EUV Spectrometer

Ideally, in order to make a calibration of both the wavelength scale and intensity of the spectrometer, a standardized source would be used and measured. Typically, a plasma discharge lamp of some kind is used to provide spectral lines across the EUV which can be used for the wavelength and relative sensitivity calibration [98] [99]. These EUV sources are often themselves calibrated against a synchrotron storage ring, which can provide an absolute sensitivity calibration. Unfortunately, calibration against such an external source was not possible for the EUV spectrometer at NLCTA. This was due primarily to the fact that the EUV spectrometer was constructed in-place on

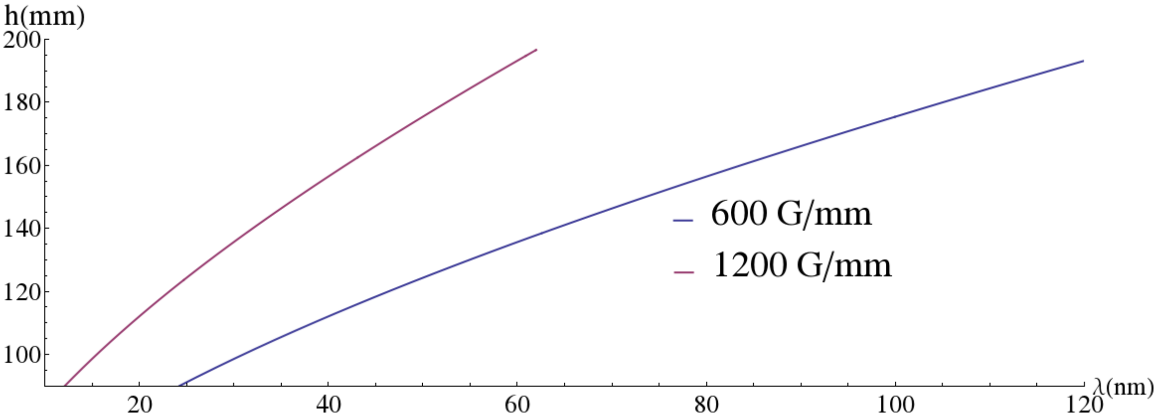


Figure 3.11: Theoretical wavelength calibration for the EUV spectrometer along the h axis from equation 3.2.

the beamline, and it was not possible to quickly manipulate beamline components in order to insert or remove a calibration source. The secondary reason for lack of absolute calibration, as with several aspects of the spectrometer construction, was the time constraints. Therefore, a calibration was made based on diffraction grating theory and the known properties of the radiation.

The theoretical foundation for the spectrometer is the underlying grating equation, which states that constructive interference is obtained only when the grating geometry satisfies [100],

$$m\lambda = d\left(\sin\alpha + \sin\beta\right), \tag{3.1}$$

where m is the integer diffraction order, λ is the incident wavelength, d is the linear groove density, and α and β are the incidence and exit angles (with respect to normal). From the diffraction grating geometry 3.6, we aim to find the linear position along the focal plane as a function of incidence wavelength at the first order diffraction $m = 1$. The simple trigonometry furnishes the condition,

$$h = r_{\text{focal}}\left(\frac{\sqrt{1-\left(d\lambda-\sin\alpha\right)^2}}{d\lambda-\sin\alpha}\right), \tag{3.2}$$

where r_{focal} is the distance along the Y axis to the focal plane and h is the linear

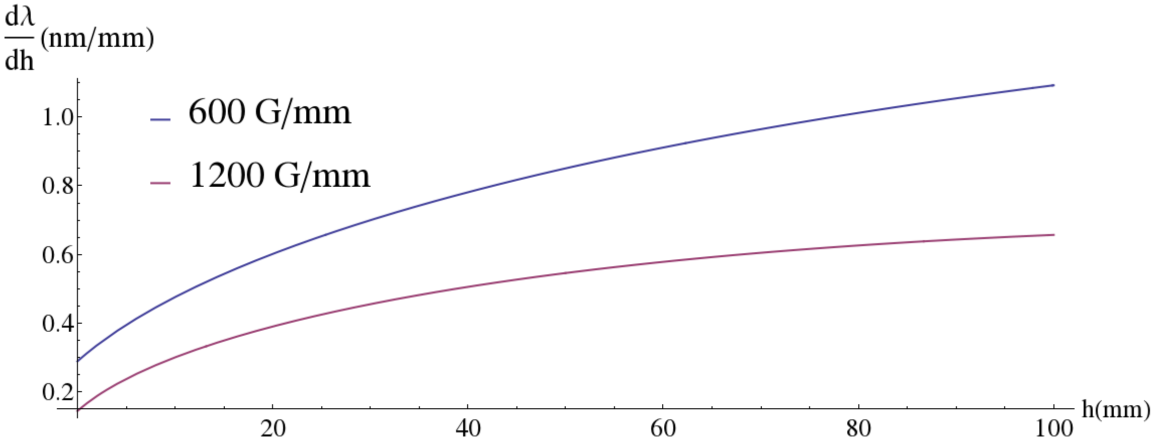


Figure 3.12: The calculated dispersion of the Hitachi gratings as a function of distance along the focal plane h.

distance along the focal plane in the X direction (see 3.6). This relationship provides a theoretical calibration for the EUV spectrometer, assuming perfect alignment and transport, and that all quantities are known to sufficient precision. A plot of this theoretical calibration is shown in figure 3.11. Furthermore, this relationship can be inverted to find the dispersion $d\lambda/dh$ as,

$$\frac{d\lambda}{dh} = \frac{hL}{\left(h^2 + L^2\right)^{3/2} d}. \tag{3.3}$$

Note that due to the flat field criterion, the dispersion along this plane is not linear, and this relationship is plotted in figure 3.12.There is significantly higher spectral resolution in the low wavelength region (low h) than the high wavelength region. In the operation of the EUV spectrometer, the location of the zeroth order diffraction was marked on the MCP screen as a reference zero pixel. Then, using the theoretical relationship 3.2, the linear stage was moved a corresponding distance to the desired central wavelength. The linear stage can be controlled to a fraction of a mm, leading to a pessimistic pointing accuracy of $\lesssim 0.1$ nm. Based on the central pointing wavelength and the known position of the zeroth order, a seamless correspondence between pixel position and wavelength position was established over the entire spectral range.

The actual source for the wavelength calibration was the harmonics produced

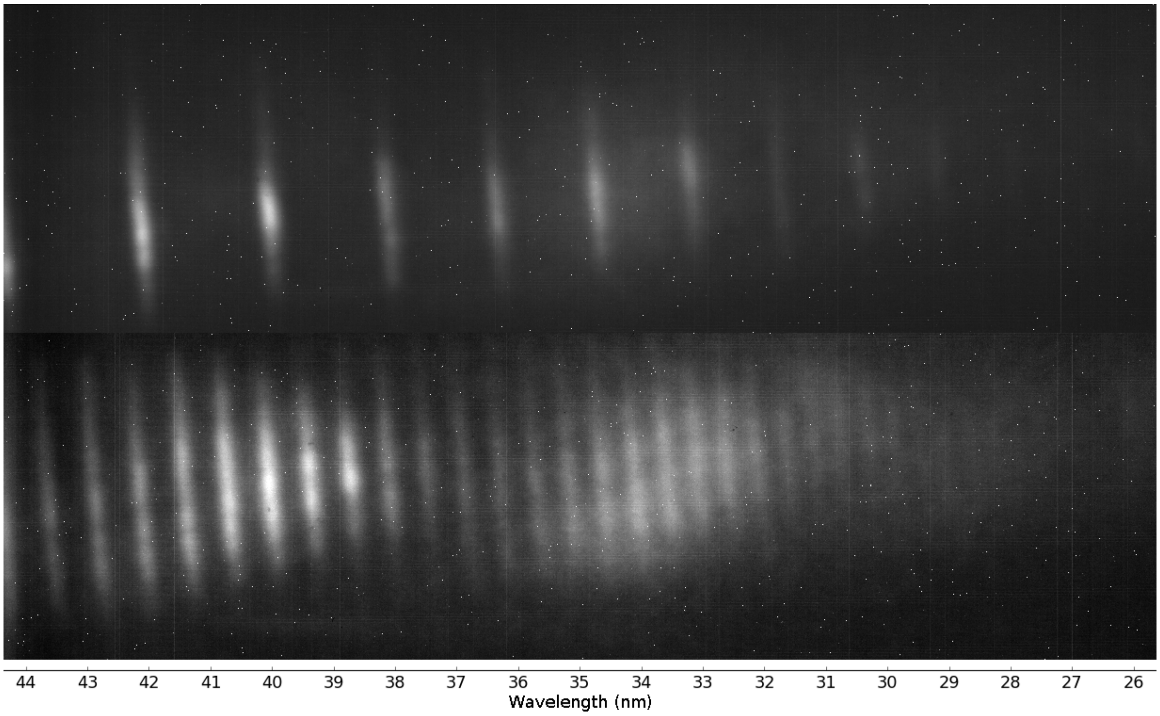


Figure 3.13: Wavelength calibration of the EUV spectrometer using HGHG light produced with either an 800 nm (top) or 2400 nm (bottom) laser. The vertical axis is the physical vertical extent of the radiation. Individual harmonics are easily distinguishable well through the wavelength band of interest. The spectra have a slight tilt owing to a slight misalignment of an optical component in the spectrometer.

by the seeding process itself. As the process requires only bright, stable spectral lines, the beamline was operated for HGHG in which the first laser is turned off. The second EEHG laser and chicane combination were then tuned to provide HGHG radiation in the wavelength region of interest. In practice, because of the HGHG optimization curves shown in figure 2.24, this involved setting the laser power and corresponding chicane strength as high as possible in order to generate the brightest harmonics. As the seed laser wavelength was known with precision, the radiation was at well-known harmonics λ/h which defined wavelength posts throughout the entire range of interest. While performing the calibration measurements, it was essential to monitor closely the electron beam chirp, as a chirped beam will produce bunching at slightly off-center laser harmonics [101]. While a single set of harmonics from one seed laser is sufficient for a wavelength calibration, this choice can be checked against an alternate set of harmonics. The second EEHG laser was normally passed through an OPA and downconverted from 800 nm to 2400 nm, but this stage was bypassed to directly interact with 800 nm radiation in the second modulator. This provided a second set of harmonics with which to compare, and spectrometer images as well as the resultant wavelength calibration are shown in figure 3.13. In these images, individual harmonics are easily identifiable and can be correlated to yield an accurate wavelength calibration. Also of note in these images is the presence of multiple undulator harmonics, centered at $\approx$ 39 nm and $\approx$ 32 nm as the 4th and 5th undulator harmonics of the electron beam at $E \approx 162$ MeV. The characteristic boomerang shape is due to the off-axis emission at longer wavelengths from equation 2.43.

In practice, when the first calibrations were being performed, it was quickly realized that small misalignments and uncertainties in the grating geometry (distance to focal plane, angle, etc.) conspired to give an uncertainty via equation 3.2 of $\approx$ 1 nm. Given only the set of harmonics from the 2400 nm laser at the bottom of figure 3.13, there was considerable ambiguity between harmonics h and $h + 1$. To take a step beyond this calibration, then, we admitted that the alignment of the spectrometer and grating assembly was not perfect. The previously fixed parameters of equation 3.2 then become parameters to be fit, although they were banded by narrow margins around the measured values.

The parameters L, α, and the angle of MCP with respect to the focal plane were allowed to vary in their narrow ranges. Furthermore, given a set of k harmonic lines, there would be several possible orderings based on the central pointing wavelength, $[h, \cdots, h + k]$, $[h + 1, \cdots, h + k + 1]$, etc. The known pointing resolution of $\approx$ 1 nm limited this list to at most four different orderings for the harmonics, and the dispersion in terms of pixels/mm could be computed to sufficient accuracy. This set of parameters was scanned until a best fit solution to the entire range of harmonics was found. The results of this procedure were usually only small deviations from the design parameters of the spectrometer, as well as an unambiguous determination of all the harmonic numbers in the image. After performing the procedure on several datasets, a set of consistent nominal offsets was converged on, and with these modified values, equation 3.2 was used for all future determination of wavelength axes.

3.3 Characterization of the Electron Beam

In order to successfully attain high harmonics using the EEHG technique, it was necessary to characterize and optimize the properties of the electron beam. In particular, the experiment required many of the same electron beam optimizations which occur in a full-scale lasing FEL beamline. It is important to note however that the setup at NLCTA lacks the high electron beam current and long radiating undulator to have appreciable gain. Nevertheless, a slightly compressed electron beam is desirable to increase the uniformity of the EEHG modulation across the beam, thus minimizing effects of seed laser chirp well as the quadratic energy curvature which arises from the relatively short wavelength X-band accelerating cavities. Additionally, due to the short radiating length and difficulty in detecting EUV photons, a high beam charge is desirable for purely diagnostic reasons. Finally, a low electron beam emittance is desirable to reduce longitudinal smearing of the EEHG bunching and allow tight focusing of the beam into both the modulating and radiating undulators.

The electron beam is born at the copper cathode inside the BNL/SLAC/UCLA 1.6 cell S-band gun after it is illuminated by the 266nm tripled Ti:Sa laser pulse. The structure of the S-band gun is shown in figure 3.14. Unfortunately, due to limited diagnostics, a complete characterization of the beam from the NLCTA was not available. Simulations, matched to measured data where possible, provided some insight into the realistic electron beam quality. The simulations of the electron gun were performed with the space charge tracking code ASTRA [103].

In our setup, the frequency tripled laser pulse had an FWHM duration of $\sim$ 1ps. The maximum RF voltage in the gun produces a beam energy, measured by the immediate downstream bend magnet, of $\sim$ 5 MeV, which yielded a maximum accelerating voltage on the cathode of $\sim$ 110 MV/m. In order to provide an accurate simulation, the Schottky effect needed to be accurately accounted for. The well-known Schottky effect states that the work function of a metal is lowered in the presence of

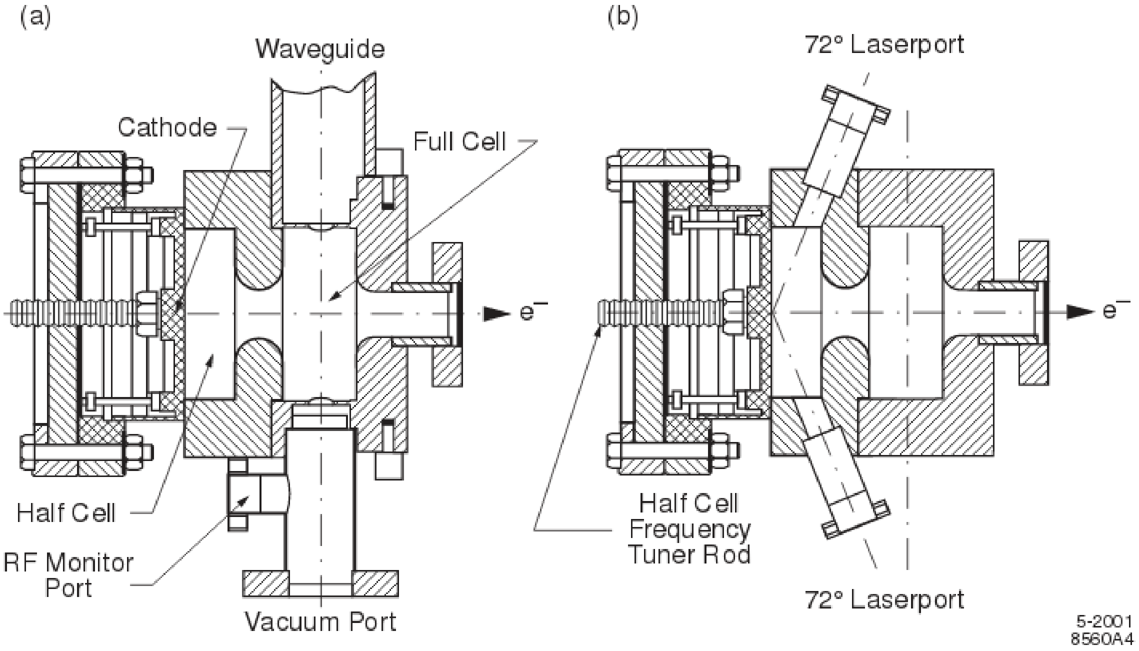


Figure 3.14: A schematic representation of the electron gun in place at NLCTA. Orientation b) shows the off-axis laser injection ports. Diagram from the LCLS CDR[102].

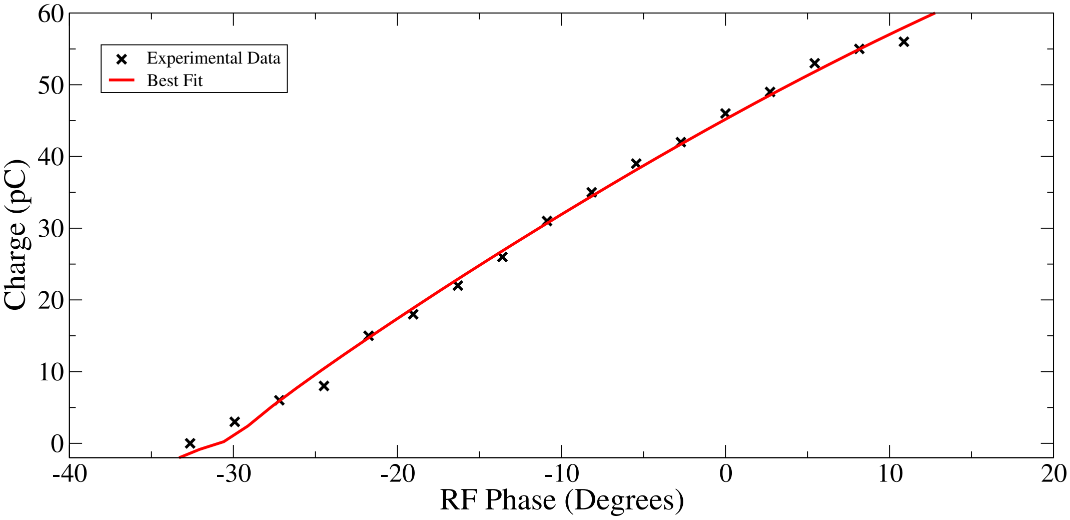


Figure 3.15: Schottky scan data from the NLCTA electron gun. The RF phase is relative, and the best fit is performed using a linear and square root model employed by ASTRA.

a strong electric field. Therefore, as the phase of the RF is changed relative to the timing of the laser pulse, the amount of charge released changes as well. In ASTRA, this effect is modeled as an enhancement in the released charge which varies linearly or with the square root of electric field at the cathode,

$$Q = Q_1 E(\phi) + Q_{1/2}\sqrt{E(\phi)}, \tag{3.4}$$

where $E(\phi)$ is the electric field at the cathode, and ϕ the relative phase of the RF field. This is an entirely empirical description of the Schottky effect, as ASTRA does not model the actual photoemission process. By scanning the RF phase and measuring the charge, the effective Schottky coefficients $Q_1, Q_{1/2}$ were measured as shown in figure 3.15. The RF phase was relative to simulation of maximum energy at 60°. The charge was measured with an insertable metallic Faraday cup directly downstream of the S-band gun. With a maximum field of 110 MV/m, the resultant Schottky coefficients were $Q_1 = 0.69\text{pC}/\left(\text{MV/m}\right)$, and $Q_{1/2} = 0.95\text{pC}/\sqrt{\text{MV/m}}$, which provided a reasonable fit over the limited RF phase in discussion. These Schottky parameters were used for all subsequent ASTRA simulations.

In order to both theoretically optimize and understand the limitations of the injector system, a genetic optimization was performed using the multi-variate optimization routine NSGA-II [104]. The genetic algorithm creates a 'population' of individual electron injectors, each described by various parameters: the RF phase, accelerating voltage, first solenoid strength, accelerating structure voltage, and so forth. Upon running ASTRA to simulate each individual configuration in the population, one defines various merit criteria in order to compare how each injector performs.

For this study, the merit criteria were, as discussed above, peak current and emittance. Since in general increasing peak current comes with an increase in emittance, this is a multi-variate optimization problem which in general has no unique 'best' solution. The NSGA-II algorithm, therefore, selects only those individuals which are *undominated*: Those solutions which do not perform strictly worse than any other individual in all merit criterion. These solutions are then 'bred' together, merging their parameters and including random variation 'mutations' in order to avoid the

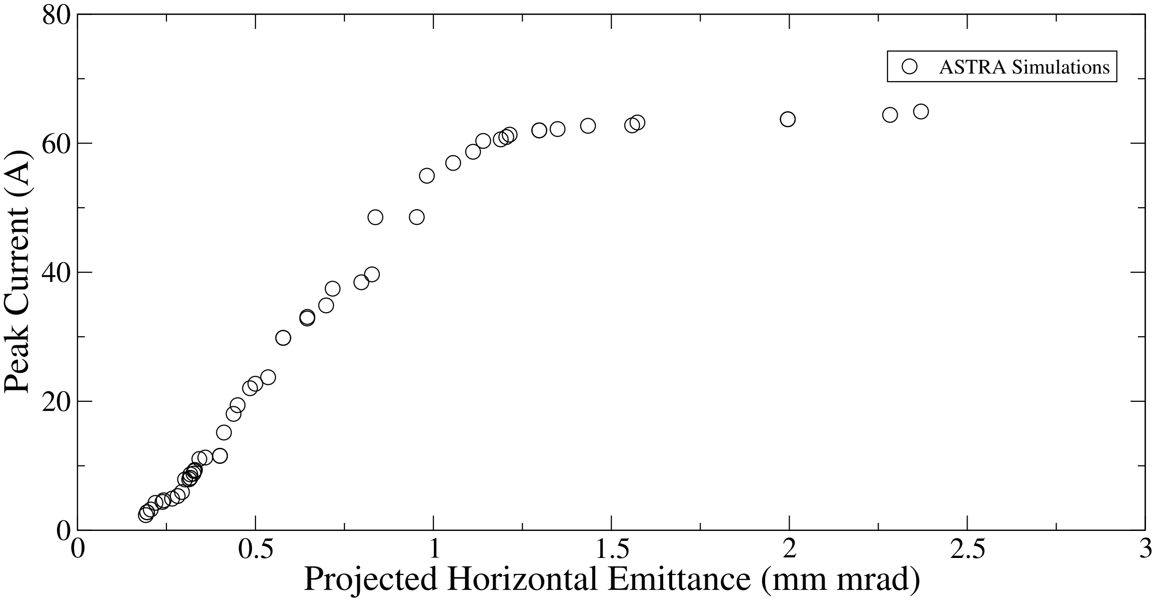


Figure 3.16: The simulated Pareto front for the NLCTA injector in terms of current and emittance.

population being stuck in a local minima of performance. After many generations, one arrives at a stable set of solutions which are no worse than any other, which is known as the Pareto front. Given this front, a human must then decide which tradeoffs are acceptable in order to find the preferred working point.

This procedure was performed for the NLCTA injector setup through the first accelerating structure. After the first structure, the beam had energy $\sim$ 60 MeV, and the detailed space charge simulation of ASTRA was no longer necessary. For this simulation, the following parameters were allowed to vary: S-band gun voltage and phase, X-band structure voltage and phase, solenoid B-field, and the total bunch charge. As previously mentioned, ASTRA does not simulate the actual photoemission process, so altering the charge parameter was a proxy for changing the laser power. For the electrons coming off the cathode, the spot size was held constant at 1 mm, the laser had 1 ps FWHM length, and the transverse momenta off the cathode were drawn from a Fermi-Dirac distribution. Of note is the fact that these transverse momenta were not scaled to account for the Schottky effect, although this played only a minor role in the total emittance.

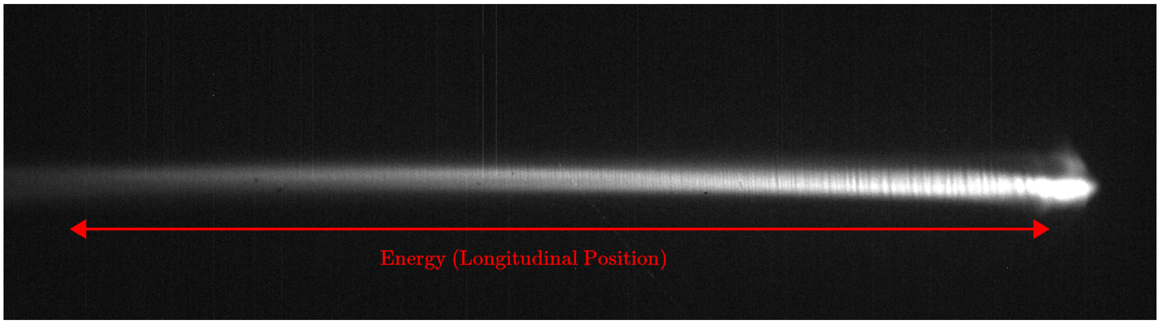


Figure 3.17: An image at the downstream energy spectrometer, where energy is projected onto the horizontal axis. Due to the induced energy chirp in X3, this axis is also longitudinal position within the bunch. The bright density structures are the HGHG-induced density bunches separated at the laser wavelength of 2400 nm.

The results of this genetic optimization procedure are shown in figure 3.16. The peak current was derived from the rms current: $I_{\mathrm{rms}} = Q_{\mathrm{total}}/\sigma_t$, which for a Gaussian bunch this differs from the peak current by a factor of $1/\sqrt{2\pi}$. In the presence of a non-Gaussian temporal profile, this conversion is off, but nevertheless gives a meaningful metric for the optimization. We observe that for the simulated configuration, it is difficult to push the peak current much higher than $\sim$ 65 A. In all these optimal configurations, the extracted charge was $\sim$ 150 pC. From these optimal ASTRA simulations, one can additionally extract the difficult to measure slice energy spread. In most optimal configurations the slice energy spread was calculated to be $1 \sim 2$ keV over the core of the electron beam.

The electron beam emittance in the low-energy (60 MeV) section of the beamline was also measured using the quadrupole scan technique. In this technique, the strength of an upstream quadrupole was varied and the transverse distribution of the electron beam monitored on a downstream OTR screen. The projected distribution at the screen along either x or y is quadratic as a function of the quadrupole focusing strength. By scanning the quadrupole strength and fitting the measured spot sizes to a parabola, the emittance was computed analytically [105]. This measurement was performed, and optimal configurations were found to yield an emittance of ≈ 2 µm, although the measured value depended on the extracted charge and emittance compensation solenoid settings which were varied across the separate EEHG-related

Quantity	Value	Units	Measured / Simulated
Maximum Energy	$162 \sim 192$	MeV	Measured
Total Charge	$10 \sim 150$	pC	Measured
FWHM Bunch Length	≈ 1	ps	Measured
Transverse Emittance (Normalized, Projected)	≈ 2	µm	Measured
Slice Energy Spread	$1 \sim 2$	keV	Simulated

Table 3.1: A summary of the relevant electron beam parameters at NLCTA.

experiments.

The electron bunch length was approximately measured at the end of the beamline using the cavity X3 as an RF streaker. In this mode, X3 was operated far off crest to impart a linear chirp to the electron beam energy, which was then converted into a horizontal position offset via the downstream energy spectrometer dipole magnet. By utilizing the echo setup to generate strong HGHG type bunching in the chicane C2, we imprinted a known longitudinal scale onto the electron bunch to use as a ruler. An image of the electron beam at the energy spectrometer is shown in figure 3.17, where we clearly identify the bunching peaks. This image was taken with gun settings optimized for a 50 pC bunch charge – a common working point for the echo experiment due to the tradeoff between emittance and total radiation generated.

Using the separation between density peaks as a known bunch coordinate distance of 2400 nm, the full electron bunch length in this image computes to approximately 1.9 ± 0.2 ps. It is clear from figure 3.17 that an accounting of the tails of the electron beam distribution is limited, so care must be taken to define a FWHM bunch length. To this end, the left-most portion of the image can constitute one side of the Gaussian distribution. After integrating over the vertical dimension, we found that the FWHM of the intensity distribution was 1 ± 0.2 ps. This is in reasonable agreement with the full bunch length if we consider the bunch visible out to ≈ 2 FWHM before dropping below visible intensity. These estimates are in good agreement with the predicted FWHM bunch length of ≈ 1 ps.

A summary of the relevant electron beam parameters, both measured and simulated, is tabulated in table 3.1.

These beam parameters are intended as a theoretical basis upon which to base the EEHG optimization which will be discussed in chapter 4. They also inform which experiments are possible with the NLCTA electron beam.

For example, given this estimate of the optimal electron beam, we can imagine an experiment to measure the FEL gain of an EEHG-seeded setup at NLCTA. To do this, we consider the most favorable parameters from table 3.1, coupled with the EEHG settings which provide the greatest possibility of lasing. We consider lasing at the relatively long wavelength of 160 nm, obtained at the $n = -1$, $m = 18$ point of the 800nm and 2400nm seed lasers. For a total electron beam energy of $\sim$ 162MeV, $\epsilon_N \approx 2$ µm, $\sigma_E/E \approx 5 \times 10^{-5}$, and the rms beta function in the VISA undulator is $\beta \approx$ 60cm, and we obtain a 1D gain length $L_{G,0} \approx$ 40cm. With such a gain length, it is marginally feasible to measure gain, as the VISA undulator represents almost 5 gain lengths. However, due to the tight electron beam focusing, the M. Xie correction [45] brings this gain length to $L_G \approx$ 90cm, a value far too high to see significant gain in the only 2 meters of VISA undulator. Without an external source of compression, then, the peak current and focusing rules out the possibility of significant gain in our configuration.

The measured and simulated parameters of table 3.1 therefore established the ground rules for EEHG experiments at NLCTA. The slice energy spread, for example, was used to estimate the laser modulations necessary to achieve the desired scaled EEHG parameters $A_{1,2}$. The measured values for the bunch length provided guidance for selecting optimal EEHG parameters in the presence of significant parabolic energy structure on the electron beam. The transverse emittance measurement established a pessimistic estimate for the emittance-smearing effect discussed in section 4.1.1. The characterization was therefore essential in illuminating the path towards obtaining high harmonics with the NLCTA EEHG setup as discussed in chapter 4.

These baseline measured parameters will also be fed into the start-to-end NLCTA simulations presented in section 5.3. While the simulations of this section considered optimal configurations, the start-to-end simulations will simulate, as accurately as possible, an electron beam similar to the one which ultimately generated the radiation profiles shown in figure 4.5.

Chapter 4

EEHG Experiments at NLCTA

> An expert is a person who has found out by his own painful experience all the mistakes that one can make in a very narrow field.
>
> Niels Bohr

The primary goal of the echo program at NLCTA from 2014-2015 was to generate and measure bunching at the 75th harmonic of the (second) seed laser. In principle, this harmonic upconversion roughly marks the point at which EEHG becomes a viable single-stage seeding technique for a full-blown soft x-ray FEL facility. The goal of the 75th harmonic was ultimately achieved, although along the way several interesting and novel studies in the lower harmonic range were also performed. The 75th harmonic represents an important milestone, because it implies a radiation wavelength, when seeding with 266 nm lasers, of 3.5 nm, which is in the middle of the so-called water window for soft X-rays. This water window is the wavelength range between the K absorption edges of carbon (4.37 nm) and oxygen (2.33 nm) which would allow for in-vitro imaging and analysis of biological samples. The ability to generate fully coherent X-ray pulses in this important wavelength range from a single seeding stage would therefore greatly enhance many of these studies.

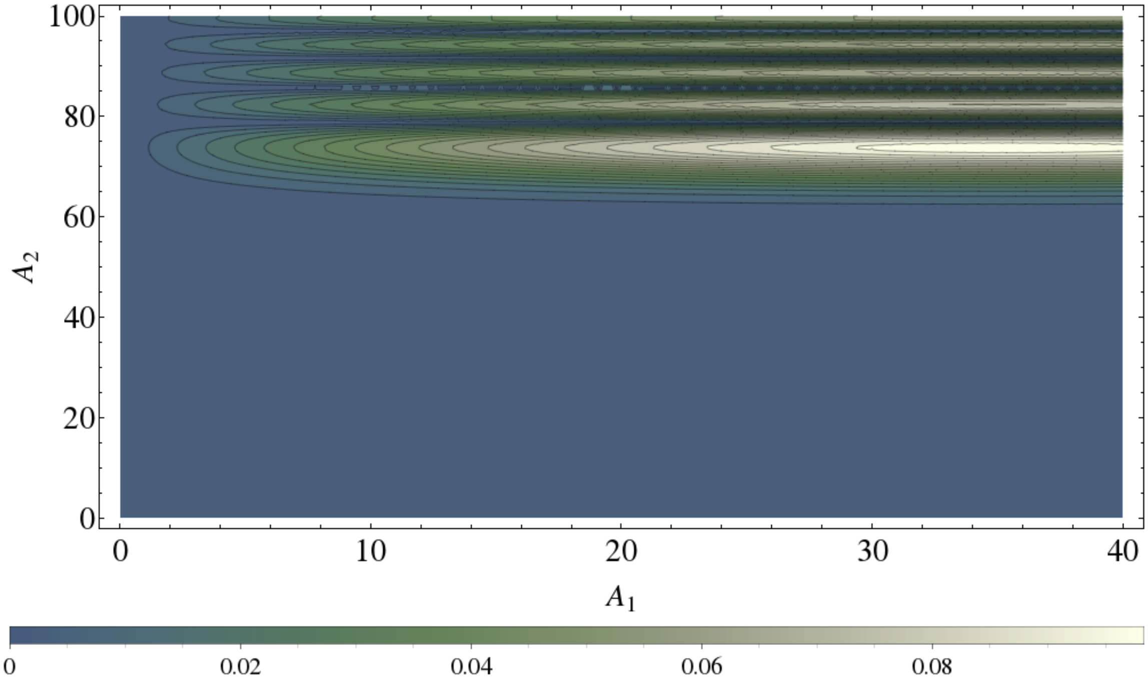


Figure 4.1: Bunching contours for Echo-60 for varying A_1, A_2 and optimal values for B_1, B_2.

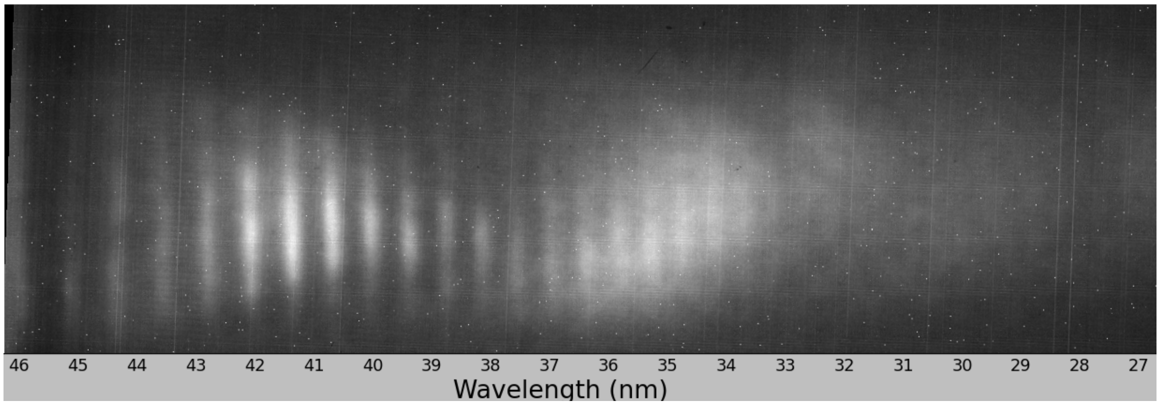
4.1 Highest Harmonics: Echo-60 and Echo-75

While gradually increasing the harmonic number and electron beam energy, the radiation wavelength around 40 nm became one of particular interest for studying EEHG. For one, 40 nm is close to both the 4th VISA harmonic with a beam energy of 162 MeV, and also the 3rd VISA harmonic at 192 MeV, which allows for efficient radiation. Second, the efficiency of the EUV spectrometer is quite good in the regime when using the low density grating (see fig 3.7). Finally, and perhaps most critically to its utility in studying EEHG, the 40 nm bunching point was at a long enough wavelength that deleterious effects did not swamp the echo effect. Nevertheless, the 40 nm bunching point at the 60th harmonic of the 2400 nm seed laser is a high enough harmonic to be novel and somewhat interesting when transferred to a full soft X-ray seeding facility.

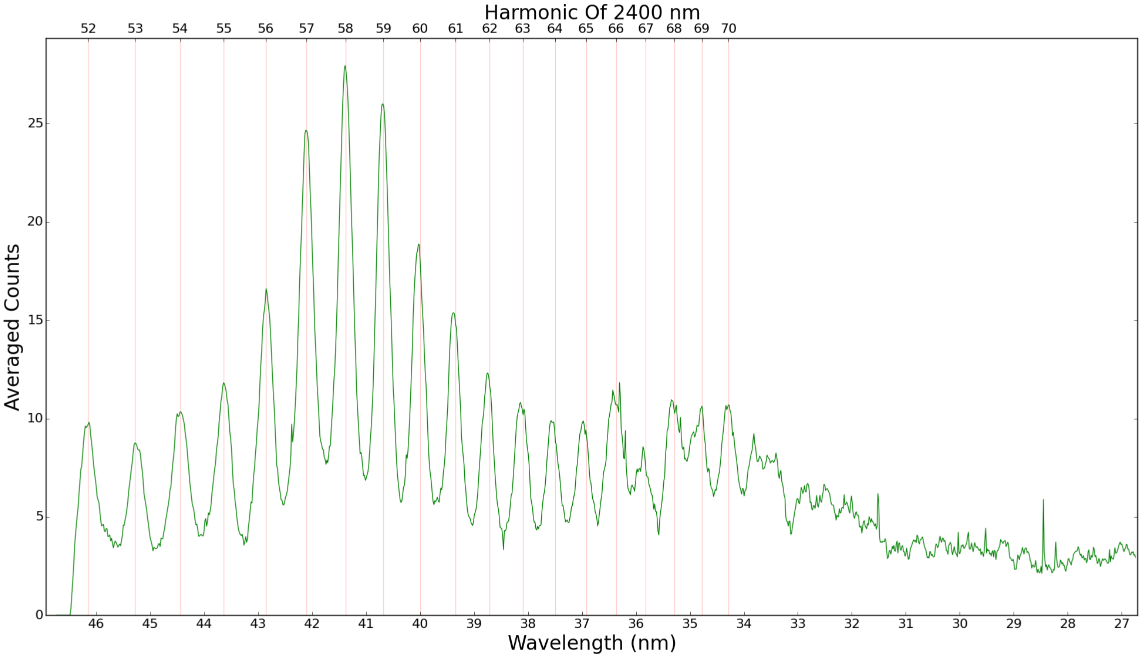
To begin with, it is useful to have an idea of what kind of laser and chicane parameters are necessary to generate bunching at a given harmonic. Seeking harmonic content at the 60th harmonic of the 2400 nm laser, we look to the approximate relation eq. 2.126, which yields $h \approx nR_{56}^{(1)}/\kappa R_{56}^{(2)}$. In almost all cases, one desires to use the largest first chicane possible in order to minimize the amount of laser power required to generate the desired harmonics. From this condition, we can set $R_{56}^{(1)} = 10.0$ mm, a relatively large but not maximal value, yielding a $B_1 \approx 0.65$ for $\sigma_E/E \approx 8 \times 10^{-6}$. As $\kappa = 1/3$ for the combination of $\lambda_1 = 800$ nm , $\lambda_2 = 2400$ nm, this means $B_2 \approx 0.65/20 \approx 0.032$, with the stipulation that the minimum attainable B_2 due to scraping with the laser ejection mirror has $R_{56}^{(2),\mathrm{min}} \approx 500$ µm, or $B_2^{\mathrm{min}} \approx 0.03$. Under these approximate restrictions, we can use the analytical theory (eq. 2.124) to optimize the laser amplitudes A_1, A_2 for bunching at the 60th harmonic, which is shown in figure 4.1.

It is clear that due to the (relatively) underpowered first chicane, the laser amplitudes must be increased substantially (compared to the slice energy spread) in order to generate significant bunching. In particular, a relative energy modulation in the second laser of $A_2 \approx 70$ is required to generate any significant bunching at these high harmonics. Although large in relative terms, since the slice energy spread is only 1 keV this represents a total energy deviation of only $\Delta E/E \approx 3 \times 10^{-4}$. This is approximately the same range (in terms of $\Delta E/E$) in which a full EEHG-seeded XFEL facility would operate, making the experiments at NLCTA a scaled testbed.

Furthermore, we recall from section 2.4.2 the relative insensitivity of the echo process to the first laser modulation A_1 after a certain point ($A_1 \approx 35$ here). The first laser modulation amplitude was generally not tuned very far beyond this point. Based on this information, the chicane values were set to their optimal values and the laser power adjusted around the neighborhood defined by $(A_1, A_2) \approx (40, 80)$. Due to the relatively large first chicane strength of $R_{56}^{(1)} = 10.0$ mm, changes to $R_{56}^{(1)}$ generate significant electron orbit changes which influence the rest of the echo line. For this reason, the first chicane is set at the beginning of the experiment and not changed during the tuning procedure. By contrast, since the value of $R_{56}^{(2)}$ is smaller, and it sits closer to the radiating undulator, the second chicane was tuned to optimize the bunching.



(a) The captured image of the MCP showing the Echo-60 signal in the vicinity of 40 nm. The vertical dimension shows the parabolic off-axis emission spectrum from the undulator. The two 4th and 5th undulator spontaneous emission spectra are clearly visible near 40 nm and 32 nm. This image is the sum of 200 consecutive shots, taken over a 20 second interval taken on 05-20-2015.



(b) The projected Echo signal with the averaged spontaneous radiation subtracted, and harmonic lines shown. Echo signals are visible out to approximately the 70th harmonic, although at low intensity.

Figure 4.2: The EEHG signal optimized for $\lambda \approx 40$nm in the vicinity of the 60th harmonic. Plot (a) shows the raw data from the MCP, while plot (b) shows the projected signal and correspondence with harmonics of the 2400 nm seed laser.

For this study (prior to the installation of the SLED cavity), the electron beam energy was set to 162.5 MeV, such that the 4th resonance of the VISA undulator sits at approximately 40 nm to efficiently radiate the 60th harmonic. The electron bunch charge was set to the relatively high value of 200 pC, which was possible due to the loosened emittance requirements relative to the 75th harmonic. The final optimal value for the second chicane was $R_{56}^{(2)} = 0.55$ mm, while $R_{56}^{(1)} = 10$ mm. The resultant spectrum captured on the EUV spectrometer is shown in figure 4.2. The images clearly show the individually resolved echo harmonics sitting atop the undulator spontaneous radiation pattern. The individual echo harmonics are clearly distinguished from 52 down to perhaps 70.

The effect of averaging over many consecutive images broadens the bandwidth of the individual echo harmonics. From the single images, the computed FWHM bandwidth of the 59th harmonic is $\Delta\lambda_{\text{EXP}} \approx 0.15$ nm, which is equal to the resolution of the EUV spectrometer. Simulations show the true echo-induced bunching bandwidth to be as low as $\Delta\lambda_{\text{TH}} \approx 0.01$ nm, which is limited by the effective electron bunch length and quadratic electron energy curvature.

The next notable feature is the amplitude of the echo harmonics relative to the spontaneous radiation. By comparing the brightest harmonics to the spontaneous radiation, we see that the brightness is increased by perhaps a factor of three. In light of the fact that the electromagnetic power relative to the spontaneous should scale as $P \sim bN_e$, the only conclusions are that either the induced bunching is incredibly low or only a small number of particles are participating in the radiation process. The inefficiencies of the radiation process which are consistently observed in these experimental results at NLCTA will be discussed in section 4.1.1.

The laser modulation amplitudes were also measured at the downstream energy spectrometer to make contact between the analytic theory and experimental results. The calculation of the bunching factor based on equation 2.124 using the consistent modulation amplitudes $A_1 = 40$ and $A_2 = 80$ is shown in figure 4.3. Note that the bunching is quite low, but still observable at the 1% level. Furthermore, the Bessel function modulation is consistent with the modulation in intensity from the observed data in figure 4.2, when accounting additionally for the modulation due to

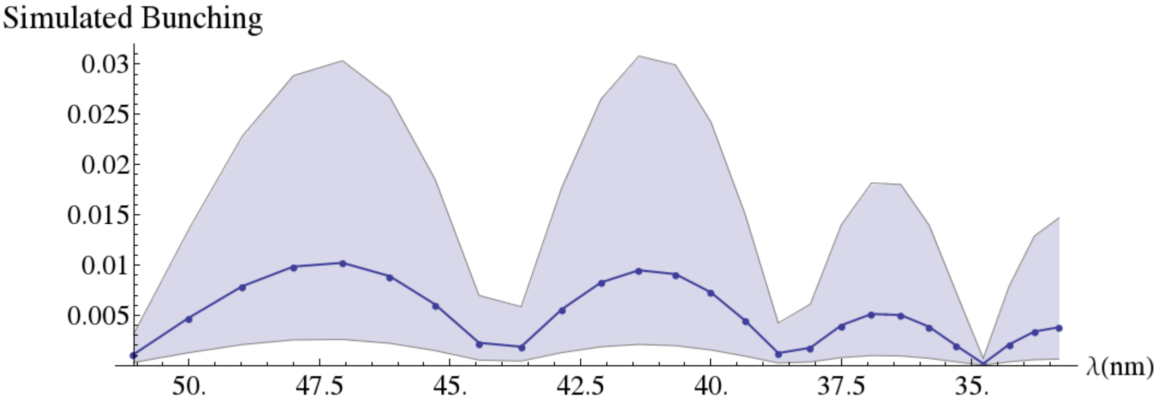


Figure 4.3: The predicted bunching in the vicinity of Echo 60 with $A_1 = 40$, $A_2 = 80$, $R_{56}^{(1)} = 10\text{mm}$, $R_{56}^{(2)} = 0.55\text{mm}$, and $n = -1$. The bands represent fluctuation of 5% in the value of A_2, which is characteristic of the performance of the OPA.

the undulator harmonics.

This demonstration of Echo-60 is not ideal, however. The bunching is quite low, and the variation in the second laser intensity implies that only select shots produce a significant signal. To analyze the stability of the signal, we plot the individual shot spectra alongside one another in figure 4.4. While the amplitude varies due to the fluctuations in intensity of the second laser modulation (on the order of 10%), the central wavelength remains unchanged in light of laser and electron beam imperfections. Finally, we note that although this is a genuine echo signal, there also exists a strong HGHG signal in this wavelength region due to the tuning of the 2nd chicane and laser. This can be observed by simply blocking the first laser, thus converting the setup into one of HGHG signal only.

In order to access the highest harmonics available to the NLCTA setup, the beam energy was increased to near its maximum of 191 MeV. This beam energy was chosen over the perhaps more practical choice of 182 MeV, where the 4th undulator harmonic would be at 32 nm. In practice, it was found that the difficulty of tuning harmonics in the 182 MeV configuration outweighed the benefits of radiating directly on an undulator harmonic. As the earlier Echo-60 studies had shown, it was still possible to get significant off-axis emission slightly above the undulator harmonic wavelength. Therefore, the highest beam energy was used to maximize the radiation power and

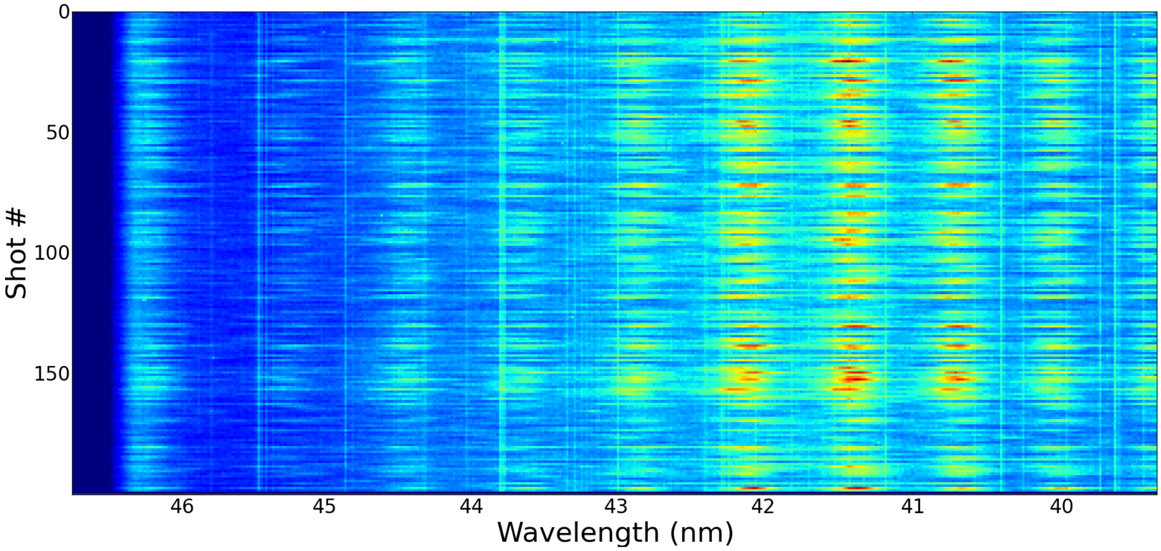


Figure 4.4: A stability plot of the Echo-60 signal over the course of 200 shots (≈ 20 seconds).

minimize deleterious effects in the undulator (discussed in section 4.1.1).

The EEHG configuration at this high harmonic suffers from an even worse form of the tuning problem found in the Echo-60 configuration exemplified by figure 4.1. It is necessary to have a large A_2 in order to generate significant harmonic bunching. Therefore, $R_{56}^{(1)}$ was further increased to near its maximum of 12.5 mm, and $R_{56}^{(2)}$ set to ≈ 0.51 mm. The resulting high harmonic spectrum is shown in figure 4.5. The observed modulation amplitudes are $A_1 = 50$ and $A_2 = 100$, which are consistent with the expected EEHG spectrum of figure 4.5. Furthermore, when the first laser is blocked to simulate the analogous HGHG configuration, no significant radiation is observed below 34 nm, confirming that the signal reported in figure 4.5 is indeed solely due to the unique EEHG phase space manipulation.

The stability of the EEHG signal can also be analyzed in the same way as before, and the stability plot over 224 shots is shown in figure 4.6. The stability of the 75th harmonic and beyond is similar in character to the stability of the highest harmonics shown in figure 4.4. The amplitude fluctuations are due both to fluctuations in the second laser, as well as charge fluctuations coming off the electron gun (which affect all harmonics as well as the spontaneous background). The predicted bunching factor,

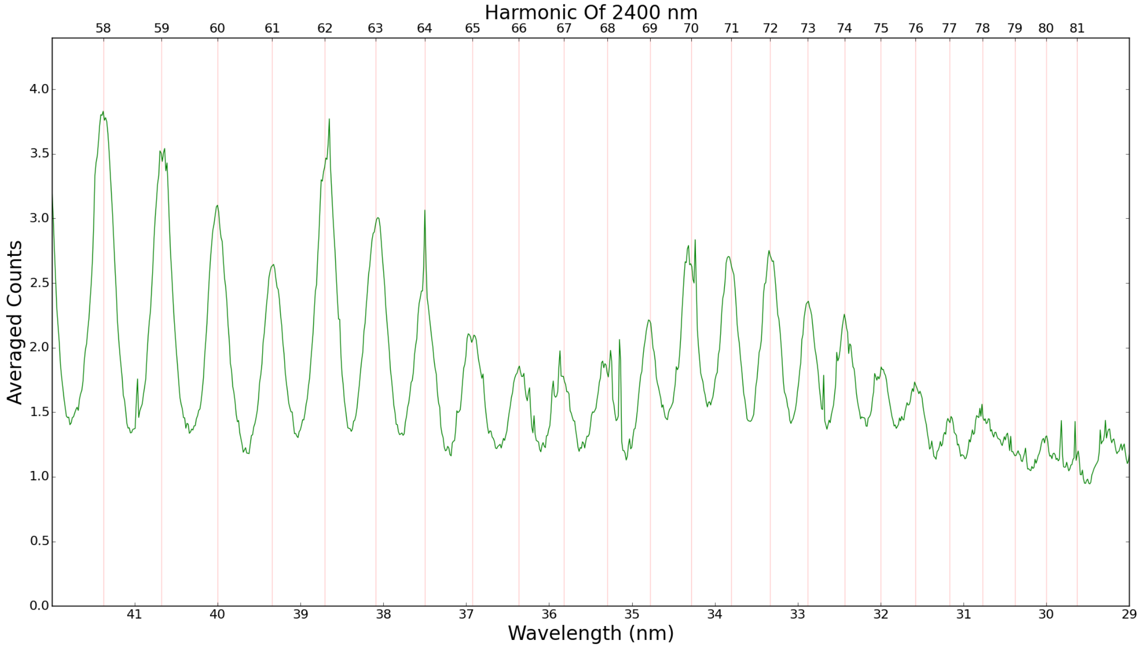


Figure 4.5: The projected Echo-75 signal with the averaged spontaneous background subtracted off. Individual echo harmonics down to the 77th are clearly distinguished from background. This spectrum is averaged over 220 shots, and was taken on 10-19-2015.

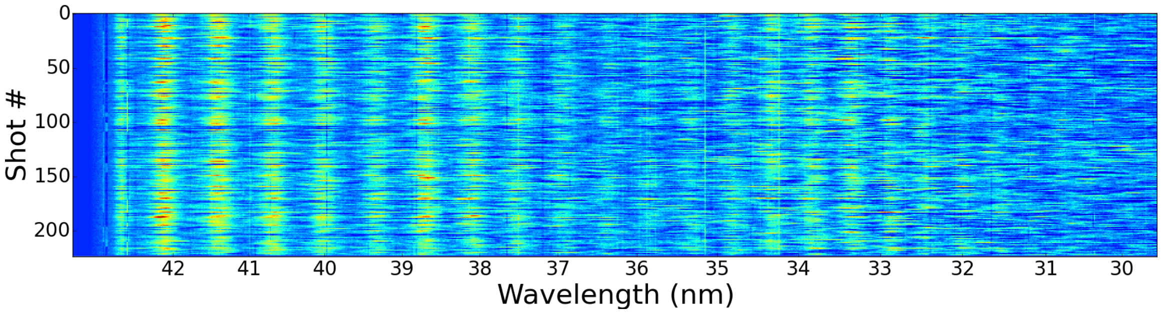


Figure 4.6: Stability plot in the vicinity of the Echo-75 signal over the course of 224 shots (≈ 22 seconds).

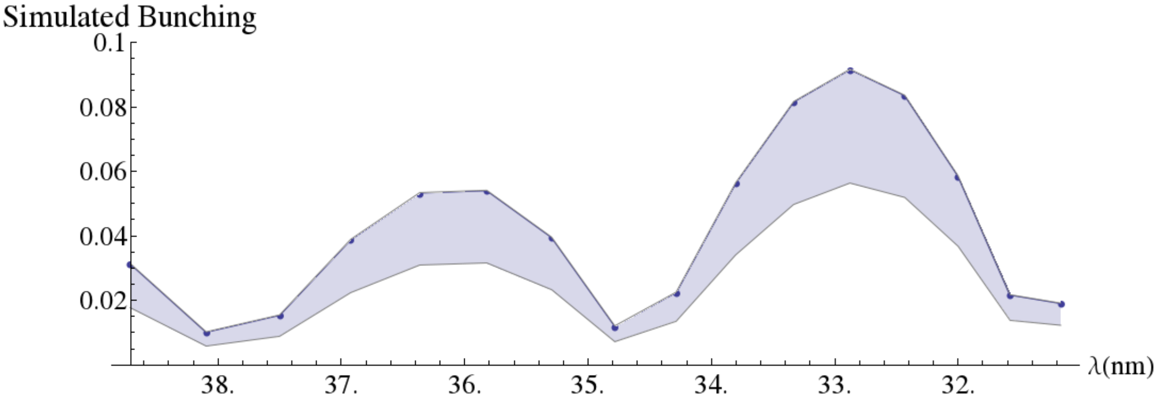


Figure 4.7: Theoretical bunching factor in the vicinity of Echo-75 with $A_1 = 50$, $A_2 = 102$, $R_{56}^{(1)} = 12.5\text{mm}$, $R_{56}^{(2)} = 0.49\text{mm}$. The bands represent a fluctuation of -5% in the 2nd laser modulation intensity from the maximum value of $A_2 = 102$.

as well as the fluctuations due to the 2nd laser intensity ($\approx 10\%$), are shown in figure 4.7. Once again, the general shape of the Bessel function envelope and fluctuations due to laser intensity are consistent with the observed spectra of figure 4.6.

The observations of radiation up to the 75th harmonic confirm that the EEHG scheme can be used to produce very high harmonic bunching of the initial seed laser. The 75th harmonic, in particular, is an important milestone in that it suggests that the EEHG technique may be straightforwardly applied to an XFEL in which a 266 nm seed laser is up-converted to bunching in the water window ($2 \sim 4$ nm range). Furthermore, the establishment of this single-stage EEHG high harmonic upconversion may suggest more ambitious, multi-stage EEHG setups which could extend the technique all the way into the hard x-ray regime.

4.1.1 Limitations on High Harmonics

Although the goal of attaining high harmonics up to the 75th was ultimately achieved by the programme at NLCTA, the resulting radiation spectra show only a coherent enhancement factor over the background spontaneous radiation of 2-10x for harmonics in the 60-75 range. This suggests that the net bunching factor, when the entire electron beam and radiation path is considered, is quite small, and has been deteriorated by one or more various effects. We now examine these effects in turn.

One possibility is that the bunching is being degraded due to incoherent synchrotron radiation (ISR) in the echo chicanes. The incoherent synchrotron radiation effect will be discussed more in chapter 6, but briefly is due to the quantum nature of photon emission as an electron traverses the bend magnet. Upon traversing a magnetic field of strength B and length L, the electron bunch will develop an rms energy variation of [17],

$$\sigma_E^{ISR} = mc^2 \sqrt{\frac{55}{24\sqrt{3}} \frac{r_e^2}{\alpha} \gamma^4 \left(\frac{eB}{mc}\right)^3 L}, \tag{4.1}$$

where α is the fine structure constant and r_e the classical electron radius. A straightforward estimation for the first chicane, which has larger R_{56} and hence B, produces an energy fluctuation $\sigma_E = \mathcal{O}(1 \text{ eV})$. This energy fluctuation is far too small to produce any meaningful bunching degradation, since the slice energy spread is approximately 1 keV.

A second, purely practical restriction on the harmonics reached relates to leaked dispersion from the EEHG chicanes. Consider the matrix element R_{51}, which relates the transverse size of the beam to an associated longitudinal slippage. For chicane constructed with four rectangular dipole magnets, R_{51} is identically zero. However, small errors in the magnetic field, positioning of magnets, or angles on the pole faces can contribute to a nonzero net R_{51} value. Consider a variation in the angle produced by the first bend magnet in the chicane (or equivalently, a total field error) of $\delta\theta$. The induced R_{51} is then,

$$R_{51} \approx -2 \frac{\delta\theta}{1 + \cos(\theta)}. \tag{4.2}$$

Other sources of errors similarly give rise to nonzero matrix elements. We should require that, in order to preserve bunching at wavelength λ_r, that there is no rms longitudinal motion Δs on a scale,

$$\Delta s \lesssim \frac{\lambda_r}{2\pi}. \tag{4.3}$$

For an electron beam with $\sigma_x \approx 200$ µm, and a target radiation wavelength of 32 nm, we therefore require $R_{51} \lesssim 10^{-5}$. In terms of equation 4.2, this implies a field variability of $\delta\theta \lesssim 10^{-5}$, which is a relative variability of $\delta\theta/\theta \lesssim 10^{-4}$. Already we see that the practical requirement of closing the dispersion can be difficult to obtain. This is one of the prime reasons for the installation of the quadrupole trim magnets in each chicane, through which any excess dispersion in the chicanes can be tuned away. Nevertheless, this tuning procedure can never be perfect, and we see the possibility for a small amount of leaked dispersion to completely overwhelm the bunching at high harmonics.

Another, far more insidious source of degradation is the smearing effect caused by the emittance of the electron beam propagating through the focusing lattice [106]. Although the ensemble of particles is well described by an envelope with a given set of Twiss parameters (see section 2.1.4), each individual particle still moves on its own ellipse in phase space. As a result of this finite size in phase space, each electron takes a slightly different (transverse) trajectory in traveling from beamline position z_1 to z_2. This spread in trajectories can be characterized as either a differential longitudinal velocity, or a differential path length.

In the longitudinal velocity description, those particles on larger betatron orbits appear to slip backwards relative to the reference particle due to a decreased (relativistic) $\bar{\beta}$ (in the same sense as equation 2.38). This approach is taken in the standard description of FEL theory including focusing [37]. The phase evolution equation 2.71 gains an extra term to reflect this fact,

$$\frac{d\theta}{dz} = 2k_u\eta - \frac{k_r}{\langle\beta\rangle}\left(\mathcal{W}_x + \mathcal{W}_y\right), \tag{4.4}$$

where $\mathcal{W}_{x,y}$ is the Courant-Snyder invariant for an individual particle (equation 2.22) in either the x or y dimension, and here $\langle\beta\rangle$ is the averaged Twiss beta. When considering the average across all particles, these Courant-Snyder invariants become emittances, $\mathcal{W}_{x,y} \to \epsilon_{x,y}$, and we can compute an rms drift rate due to the betatron motion. This approach leads to a longitudinal drift in real space (Δs) after a distance L of,

$$\Delta s \approx \frac{\sqrt{2}\epsilon_n L}{\gamma\langle\beta\rangle}, \tag{4.5}$$

where we have assumed equal horizontal and vertical emittances and used the normalized emittance ϵ_n. This is the standard result and is approximately correct for the types of FODO lattices employed in large-scale FEL facilities. In particular, the derivation of equation 4.4 assumes that $\alpha = \pm 1$ in the lattice (see Ref. [37] for a full derivation).

Another way to consider this effect, outside the embedded context of the FEL equations implicit in equation 4.4, is to consider the path length difference for particles on different orbits [107]. The difference in path length from the reference orbit can be computed as,

$$\Delta s = \int_0^L \left[\sqrt{\left(1 + \frac{x}{\rho(z)}\right)^2 + x'^2 + y'^2} - 1 \right] dz. \tag{4.6}$$

We consider a bend-free region composed only of quadrupoles and drifts, such that $\rho \to \infty$. The only contributions then are those due to the angle differences in equation 4.6. Assuming that these angles are small, we can expand the square root and arrive at separate, uncorrelated terms due to the x and y motion. For the y dimension, this equation is,

$$\Delta s = \int_0^L \frac{1}{2} y'^2 dz. \tag{4.7}$$

In terms of the (single-particle) Twiss parameters, the angle can be written as,

$$y'(z) = \sqrt{\frac{2\mathcal{W}_y}{\beta(z)}} \left(\sin(\psi(z)) + \alpha \cos(\psi(z))\right). \tag{4.8}$$

Inserting into equation 4.7 and averaging over the particles, we have,

$$\Delta s = \langle \mathcal{W}_y \int_0^L \frac{1}{2} \left(\frac{1+\alpha^2}{\beta}\right) dz \rangle = \frac{1}{2} \epsilon_y L \langle \gamma_T \rangle. \tag{4.9}$$

Note that due to confusing notation, the $\langle \cdots \rangle$ on the left hand side of this equation represent an ensemble average over the particles, while on the right it is the averaged γ_T across the lattice of length L. Considering now both x and y, and switching to normalized emittance, we finally have,

$$\Delta s = \frac{\sqrt{2}\epsilon_n L}{2\gamma}\langle\gamma_T\rangle. \tag{4.10}$$

From this point we observe that setting $\alpha = \pm 1$ collapses the result to the standard FEL result of equation 4.5. However, in the more general case, this slippage is dependent on the detailed construction of the lattice. The transport inside the VISA undulator is dominated by the strong focusing lattice described in section 3.2, which has a matched $\langle\beta\rangle \approx 60$ cm, but a large α due to the short-period nature of the lattice. The result is that $\langle\gamma_T\rangle \approx 5\ \mathrm{m}^{-1}$ for this section of the lattice. For the parameters of the 2 m VISA at 192 MeV, and taking $\epsilon_n = 1.7$ µm, this gives a resultant slippage of $\Delta s \approx 32$ nm.

This result indicates that the strong focusing in the VISA at 192 MeV will completely smear out any bunching at the high harmonics of interest by the end. This represents the inherent difficulty of transporting a finely bunched beam through a tight focusing lattice, and is thought to be the prime source of diminished bunching observed in figures 4.2 and 4.5. We note that mismatches into the undulator, although they tend to increase $\langle\beta\rangle$, do not decrease the effect. In fact, mismatched Twiss parameters into the undulator tend to increase the resultant $\langle\gamma_T\rangle$, thus providing even more smearing.

There are some subtleties to equation 4.10 which are critical to understanding its impact on the NLCTA results. The first of these is that the emittance ϵ is taken over the entire ensemble, and is therefore understood to be the projected emittance of the beam. In reality, individual longitudinal slices of the beam with lower emittance will experience less longitudinal smearing, and hence will have a higher preserved bunching. Simulations show (see section 5.3) that this slice emittance can be significantly lower than the projected emittance measured, and therefore certain sections of the beam will perform better than others.

The second issue is that the result in equation 4.10 is cumulative, that is, it is integrated out through the full distance L. In reality, the smearing is compounded as the electron bunch travels through the focusing lattice. This will generally result in a strongly radiating bunch for the first section of the undulator, until Δs becomes comparable to the radiation wavelength at which point the beam quickly ceases to radiate coherently. What is measured by comparing, say, the coherent radiation intensity to that of the spontaneous background, is then an effective bunching factor accounting for this fact.

This smearing is believed to be ultimately responsible for the low signal in the high harmonic regime observed at the NLCTA. Overcoming this barrier is difficult, and would require either increasing the beam energy or improving the beam emittance – the focusing lattice cannot be changed. However, with regards to a full FEL facility, we note that equation 4.10 can be considered from the start when designing a seeding section. In particular, because of the cumulative nature of the effect, such a facility would only need to protect the bunching through the first few gain lengths of the undulator, after which point the gain process in the FEL will take over and reinforce any lost bunching. In analytical form, the bunching criterion is very similar to the emittance requirement in an FEL that,

$$\epsilon < \frac{\lambda_r}{4\pi}\frac{\langle\beta\rangle}{L_{G,0}}, \tag{4.11}$$

where $L_{G,0}$ is the one-dimensional gain length. Although similar in form, the FEL criterion comes about due to the requirement that the resonant wavelength spread not be larger than ρ [37]– a separate argument from the one which leads to equation 4.10. Therefore, since this emittance requirement is typically satisfied in an FEL (with $\langle\beta\rangle$ to maximize coupling to the radiation), the bunching preservation requirement will also be satisfied. Furthermore, the general scaling of 4.10 implies that it is a larger nuisance for small scale facilities than for large ones, and the effect is not anticipated to be a limiting factor for any full-scale EEHG FEL facility.

4.2 Dispersive Tuning of the Echo Signal

One attribute which separates the EEHG signal from, for example, that due to HGHG is the ability to control the relative amplitudes of individual, nearby harmonics. As shown in section 2.4.1, the HGHG bunching amplitude is a strictly decreasing function of harmonic number, no matter what the tuning parameters are. EEHG, however, has the ability to excite small islands in the bunching space while suppressing other harmonics (see Fig. 2.29). By extension then, EEHG has the ability to tune the relative amplitudes of nearby harmonics by adjusting the configuration slightly.

The EEHG signals observed at the 60th and 75th harmonic, and shown in figures 4.2 and 4.5, both demonstrate the harmonic envelope of the EEHG bunching spectrum due to the Bessel function dependence. However, the relative amplitudes are convolved with both the emission spectrum of the undulator and the efficiency of transport, which is a function of wavelength. Therefore, while a clear envelope structure was observed, it is difficult from this data alone to adequately match to the underlying theory.

To demonstrate control over the harmonic envelope, the laser configuration was changed to bypass the OPA with the second laser. This leaves a configuration with two 800 nm lasers, in contrast to the 800/2400nm setup employed for the highest harmonic studies. Because the nonlinear OPA generally amplifies any jitters in the input laser beam, by bypassing it the overall stability of the EEHG modulations is greatly improved for precision studies. For this study, the final beam energy was decreased to 162 MeV, such that the 4th undulator harmonic sits at 40 nm, or the 20th harmonic in the 800/800nm EEHG configuration. The first and second laser modulations were established with modulation amplitudes of $\Delta E_1 = 29$ keV and $\Delta E_2 = 27$ keV, while the first chicane was set to $R_{56}^{(1)} = 10$ mm. With these settings, the optimal bunching at 40 nm is found with $R_{56}^{(2)} = 660$ µm. From this configuration, the value of $R_{56}^{(2)}$ was varied, and the spectrum in the vicinity of 40 nm measured. The resulting spectrogram is shown in figure 4.8.

For each $R_{56}^{(2)}$ value, approximately 50 separate shots were averaged to give the resulting spectrum. For these images, only a blank image with no electron beam has been subtracted off, and thus the spontaneous radiation is still present. The simulated bunching is created by numerically computing the electron coordinates using equations 2.118-2.121, rather than a full tracking simulation. This simulated bunching, using the full electron beam length, reveals the small predicted bandwidth

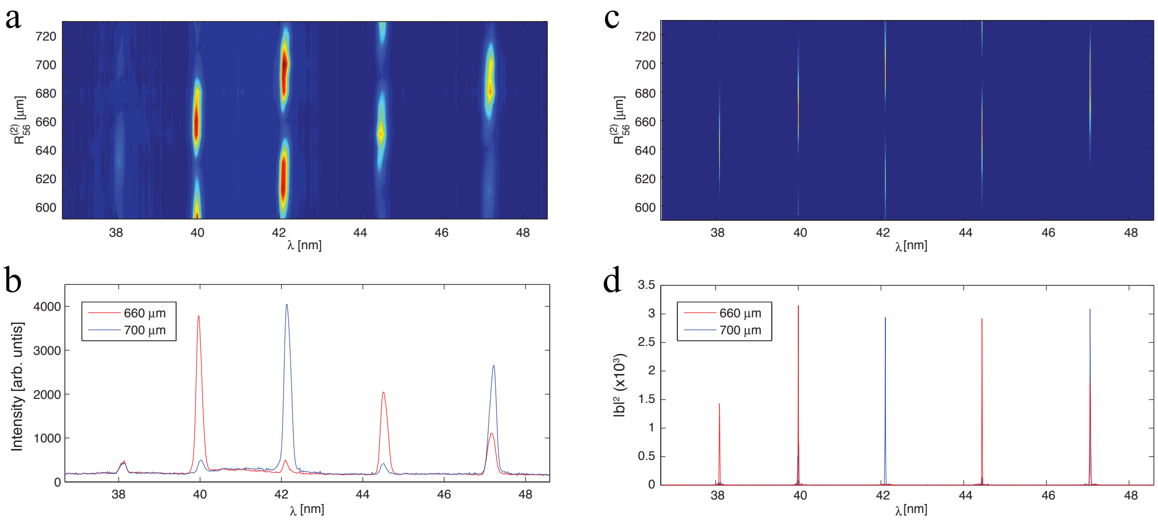


Figure 4.8: Tuning echo harmonics in the vicinity of 40 nm with the second dispersion. a) Measured individual echo harmonics of the 800 nm lasers as the second dispersion is changed. Subfigure b) shows a line-out for two specific dispersion values indicating the near extinction of adjacent harmonics due only to a small shift in the second dispersion. Plots c) and d) correspond to a) and b) but with simulated bunching data using the true bunch length, thus revealing the theoretical bandwidth of each harmonic. Figure from [108].

which is not resolved by the EUV spectrometer.

These results clearly show that the amplitude of even adjacent harmonics can be controlled by varying the second dispersion in the EEHG setup. To appreciate the magnitude of this control, we note that this effect manifests in the complete control of spectral features 2 nm apart, controlled via a 40 µm change in $R_{56}^{(2)}$, all deriving from an 800 nm laser. Although this property of EEHG has been known since the bunching spectrum of equation 2.124 was first discovered, this is the first time it has been both experimentally observed (Figs. 4.2, 4.5) and also exploited and controlled.

In the context of an FEL, this result can imply a quickly tunable echo configuration by changing only the second dispersion. As was observed in our experimental setup, changes in the first, large chicane may have a significant effect on the downstream electron orbit. Therefore, the ability to quickly re-tune to a nearby echo harmonic via only small changes in the second chicane could provide a means to quickly change the wavelength of an EEHG seeded FEL (via, for example, a variable K in the downstream undulators as well).

Since the difference in wavelength between adjacent harmonics at harmonic h scales as, $\Delta\lambda/\lambda \sim 1/h$, while the FEL amplification bandwidth $\Delta\lambda/\lambda \sim 1/\rho$, either a large ρ or h is necessary for adjacent harmonics to be within the FEL amplification bandwidth. Achieving echo harmonics higher than $100 \sim 200$ seems difficult due to many practical considerations [109], requiring an FEL parameter $\rho \gtrsim 10^{-2}$. This requirement could possibly be met in high efficiency EUV FEL configurations (in contrast to the current generation of XFELs), and would enable the precision control of multi-color pulses within the FEL bandwidth via EEHG dispersion tuning.

4.3 Multi-color Echo Effects

Having demonstrated the precise control of individual harmonics via $R_{56}^{(2)}$ tuning, we turn now to the generation of multi-color echo signals. By multi-color, we refer in general to the splitting of a single EEHG harmonic into nearby, separate peaks. These peaks, unlike the control demonstrated in section 4.2, may well exist within the FEL bandwidth, thus allowing a true multi-color seeded FEL.

One method of generating an echo bunch with multicolor components involves introducing a linear chirp to the electron beam. The linear energy chirp on the electron beam can be characterized by the parameter $h_1 \equiv \frac{1}{k_1}\frac{dp_0}{ds}|_{s=0}$, where p is the scaled momentum coordinate, k_1 is the first laser wavenumber, and s the longitudinal co-moving coordinate (these are the same conventions as in equations 2.118-2.121). It has been known for some time [101] that a linear chirp has the effect of shifting the echo harmonic a for a given echo configuration to the location,

$$a' = \frac{n + m\kappa\left(1 + h_1 B_1\right)}{1 + h_1 B_1}. \tag{4.12}$$

For $h_1 = 0$, this simply reduces to the location of the various echo harmonics, while $h_1 \neq 0$ will shift their locations. For the $n = -1$ tune, this has led to the conclusion that EEHG is less sensitive in terms of wavelength shift than HGHG. In particular, the shift in the $n = -1$ echo configuration is only half that of an HGHG beam with

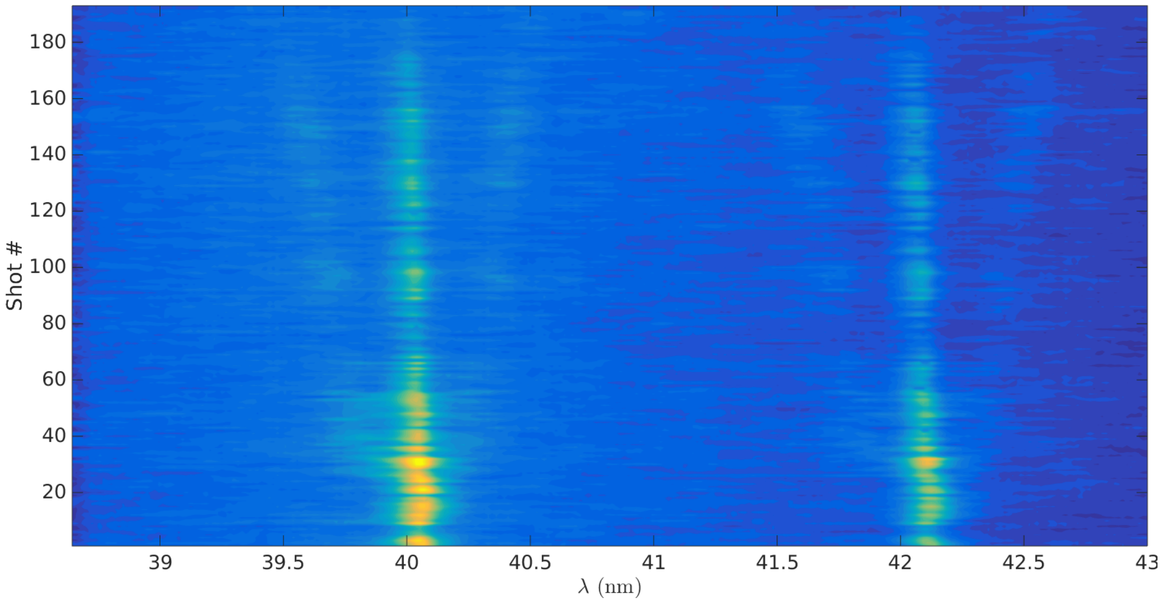


Figure 4.9: The splitting of the main echo peaks near 40 and 42 nm due to an increase in linear chirp. Clearly visible are faint sidebands with wavelength both above and below the the main peaks at 40 and 42 nm. The shot number is correlated with RF phase increase such that the difference between shot 0 and 200 is approximately 15 degrees in RF phase (for the first RF structure).

identical linear curvature. However, equation 4.12 clearly shows a dependence on the n mode of the echo configuration, and in fact, higher $|n|$ modes exhibit larger shifts for a fixed h_1. In particular, by the time $n = -3$, the shift of an EEHG and HGHG signal are comparable [76].

In an EEHG configuration with no energy structure on the electron beam, all various n-modes (with correspondingly different m values) are degenerate in wavelength. Thus, even if multiple n-modes are excited with a given EEHG configuration, they all simply stack atop one another in the resulting spectrum. In the presence of a non-zero chirp h_1 however, the degeneracy is broken as the higher $|n|$ modes move farther away from the central, unchirped value. This provides the first method for creating closely spaced, multicolor bunching using EEHG.

To test this idea at NLCTA, we again used an EEHG configuration with 800nm/800nm lasers to achieve the greatest stability and tuned the configuration for Echo-20 at 40 nm. For this setup, the dispersions were set as $R_{56}^{(1)} = 12.5$ mm and $R_{56}^{(2)} \approx 0.57$ mm, and the beam energy for was 162 MeV. Then, having established a reliable signal on the 20th harmonic, the RF phase in the first accelerating structure (X1 in figure 3.10) was adjusted by approximately 15 degrees over the course of 20 seconds, or 200 shots. The resulting spectrum in the vicinity of 40 nm is shown in figure 4.9.

The RF phase (measured as an offset from the on-crest value) of the first accelerating structure was increased monotonically with shot number in figure 4.9, but not necessarily linearly. Furthermore, the direct control of the RF phase was not calibrated to an absolute change in the arrival RF phase. Comparing the shift in the central wavelength of the $n = -1$ harmonic, however, yields general agreement with the uncalibrated RF phase that between shot 0 and 200 approximately 15 degrees of RF phase were scanned.

We observed that as the RF phase was increased, two distinct sidebands formed and moved away from the central $n = -1$ peak. Numerical simulations of this EEHG setup show that the approximate maximum for the $n = -1, m = 21$ 40 nm echo bunching is given when $A_1 \approx 25$ and $A_2 \approx 31$, which is consistent with the measured values of the laser modulations. In this configuration, there is nonzero bunching in the $n = -2, m = 22$ and also the $n = 0, m = 20$ modes at approximately the 1-2% level, while all other modes have bunching at $< 0.1\%$ and were not visible. From equation 4.12, the $n = 0$ and $n = -2$ modes shift with opposite sign and equal magnitude away from the central peak, which gives rise to the equally spaced sidebands visible in figure 4.9. A similar analysis holds true starting with the $n = -1, m = 20$ 42 nm echo bunching which also splits in figure 4.9.

In this particular configuration, the magnitudes of the sideband bunching peaks ($n = -2$ and $n = 0$) were approximately equal, however this need not be the case. For example, choosing instead $A_2 = 34$ reduces the $n = -1$ bunching by $\approx 40\%$, and leaves a non-negligible subsidiary bunching factor only in the $n = -2$ mode. This would, in contrast to figure 4.9, result in an asymmetric spectrum around the central $n = -1$ peak. In principle, by carefully tuning the echo parameters, a wide variety of spectra are possible.

It is intriguing to consider how this effect might be exploited at an FEL facility. In order to amplify multiple colors in an FEL the separate colors must be separated from the resonant wavelength by $\Delta\lambda/\lambda \lesssim \rho$. To take a numerical example, consider an EEHG-seeded EUV FEL with $\rho = 10^{-3}$, $\sigma_E = 100$ keV, and lasers with $\lambda_1 = \lambda_2 = 266$ nm operating at the 50th harmonic (5.32 nm). Full control of the $n = -1$ optimized wavelength within the FEL bandwidth could be obtained by a zero-crossing X-band cavity with power to provide a 100 MeV on-crest increase in beam energy. As the $n \neq -1$ peaks furnish an even greater sensitivity to energy chirp, this provides an upper limit on the RF power necessary to provide wavelength control using this technique. Thus, the requirement of this zero-crossing X-band cavity seems eminently reasonable for a full FEL facility. Consider, by way of comparison, that the structures employed at the NLCTA were capable of providing an almost 100 MeV/m gradient, so little floor space need be devoted to such functionality.

4.4 Simultaneous EEHG and HGHG Signals

As was previously mentioned, one significant difference between EEHG and HGHG signals at the same wavelength is their response to electron beam energy chirp. In general, HGHG signals tend to respond more strongly to changes in the electron beam longitudinal phase space, at least for low values of $|n|$. Therefore, when there is significant jitter due to time of arrival or RF power which cause fluctuations in the energy chirp of the electron beam, the EEHG signal is significantly more stable than the HGHG signal. This effect was previously observed at the 15th harmonic of the 2400 nm seed laser [83]. In that experiment, however, the configuration was changed between the EEHG and HGHG setup in order to study the difference. Here we report a similar measurement using an EEHG and HGHG signal contained within the same electron beam.

The possibility of generating separate regions of EEHG and HGHG bunched beams, and separating them, exists because the laser beams are in reality finite in temporal extent. There may be areas where they overlap extensively, and others in which one laser or the other dominates. To extend this notion, one can consider instead of the global bunching factor of equation 1.1, a local bunching factor,

$$b(k, z_0, \delta_z) = \frac{1}{N_{e,z}} \left| \sum_{i=1}^{N_{e,z}} e^{ikz_i} \right|_{|z_0 - z_i| < \delta z}, \tag{4.13}$$

that is, a bunching factor in which one confines the sum to only those electrons within a distance δ_z from the reference position z_0, and $N_{e,z}$ is the number of electrons in such an interval. Then, the bunching factor becomes a function of position along the electron bunch. The scale δ_z is arbitrary, but should be chosen to be consistent with the length scales of interest. For example, δ_z might be the cooperation length in a free electron laser. Care should be taken, however, in the limit that δ_z becomes small enough that it does not contain a representative number of electrons. For a given δ_z, as the number of electrons is increased to infinity, a definite bunching factor is obtained. As the number of electrons is lowered, eventually the shot noise statistics overwhelm the 'true' bunching factor. The interval δ_z should thus be chosen to be well above the distance where shot noise overwhelms the fluid bunching.

In the fluid description of EEHG, however, the bunching spectrum computed from equation 2.124 can be promoted to a local quantity as well by making the transformation $A_{1,2} \to A_{1,2}(z)$. Some care should be taken in interpreting this bunching factor, as it assumes that the each individual subsegment at location z is infinite in longitudinal extent. However, as long as one is not concerned with the bandwidth of the resulting signals (and, due to the already demonstrated low resolution of the EUV spectrometer, we are not), and the typical length scale of change for $A_{1,2}(z)$ is significantly longer than the radiation wavelength it remains a reasonable approximation to consider.

To this end, we consider an infinite, uniform electron beam modulated by two laser beams, which are both Gaussian in temporal extent with scaled length $\sigma_z = 0.3$. The peak modulation amplitudes are $A_1 = 44$ and $A_2 = 36$, and the centers are offset by a scaled distance of 0.6. These laser profiles are shown in figure 4.10 a). The dispersions in both chicanes for the echo setup are set such that $B_1 = 0.818$ and $B_2 = 0.13$. The resulting bunching spectrum, both due to EEHG and HGHG

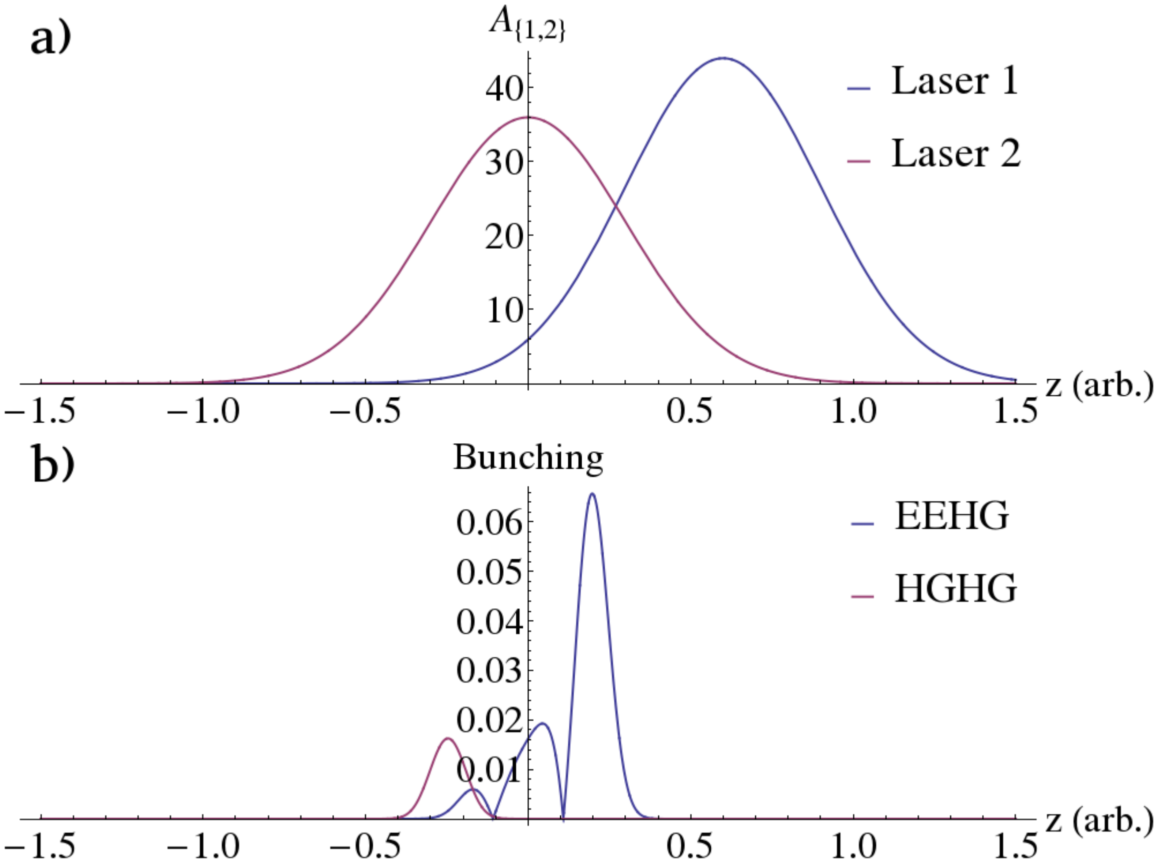


Figure 4.10: The EEHG and HGHG signals for Gaussian, temporally offset laser beams. Figure a) shows the laser profiles, each of which is a Gaussian with $\sigma_z = 0.3$. Figure b) shows the resulting bunching, both due to EEHG and HGHG. There are clearly separated regions of EEHG and HGHG bunching which may radiate independently in a downstream undulator.

contributions, is shown in figure 4.10 b). The bunching is computed assuming a laser wavenumber ratio $\kappa = 1/3$ and $n = -1$, $m = 21$. Note the z axis is of arbitrary scale, so long as the laser profiles are significantly wider than the radiation wavelength.

In the region near $z = 0.2 - 0.3$ where there is significant overlap in the laser profiles, a strong EEHG signal is established. However, at the tail of the laser 1 profile near $z \approx -0.15$, the laser is not strong enough to achieve any significant striation in the phase space and serves only to effectively increase the slice energy spread by a small amount. In this region, the second laser operates essentially in HGHG mode on a

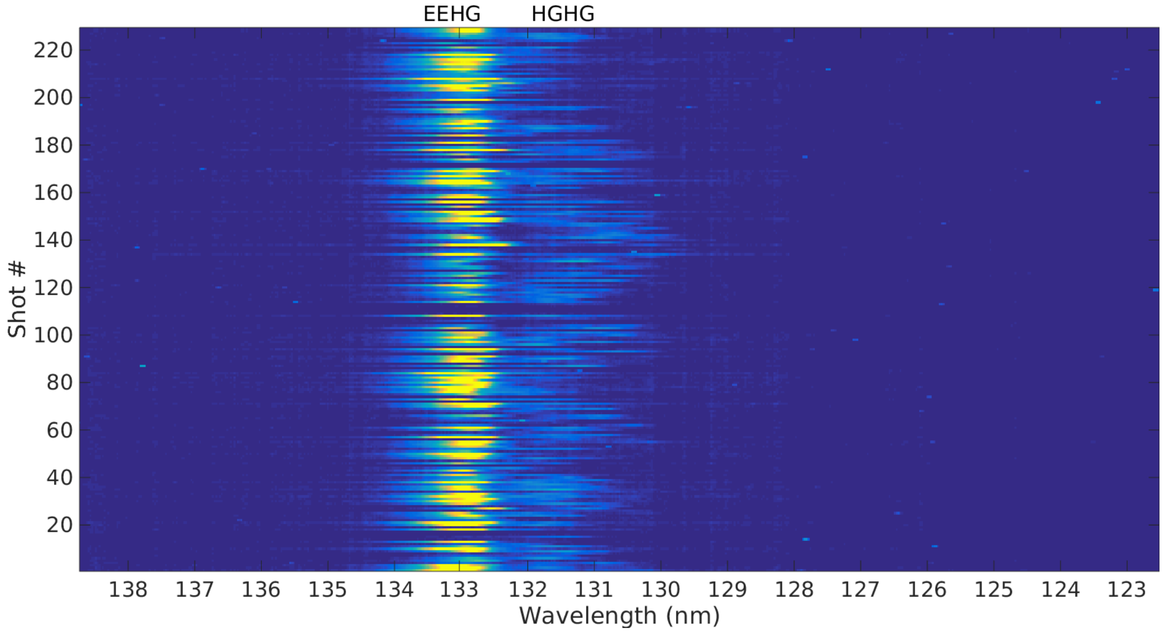


Figure 4.11: Simultaneous EEHG and HGHG signals in the vicinity of 133 nm. Both signals represent the 18th harmonic. This image was captured from the small VUV spectrometer.

slightly heated beam, and is able to establish a weaker, but significant, HGHG signal. These two signals essentially exist independently, and therefore could be considered as if there were two separate electron beams: one with an EEHG modulation and one with an HGHG modulation.

To investigate this possibility of simultaneous EEHG and HGHG signals, we set up the echo configuration at NLCTA to approximately reproduce the situation in figure 4.10. For this configuration, the lasers were in the 800nm/2400nm configuration, such that $\kappa = 1/3$. Furthermore, we investigated the signal in the low-harmonic regime near the 18th harmonic, or 133 nm, and the beam energy was approximately 185 MeV. The chicanes were re-tuned from their previous high harmonic settings to accomodate this lower harmonic upconversion such that $R_{56}^{(1)} = 12.5$ mm and $R_{56}^{(2)} = 2$ mm. Once an EEHG signal was established near 133 nm, the laser timings were slightly detuned to create a situation similar to figure 4.10. The resulting spectrum over approximately 220 shots is shown in figure 4.11. Due to the long wavelength, this spectrum was captured with the VUV spectrometer as opposed to the EUV spectrometer data shown in all other figures.

Due to a slight linear chirp on the electron beam, both the EEHG and HGHG signals are slightly offset from the 18th harmonic at 133.3 nm. Clearly, however, the HGHG signal is farther offset due to the previously discussed increased sensitivity to linear chirp. The HGHG signal is also noticably less stable in this offset wavelength, which is consistent with the results of [83]. Indeed, in attempting to establish an EEHG signal, the stability of the central wavelength serves as a diagnostic tool to rule out the possibility that it is due only to an HGHG interaction. We also acknowledge the possibility that the HGHG section of the beam is simply on a larger chirped portion of the electron beam. Since the electron beam is parabolic in energy due to the X-band structures, the EEHG region may sit at the relatively un-chirped center, while the HGHG region could exist at the largely chirped wings. Unfortunately, the experimental data taken at the time does not allow us to further investigate this possibility.

The technique described here is a crude version of the so-called fresh bunch technique [110] which can be used to increase FEL power, especially for multi-staged EEHG or HGHG techniques [71]. The particular configuration studied in figure 4.11, however, is unlikely to be a useful configuration due to the large laser modulation and low harmonic upconversion. Nevertheless, these results suggest the possibility of manipulating the laser modulation envelopes in order to create distinct modulation within a single electron beam. Coupled with the tuning of the dispersion (sec 4.2), and possibly linear chirp (sec 4.3), this creates a wealth of possibilities for fine-scale tuning of the seeded electron beam bunching spectrum.

Chapter 5

EEHG and FEL Simulations

> To err is human, but to really
> foul things up you need a
> computer.
>
> Paul R. Ehrlich

In order to fully simulate all the physics of the EEHG process, it is necessary to accurately model the electron beam from its birth at the cathode up until and through the final radiating undulator. In general, the final radiating undulator should take into account any possible gain from the EEHG prepared beam, so the full simulation process is similar to that which occurs when simulating SASE FELs. This process is known as a start-to-end (S2E) simulation, in that the electron beam is fully simulated from the gun through the FEL radiator. Detailed studies of the radiation, such as its diffraction through long or complicated optics elements, can be performed separately, although these effects will not be included in these simulations.

The business of simulating the electron beam through the entire accelerator complex is difficult because of the large disparity in energy between the beginning and end of the beamline. Near the electron gun, the electron dynamics in the non-relativistic regime are important to accurately capture the emittance and energy spread of the resultant beam. In this regime, great care is taken to model the space charge forces within the nascent electron bunch. After the first accelerating structure, however, the beam energy is generally high enough that space charge forces can either be ignored completely or modeled with simple approximations. The simulation therefore often proceeds in discrete steps where first the electron gun is handed with a space charge code, e.g. ASTRA [103] or PARMELA [111]. Simulations of the NLCTA gun using ASTRA were previously discussed in Chapter 3 in the context of beam characterization, and ASTRA is used throughout this chapter as well.

Afterwards, the beam is then propagated through the transport line and subsequent acceleration structures with a faster code designed for beam optics, e.g. ELEGANT [112]. Some codes provide a unified functionality, in which the computation of space charge can be seamlessly carried out throughout the entire beamline, e.g. IMPACT [113]. It is during this portion of the simulation that any EEHG simulations can be included. Generally, the interaction between the electron beam and laser in the modulating undulators can be accurately modeled analytically [40], and the dispersion in the chicanes can be calculated using 2nd order transport.

There are some notable exceptions, however, when these simple results cannot be applied to the modulation and dispersion portion of the beamline. The first exception is when there exists significant gain in the modulation section, as is often the case in a multi-stage HGHG beamline. In this case, one must employ a full FEL code to model the interaction, and possibly switch back to a beam transport code to model the transport between modulation sections. The second exception is when the beam density is high enough that either coherent synchrotron radiation [114] [49], intra-beam scattering [115] [116], or the stochastic CSR term (chapter 6) become important effects. While simple analytic models for these effects exist, a complete description requires an intensive computational model. Therefore, one must carefully anticipate the magnitude of any smearing effects in the chicanes and deal with them appropriately, either through a simplified but sufficient model, or with a full and expensive numerical simulation. For all simulations presented here, the code ELEGANT is employed for the beam transport and simulation of the EEHG undulators and chicanes. Due to the relatively low beam density of the NLCTA beam, it is not necessary to take account for any of the aforementioned radiation effects.

Finally, the electron beam arrives at the FEL undulator where the electrons are again handed off to a third simulation code. The popular FEL simulation code GENESIS [117], employs the Slowly-Varying Envelope Approximation (SVEA) [118]. The SVEA allows one to average the radiation field envelope over a radiation wavelength, which greatly speeds up the simulation of the FEL process. The radiation and electron beam, then, can be sliced up into discrete portions, each with a length of the radiation wavelength. This approximation is valid and has been extensively tested over a wide range of SASE parameters, leading to its widespread adoption for machine study and optimization simulations. However, there has been some doubt that this approximation gives valid results in the case of the complicated EEHG phase space, which may contain variations on the radiation wavelength scale. Additionally, the implementation of the slicing procedure can conflict with electrons moving substantial longitudinal distances in the radiating undulator due to the laser modulation, requiring awkward work-arounds. For these reasons, the FEL simulations in this thesis are performed with the un-averaged FEL code PUFFIN [119].

5.1 Benchmarking FEL Codes

5.1.1 Harmonic Lasing Benchmarks

Before discussing simulations of NLCTA or novel simulations of EEHG, we first present new benchmarking results which extend the correspondence between averaged and unaveraged FEL codes. Several averaged codes were previously compared in a variety of harmonic lasing studies [120], and we extend this result with the inclusion of new unaveraged, PUFFIN results. In the previous study, the codes FAST, GENESIS, and GINGER were compared, so we briefly discuss each in turn.

The 3D version of the FAST code [121] was developed in the late 90s to allow a non steady state, 3D simulation of FEL physics with the computing resources of the time. As FAST is an averaged FEL code, the electron beam is sliced up into elementary volumes of at least one resonance wavelength in longitudinal extent. The radiation field is expanded into azimuthal modes, rather than sampled on a Cartesian grid, with the number of azimuthal modes m defining the resolution of the simulation. In situations which have smoothly varying field profiles, and especially near saturation, this approximation may allow many fewer sampling points to be used than would be required to achieve the same resolution on a Cartesian grid.

GENESIS [117], by contrast, discretizes the transverse field onto a Cartesian mesh. In describing the dynamics of the electrons through the undulator, the wiggling motion on an individual period scale is averaged over to produce the familiar $\mathcal{J}$ factor (see eq: 2.66). The result is that the coupling to harmonics of the fundamental can be achieved by simply solving the relevant equations with the substitution for $\mathcal{J}_h$ [122].

Finally, GINGER [123], like FAST, employs cylindrically symmetric azimuthal modes in its field decomposition. The grid spacing, however, is not linear in this space, but instead follows a sinh function formula which yields more grid points near the dense center of the beam. Like GENESIS, GINGER models harmonics through the effective coupling of $\mathcal{J}$.

These three codes all differ on a pragmatic level in their written language, parallel implementation, and specific algorithms in other ways which are not listed here. The important distinction, however, is that they all employ the SVEA approximation. PUFFIN, on the other hand, does not.

PUFFIN is the 3D successor to an earlier 1D unaveraged code. As it does not make the SVEA approximation, PUFFIN allows the electromagnetic field to be consistently modeled at the sub-resonance scale, limited only by the Nyquist criterion due to the sampling. Since the field is discretized in all three dimensions, it is quite similar in spirit to a Particle In Cell (PIC) plasma physics code.

The diffraction of the electric field is performed via the Fourier split-step method, in which the field first diffracts in free space, and then in a second step propagates with no diffraction but driven by the electrons. Electron motion, as the field, is individually tracked on a sub-period scale. Therefore, nowhere in the equations does the $\mathcal{J}$ factor exist – it is implicitly included by modeling the detailed motion of the electrons in the undulator. A corollary of this, however, is that in order to accurately describe this motion, $\gtrsim 10$ integration steps must be performed per undulator period. Combined with the inclusion of sub-harmonic longitudinal grid spacing, it is easy to see that PUFFIN is enormously more computationally expensive than its averaged cousins.

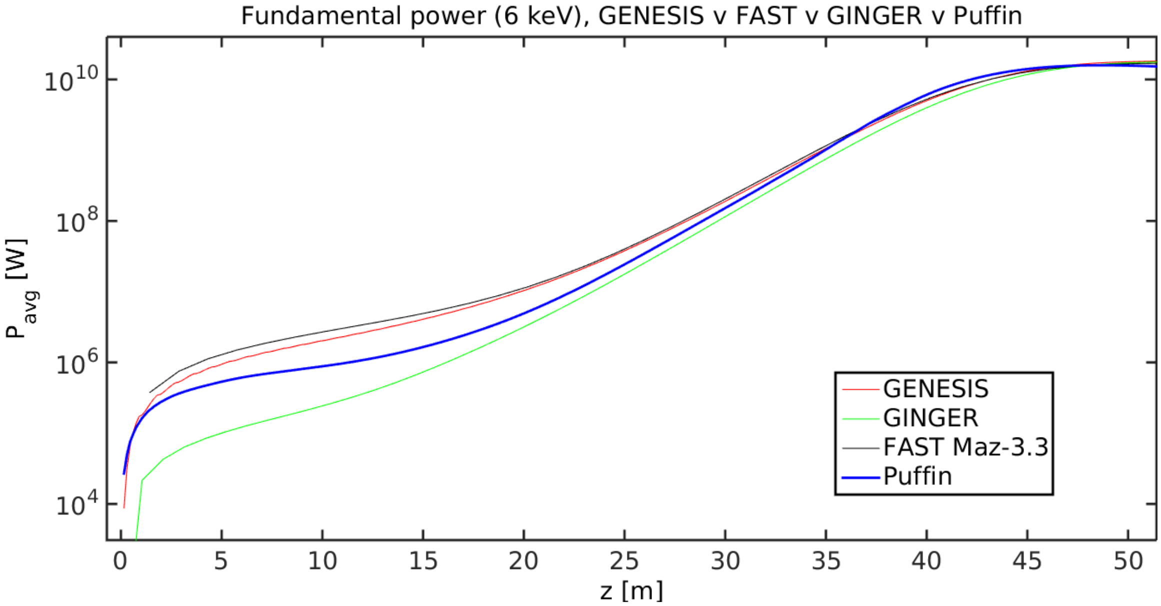


Figure 5.1: A comparison of the gain curves for the LCLS-I like scenario at the fundamental of 6 keV between codes. The Maz label for the FAST curve specifies the azimuthal modes employed.

However, because of this extra information, it is implicitly assumed that simulations in PUFFIN should be more accurate than averaged simulations, and therefore it can be used to benchmark these speedier alternatives.

The first situation to benchmark, and one which was also studied in reference [120], is a simple SASE and artificial harmonic lasing setup. For this case, the facility setup is an approximant to LCLS-I operating at 6 keV fundamental radiation energy. The electron beam energy is 11.62 GeV with $\epsilon_N \approx 0.4$ µm, while the undulator has $K = 3.5$ and $\lambda_u = 3.5$ cm. The slice energy spread is 1.4 MeV, and the current is 3 kA. This electron beam is artificially created, with assumed Gaussian dimensions to match these quantities, although in the longitudinal plane it is assumed a flat top distribution. Because of limitations in the PUFFIN code suite at the time these simulations were performed, and to facilitate comparisons between codes, the focusing was assumed to be done by an artificial channel which would keep an approximate beta function $\langle\beta\rangle \approx 26$ m. This focusing is artificial in that it cannot be replicated by a quadrupole lattice as it simultaneously focuses in x and y. The resulting gain curve for the fundamental mode is shown in figure 5.1.

Due to the high current and low value for emittance and $\langle\beta\rangle$, the fundamental radiation saturates quickly at $z \approx 45$ m for all codes. Indeed, all codes are in excellent agreement once in the exponential gain regime. Some disagreement does exist in the startup process. This disagreement is thought to be due to the various competing modes in the SASE startup, and the limit of any given code to resolve them. Therefore, when using a code such as FAST in which the number of azimuthal modes specifies the field resolution, the start-up power is in general highly dependent on the mode resolution. The same is true for the grid mesh used in PUFFIN to discretize the field. However, since there is not significant FEL interaction with these modes, one is allowed to use a lower resolution and still obtain an accurate high-gain FEL analysis, as is verified by the high-gain behavior. We briefly note, however, that although the exponential gain and saturated behavior is not affected by this discrepancy, self-seeded setups in which the process is terminated well before saturation may well be sensitive to a factor of $2 \sim 4$ in power suggested by figure 5.1. Therefore, a future avenue of study may be to conduct a more detailed comparison between a wider suite of codes focusing on this issue. Here we will not take this tack and instead focus on the comparison between the averaged and unaveraged nature of the codes.

Having established a baseline SASE benchmark, we now turn to an artificial harmonic lasing scenario. As mentioned above, the averaged codes generally employ the $\mathcal{J}_h$ factor to describe couplings to the various harmonic modes. Therefore, by artificially zeroing out these factors, one can examine only the production of, say, the third harmonic. This is precisely how the averaged codes achieve this ‘artificial harmonic lasing’. PUFFIN, however, models the complete dynamics of the electron beam, so no such artificial toggle is available. Instead, at every diffraction step, PUFFIN Fourier transforms the field in order to propagate it, which allows a convenient point of filtering. In order to replicate the artificial harmonic lasing, we apply a bandpass filter around the 3rd harmonic in PUFFIN each diffraction step. The width of this bandpass filter is arbitrary so long as it is large enough to capture the full FEL bandwidth ($\approx \rho$), and does not intersect nearby harmonics. For reference, PUFFIN performs a diffraction step once every 10 undulator periods in these simulations, as

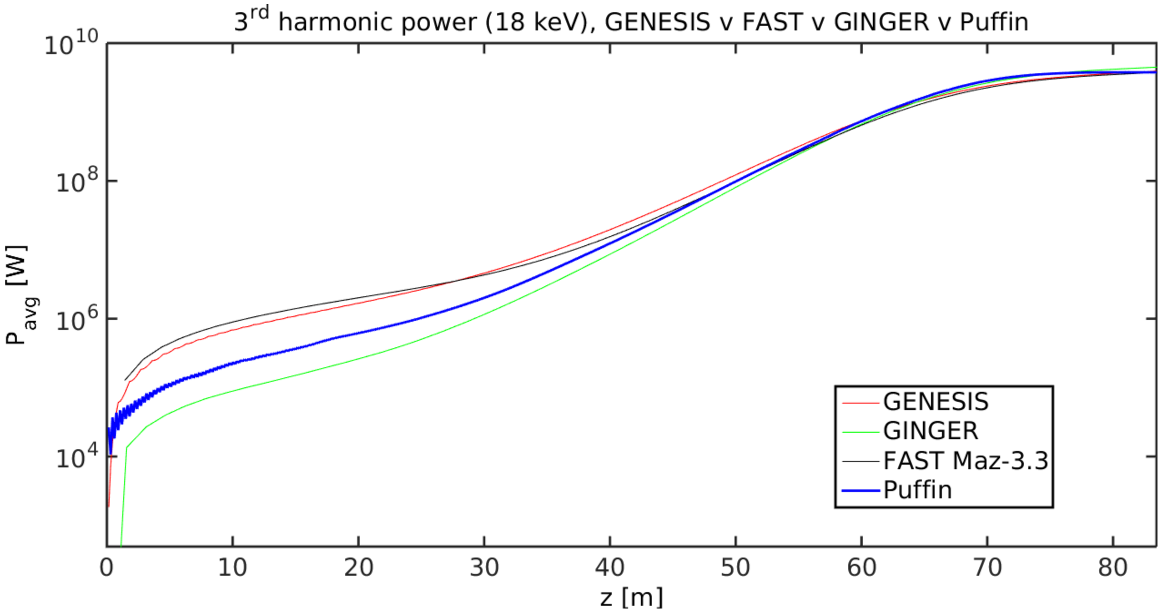


Figure 5.2: A comparison of the gain curves for the LCLS-I like scenario at the third harmonic of 18 keV between codes. The Maz label for the FAST curve specifies the azimuthal modes employed.

the short wavelength radiation does not diffract quickly (its associated Rayleigh range is approximately 24 m), so all radiation but the third harmonic is filtered out every 10 undulator periods. The comparison of these artificial harmonic lasing scenarios is shown in figure 5.2.

As before, the codes are all in excellent agreement well into the high-gain regime, but show disagreement during the startup process. One additionally observes a small amount of noise during the PUFFIN startup process, which is due to the discrete nature of the bandpass filter as described above. This comparison shows that, at least in this relatively simple case, the approximation of using $\mathcal{J}_h$ to describe the coupling to harmonics is reasonable.

Having established correspondence for a simple, artificial harmonic lasing scenario, we now turn to the real thing. The main idea behind harmonic lasing is to suppress the growth of the fundamental radiation mode by installing a series of periodic phase shifters [124]. These phase shifters can take the form of small chicanes or partial undulator periods, which serve only to delay the electron beam relative to the photons

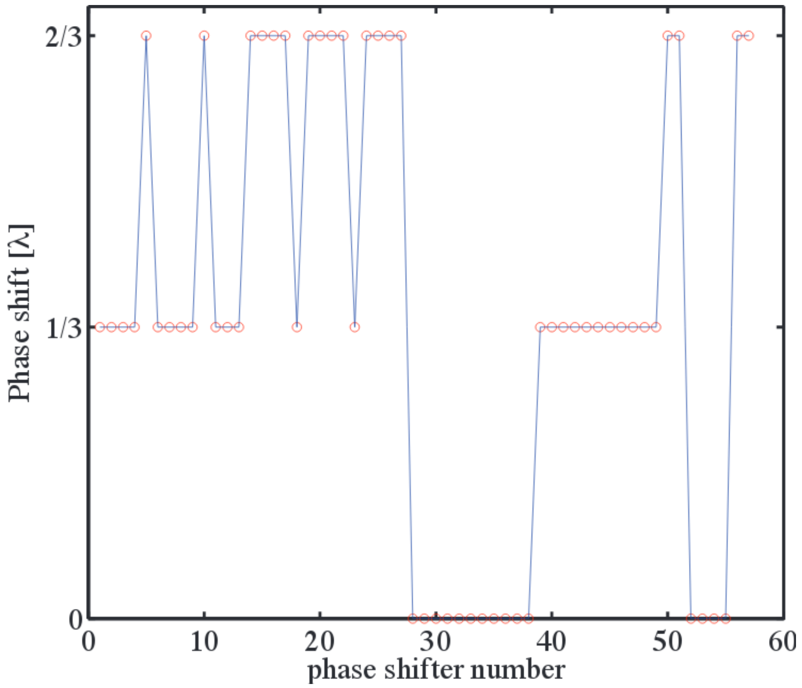


Figure 5.3: The phase shifts for the realistic harmonic lasing in the LCLS-II case as a function of shift #. Higher shift # corresponds to farther along the undulator line. Note all shifts are a multiple of the target harmonic: $\lambda/3$. Figure from [120].

by a fraction of a wavelength. Consider both the fundamental radiation wavelength λ and its third harmonic $\lambda/3$. As the electron beam radiates and begins the process of FEL microbunching, the electrons are pushed into buckets spaced by one (fundamental) radiation wavelength. Then, if they are shifted longitudinally by a distance $\lambda/3$ the electrons will radiate partially out of phase, thus canceling out some of the fundamental radiation, while they continue to radiate in phase for the third harmonic. Thus, if repeated phase shifts at a multiple of $\lambda/3$ are performed, the growth of the fundamental is stunted while the third harmonic proceeds unfettered.

It is precisely this segmented structure that was studied to benchmark various codes in a true harmonic lasing regime. For this case, however, a beam more similar to LCLS-II was adopted, with an energy of 4 GeV, emittance of 0.4 µm, slice energy spread of 0.5 MeV, and peak current of only 1 kA. The undulator was similar to LCLS-II as well, with $K = 2.23$ and $\lambda_u = 2.6$ cm, and a strong artificial focusing

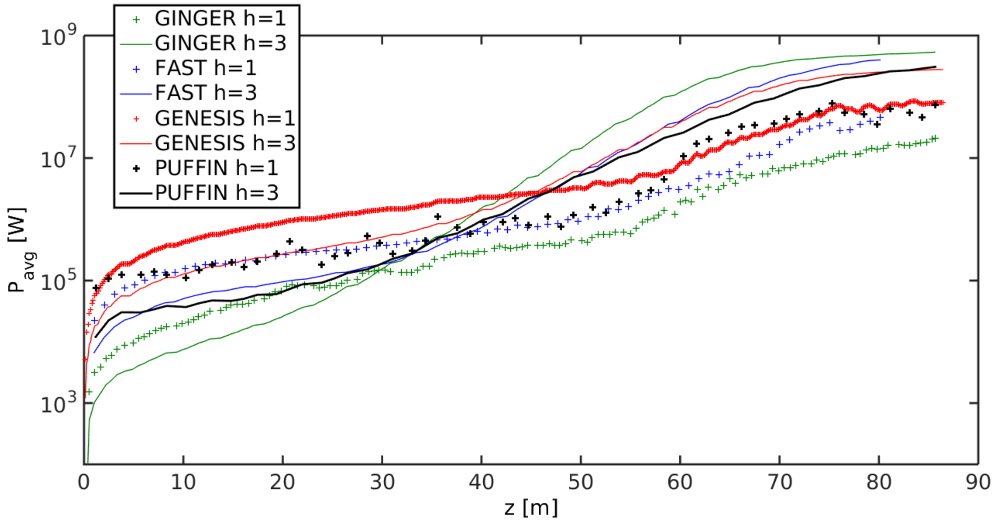


Figure 5.4: A comparison of the gain curves for the LCLS-II like scenario at the first ($h = 1$, stars) and third ($h = 3$, solid lines) harmonics. Note that due to the close spacing of the Genesis $h = 1$ points, this curve appears as a solid line.

lattice with $\langle\beta\rangle = 12$ m. For these parameters, the fundamental energy is 1.67 keV, while the third harmonic is at 5 keV. Note that, for the PUFFIN, FAST, and GINGER models, a continuous artificial focusing lattice was employed throughout, while the GENESIS model had a realistic FODO lattice which merely generated the requisite average beta. The undulator lattice was modeled as continuous, 45 period segments, which were broken up by short or long drift sections. The short sections were 5 undulator periods long, while the long drifts were 30 periods long, and they alternated in a short-short-long pattern. Each of these break sections also contained a phase shifter, whose shift value was chosen according to the distribution shown in figure 5.3. All phase shift values are a multiple of $\lambda/3$, and the perscription for choosing them comes from reference [125], with detailed considerations of the 3D FEL physics giving rise to the non-uniform distribution of phase shift values. The result of this simulation is shown in figure 5.4.

With this realistic harmonic lasing simulation, we begin to see differences in the results of each code. To begin with, the problem of resolving startup modes persists. Each code shows that third harmonic gain begins in earnest between 30 and 40 meters, and saturates somewhere near 80 m. The fundamental also appears effectively suppressed until about 60 m in each simulation, which suggests that this is a feature of the phase shifting lattice shown in figure 5.3 rather than a peculiarity in any particular code. Nevertheless, especially when one considers the ratio of fundamental to third harmonic power, there is substantial disagreement between the codes. GINGER, in particular, sees the largest disparity between the two, while the other three are more similar but show small deviations. The root cause of this disagreement is unknown, although the resolving of startup modes, by changing the grid resolution, was investigated and does not appear to affect this final harmonic-to-fundamental power ratio.

It is clear that in the case of a realistic, harmonic lasing simulation, the exact simulation parameters (grid slicing, shot noise handling, macroparticle number, etc.) may have some noticable effect on the final result. Further study is apparently necessary to pin down appropriate simulation parameters to generate a sufficiently accurate simulation of these scenarios. Nevertheless, especially from a glance at figure 5.4, one certainly cannot distinguish between the averaged and unaveraged models. Therefore, in so far as this distinction is concerned, it appears the averaged models are doing a sufficient job at modeling the physics of harmonic lasing.

5.1.2 Seeded FERMI Benchmarks

We now turn to benchmarking averaged and unaveraged FEL codes in the more advanced case of a seeded electron beam. We compare only GENESIS and PUFFIN, and choose accelerator parameters intended to model the FERMI accelerator [5]. The FERMI machine has been operating in both single-stage HGHG [70] and two-stage HGHG modes [71] for some time. Furthermore, there are currently proof of principle experiments in the design and planning stages for an EEHG configuration at FERMI [126]. Therefore, this machine makes an excellent test case against which to compare simulation codes.

The first scenario we imagine is a replica of the original HGHG configuration implemented FERMI [70]. For this study, the electron beam energy is 1.24 GeV with slice energy spread of 150 keV, emittance of 2 µm, and peak (flat top) current 300

A. The undulator used is planar and has $K = 3.452$, $\lambda_u = 5.5$ cm, and an artificial focusing lattice with $\langle \beta \rangle = 10$ m.

The HGHG manipulation is performed in an analytic way, rather than tracking a laser-modulator interaction and a full chicane simulation. Equations 2.110-2.111 are therefore used to model the HGHG manipulation, which amounts to having a perfectly sinusoidal modulation and perfectly linear R_{56} dispersion. The seed laser has wavelength $\lambda = 260$ nm, and its intensity is chosen to produce a modulation amplitude of $A = 6$, or about 900 keV. By using the analytical formula for the modulation, we implicitly assume the laser pulse to be infinite in temporal and transverse extent.

The subsequent chicane is tuned to produce optimal bunching at the 8th harmonic of the seed laser, but is chosen to slightly under-bunch the electron beam with a value $B = 0.164$, as opposed to the optimal choice of $B = 0.182$. This slight detuning is to account for the significant R_{56} of the radiating undulator itself, such that the electron beam does not over disperse before significant gain is accumulated. This value was obtained by running several simulations with various detuned B values and taking the approximate maximum of the output power. This procedure is analogous to how the chicane values would be optimized in practice, by varying the chicane R_{56} and optimizing for the highest FEL output power on a downstream detector. While in this example, there was a significant difference in optimal B values ($\approx 10\%$), this is a peculiarity of the high-modulation HGHG scheme and relatively long radiation wavelength. EEHG experiments, and in general harder X-ray FELs, will be much less sensitive to this additional R_{56} and the optimal B values will be closer to those computed from the bunching factor.

The resulting gain curves for both GENESIS and PUFFIN simulations, as well as the experimental data from [70] are shown in figure 5.5. The GENESIS simulation includes the realistic undulator module lengths, as well as break sections with quadrupole focusing. The PUFFIN simulation, by contrast, consists of a single undulator with an artificial focusing lattice. To make a meaningful comparison to both the data and the segmented undulator, the PUFFIN horizontal coordinate is therefore scaled by the ratio of physical beamline length to active undulator length in order to approximately account for this. This scaling should fairly accurately capture the

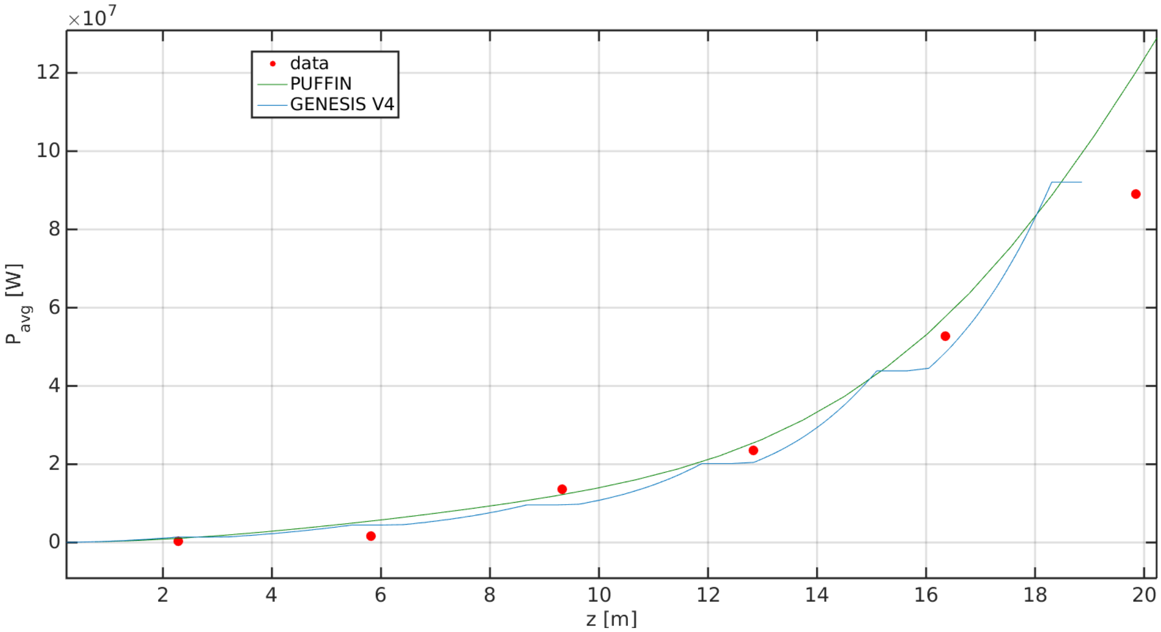


Figure 5.5: The gain curve for the HGHG modulated beam with FERMI-like parameters compared between PUFFIN, GENESIS, and the experimental data. The PUFFIN curve is scaled along the undulator axis to account for the lack of break sections.

dynamics so long as the diffraction effects in the break sections are not large. For this HGHG setup, the diffraction effects are small until the beam enters the saturation regime, which is not modeled here.

The GENESIS simulation in figure 5.5 is performed with Version 4 of the code: A new, experimental version compared to the one in common use. This new version has many new features, but importantly here, it allows the current of the electron bunch to be sliced up on a sub-wavelength scale. Thus, the current spike due to compression (see the final plot in figure 2.22) can be resolved, while in the old version of GENESIS it is simply averaged over. In principle this may affect the fine structure of the time-dependent simulation, but a separate study suggests that for this case the final output power and spectrum are generally unchanged between GENESIS versions 3 and 4.

With these caveats in mind, the agreement between PUFFIN, GENESIS, and the experimental data is quite good. The final data point seems to suggest a slight

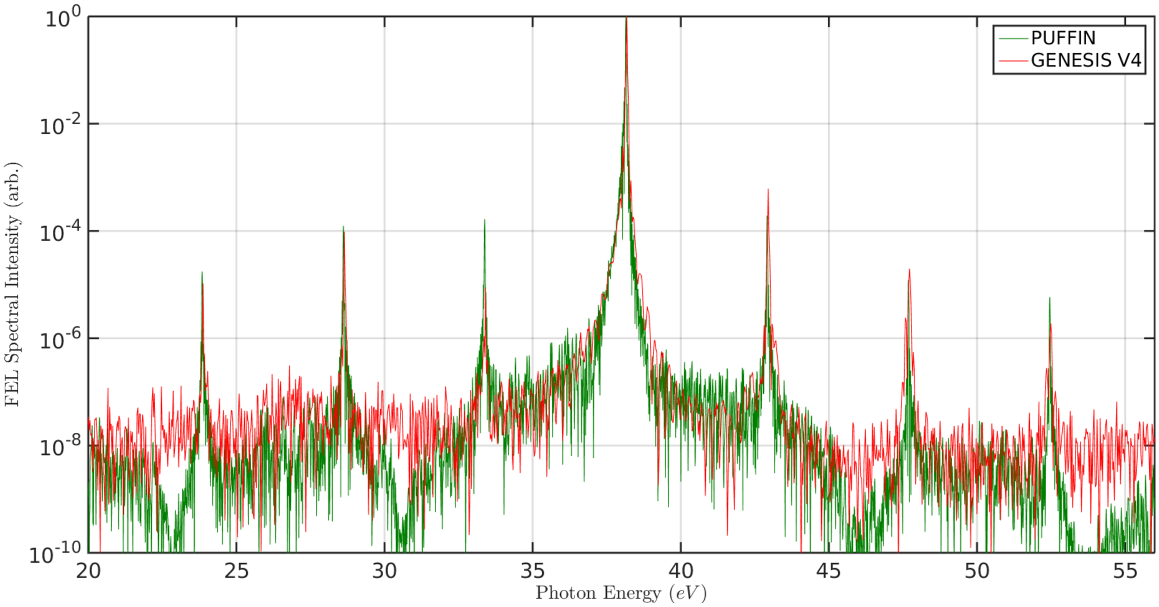


Figure 5.6: The FERMI HGHG spectrum compared between PUFFIN and GENESIS at $z = 18$ m. The spectral intensity is normalized to its maximum.

saturation based on the simulated gain curves in both PUFFIN and GENESIS.

We consider now the radiated spectrum near $z = 18$ m, or the approximate end of the undulator, which is shown in figure 5.6. The 8th harmonic of the seed laser, which is the one resonant with the radiating undulator, has energy 38.15 eV. In addition to this, we see all the HGHG harmonics which are produced as a result of the modulation and dispersion (which are all present, see fig: 2.24). These side harmonics are far outside the FEL bandwidth, however, and thus are not amplified and are at a level of 10^{-4} compared to the main peak (which is approximately the gain of this FEL). The width of the spectral peaks, as well as their relative amplitudes, are in good agreement between the codes.

These results suggest that GENESIS is quite accurate in modeling the simple single-stage phase space manipulations required for an HGHG-seeded FEL. We therefore turn to the more complex case of an EEHG FEL.

For the EEHG simulations, the setup is similar but slightly altered from the HGHG setup to account for the changes in the machine from 2012-2017. The electron beam energy is now 2 GeV, the slice energy spread 200 keV, slice emittance of 1 µm, and

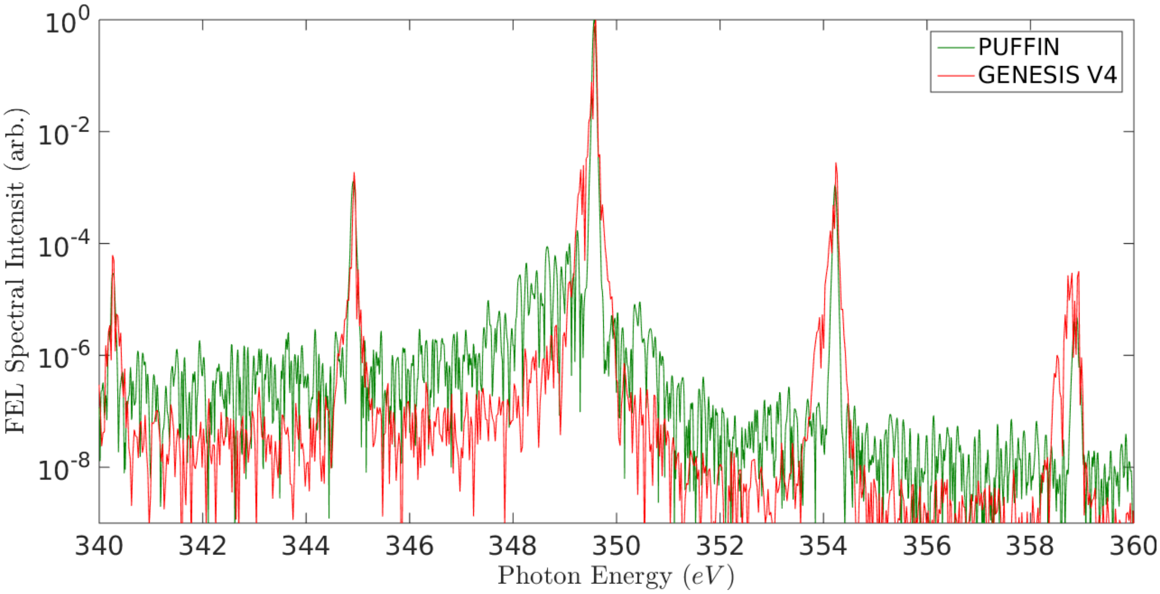


Figure 5.7: The FERMI-like EEHG spectrum near saturation for GENESIS and PUFFIN simulations. The spectrum is normalized to its maximum.

the peak (flat-top) current is increased to ≈ 650 A. The undulator is also changed, to account for the difference between the FEL-1 and FEL-2 beamlines at FERMI, and now has $K = 2.28$, $\lambda_u = 3$ cm, and $\langle\beta\rangle = 10$ m.

There are now two seed lasers, each of which has $\lambda = 266$ nm. The EEHG configuration is designed to optimize the bunching at the 75th harmonic of this seed, or 3.54 nm. To this end, the EEHG parameters selected have $A_1 = A_2 = 3$, $B_1 = 25.9487$ and $B_2 = 0.353074$. These selected parameters optimize the bunching at the 75th harmonic, but may not be attainable in practice. This is especially true for the large value of B_1, which requires a large and well-designed first chicane. Nevertheless, these parameters provide a reasonable testing ground for comparing averaged and unaveraged codes in handling a complex phase space.

A comparison of the output spectrum near saturation between both PUFFIN and GENESIS is shown in figure 5.7. The clear peak in the middle is the 75th harmonic of the seed lasers, while the subsequent side peaks are nearby excited modes from the EEHG manipulation which are outside the FEL bandwidth. The relative amplitude, as well as widths, of the main and subsidiary harmonics are in excellent agreement between the codes. The logarithmic scale of figure 5.7 enhances what small differences exist, but the peaks generally show agreement out to 10% of the main peak height. GENESIS, however, seems to show a small redshifted sideband for the peaks of the 75th, 76th, and 77th harmonics which does not appear in the PUFFIN simulations.

With this same EEHG configuration, we also investigated the effects of additional structure on the electron beam. In particular, we choose a single sinusoidal oscillation imposed on the electron beam prior to the first EEHG modulation as (notation as in equations 2.118-2.121),

$$p_0 = p_{-1} + A_0 \sin(\kappa_0 \zeta_0), \tag{5.1}$$

where $\kappa_0 = k_0/k_1$, and k_0 is the wavenumber of the mode of interest, and now p_{-1} is the initial (scaled) momentum. This represents a single, sinusoidal mode of amplitude A_0 and wavenumber $k_0 = 2\pi/\lambda_0$. In a Fourier sense, these modes can be considered the building blocks of more complicated energy structure, such as that present in the microbunching instability (section 2.3.2). Note, however, that the microbunching instability involves not only energy modulations but their associated density modulations, which are not captured by equation 5.1. Nevertheless, the system's response to modulations of the type in equation 5.1 is an interesting test case both in regards to the microbunching instability and also how sidebands are treated in averaged and unaveraged codes.

With the same EEHG configuration as above, we add an additional modulation with $A_0 = 2$ and $\lambda_0 = 3$ µm. The resulting spectra near saturation are shown in figure 5.8. This additional structure on the electron beam manifests in both the PUFFIN and GENESIS simulation as a sideband for each EEHG harmonic. The location of the sideband is related to the wavenumber of the mode, such that ,

$$k_{\text{sideband}} = k_E \pm q k_0, \tag{5.2}$$

where k_E is the wavenumber of the relevant echo harmonic and q and integer. Generally, only those modes with $q = \pm 1$ contribute substantially, as those with larger frequency separations become exponentially damped, which can be seen by examining excitation bandwidth equation 2.128. In the context of this excitation bandwidth,

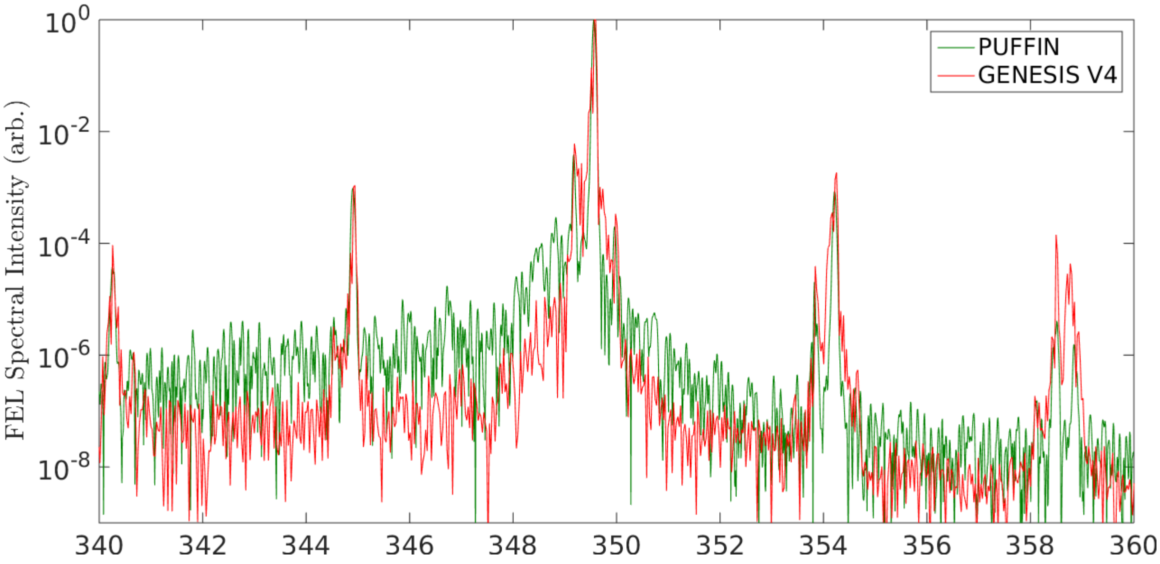


Figure 5.8: The FERMI-like EEHG spectrum near saturation for GENESIS and PUFFIN simulations with an included single-frequency modulation. The modulation has parameters $A_0 = 2$ and $\lambda_0 = 3$ µm. The spectrum is normalized to its maximum.

the presented configuration has $\xi > 0$, and therefore preferentially excites the low-frequency, $q = -1$ sideband from equation 5.2. This explains the absence of the high frequency sideband from the spectra in either PUFFIN or GENESIS simulations shown in figure 5.8.

The relative amplitude of this sideband peak, especially for the main harmonic near resonance, is in good agreement between the two codes. The presence of the sideband is also consistent in the other EEHG harmonics, although there are some amplitude variations at the 10^{-5} level for this unamplified radiation. Other discrepencies in the spectrum, while small, can be traced to individual shot noise variations in each simulation. Even upon averaging, variations may not converge as the PUFFIN simulations used a particle to macroparticle ratio of ≈ 36 in order to keep the simulation size manageable, while the GENESIS simulations were performed one-to-one.

These results strongly suggest that GENESIS V4 simulations are an adequate tool to simulate EEHG manipulated beams with the inclusion of microbunching-like energy structure. Despite the excellent agreement, due to the predilections of the author PUFFIN will be used for all subsequent FEL simulations. We will now take a closer look with PUFFIN alone at how EEHG interacts with the microbunching instability.

5.2 The Microbunching Instability and EEHG

One interesting question relating to EEHG is how its phase space transformations are impacted by structure from the microbunching instability (UBI). The UBI is believed to be the cause of the observed pedestal effects observed in self-seeded at the LCLS, for example [8] [127]. This configuration was recently analyzed analytically, wherein the FEL equations with seed radiation were solved in the presence of the UBI structure [128]. We aim to understand how a beam-based seed, such as one generated by EEHG, is affected by the same UBI and explore this question through PUFFIN simulations.

The microbunching instability was discussed in section 2.3.2, and for our purposes here we need only recall the basic form of the instability. It produces energy modulations $A(k)$ (as well as density modulations) with the form,

$$A(k) \approx \tilde{A}_0 k^2 \exp\left(-k^2/k_0^2\right), \tag{5.3}$$

for some critical wavenumber k_0 and amplitude $\tilde{A}_0$. This produces a broad band energy structure on the beam, which in general has significant contributions in wavelength space from $\lambda = 1 \sim 30$ µm or greater for choices of $\lambda_0 \sim$ µm, which is reasonable when compared to realistic UBI spectra (see figures 2.18, 2.19). The total UBI contribution can then be obtained as an integral over these modes,

$$p_0 = p_{-1} + \int_0^\infty \frac{A(k)}{k_1} \sin(ks - \phi(k)) dk, \tag{5.4}$$

for a phase $\phi(k)$ and normalization constant k_1. Naively, we might assume $\phi(k)$ to be random, but such a choice would reduce equation 5.4 to the addition of uncorrelated noise which would serve only to increase the effective energy spread of the beam. Rather, as is clearly true from visual inspection of figure 2.18, there is an inherent

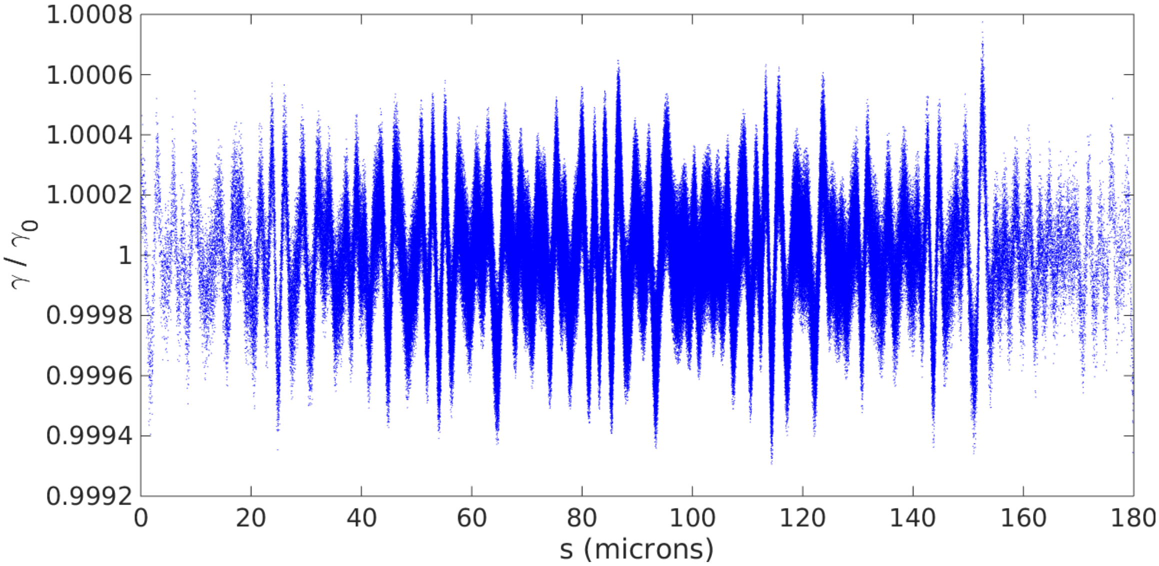


Figure 5.9: An example longitudinal phase space of the simulated microbunching instability using equation 5.5. This realization has $N_m = 40$, $\lambda_0 = 3$ µm, and $A_0 = 2$.

correlation length in the UBI structure that prevents this smearing into simple uncorrelated noise. We can ignore the complexities of this correlation by splitting the integral expression into a discrete sum of N_m separate modes,

$$p_0 = p_{-1} + \sum_{i=1}^{N_m} A(k_i) \sin(k_i s - \phi_i). \tag{5.5}$$

We can now assume that the various phases are uncorrelated, and shift the question to how many modes N_m should be used. Clearly, it make sense to only choose k values over the range where $A(k)$ is substantial. In the case where N_m is small, ($\mathcal{O}(1)$), we have a coherent modulation which is more indicative of several lasers than the UBI, which is not desirable. On the other end of the spectrum, $N_m \rightarrow \infty$ results in the same conundrum as the integral expression in which the resulting modulation simply becomes noise.

In practice, we simply choose a value of N_m which produces a spectrum that is qualitatively similar to what is observed in microbunching studies. This turns out to be a value in the neighborhood $N_m = 20 \sim 100$. The overall amplitude of modulation obviously also depends on N_m, so we must choose a method of normalization. We choose a normalized amplitude $\tilde{A}_0$ to be such that the sum of all N_m modes produces the same energy spread increase as a single sinusoidal modulation with amplitude A_0. Since each mode is independent, the total increase in energy is computed in an rms sense and we have the following relation,

$$\tilde{A}_0 = \frac{A_0}{\sqrt{N_m}}. \tag{5.6}$$

It is more convenient to therefore parameterize the modulation by the net effect through A_0 than the N_m-dependent parameter $\tilde{A}_0$. Another choice for normalization would be to parameterize to the total increase in energy spread of the beam. This total energy spread increase, Δ, is related to our choice of A_0 as,

$$\Delta = \sqrt{1 + \frac{A_0}{2}}. \tag{5.7}$$

One realization with $N_m = 40$, $\lambda_0 = 3$ µm, and $A_0 = 2$ is shown in figure 5.9. The random features of the instability are well captured in this instance, and all that remains is to choose appropriate values for λ_0 and A_0 to reproduce the overall features of the instability.

The question now is how does the combined EEHG/FEL system respond to these modulations? The FEL imagined in these simulations is an approximation to the FERMI FEL: It operates at 2 GeV, with slice energy spread of 1 MeV, and emittance of 1 um. For simplicity, we consider a continuous planar undulator which has $K = 2.28$, $\lambda_u = 3$ cm, and a constant β artificial focusing lattice with $\langle\beta\rangle = 25$ m.

The EEHG configuration is provided by two 266 nm lasers, each of which modulates the beam by 3 MeV, corresponding to three times the slice energy spread. For this study, the modulation is computed analytically, rather than tracking electrons through the laser-modulator interaction. Therefore, the modulation is perfectly sinusoidal, and the following dispersion sections represent perfectly linear chicanes. The chicanes are optimized for bunching at 3.5 nm, or the 75th harmonic of the seed

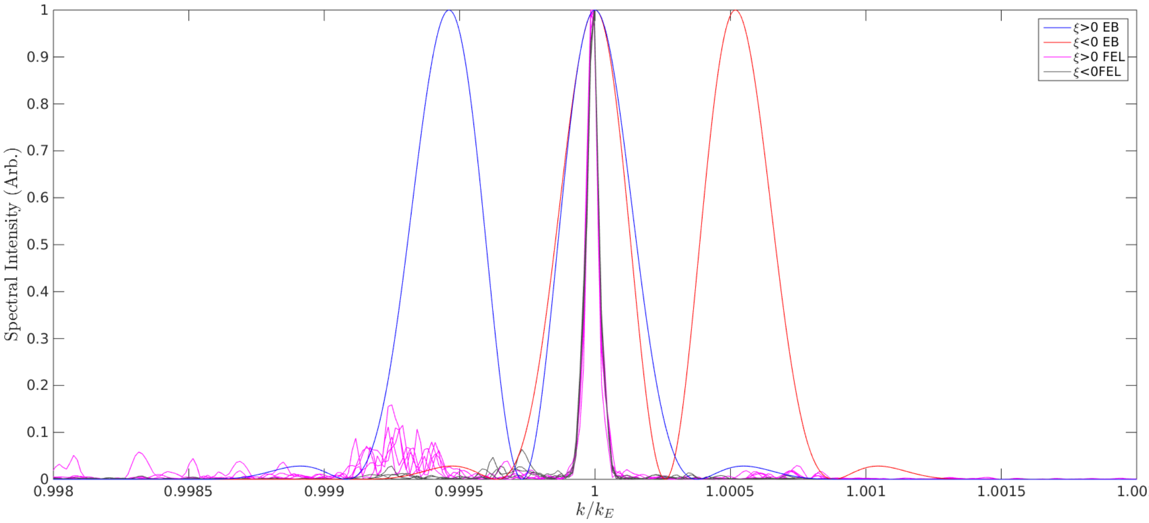


Figure 5.10: A comparison of the FEL spectra for $\xi < 0$ and $\xi > 0$ configurations with microbunching structure. Also shown are the excitation bandwidths for each configuration. Five separate realizations of the UBI noise are shown.

lasers. However, recalling that the value of B_1 has twin optima for positive and negative values of the parameter ξ (eq. 2.127), we consider setups with each configuration. The resulting FEL spectra for several shots, as well as the excitation bandwidths for the $\pm\xi$ configurations (eq. 2.128) are shown in figure 5.10.

While both ξ configurations look similar for the central EEHG peak, we observe that the $\xi > 0$ configuration has a significant side-band pedestal for low k. By contrast, the $\xi < 0$ configuration shows only minimal sideband intrusions, even in the blue region of its excitation bandwidth. Both EEHG configurations will preferentially excite the UBI modes in their own excitation bandwidth, but the FEL process is asymmetric. Due to the FEL detuning curve with energy (fig 2.14), blueshifted components compared to resonance can become highly suppressed while those which are redshifted suffer only mild degredation. In this sense, the FEL acts as a low-pass filter. Therefore, when the EEHG configuration is operated in the $\xi < 0$ mode, the UBI modes are preferentially excited in the blue and subsequently suppressed by the FEL mechanism. This observation is important to the operation of an EEHG-seeded FEL in which the mitigation of the UBI may be a key design constraint.

The microbunching instability also impacts the performance of non-optimal EEHG

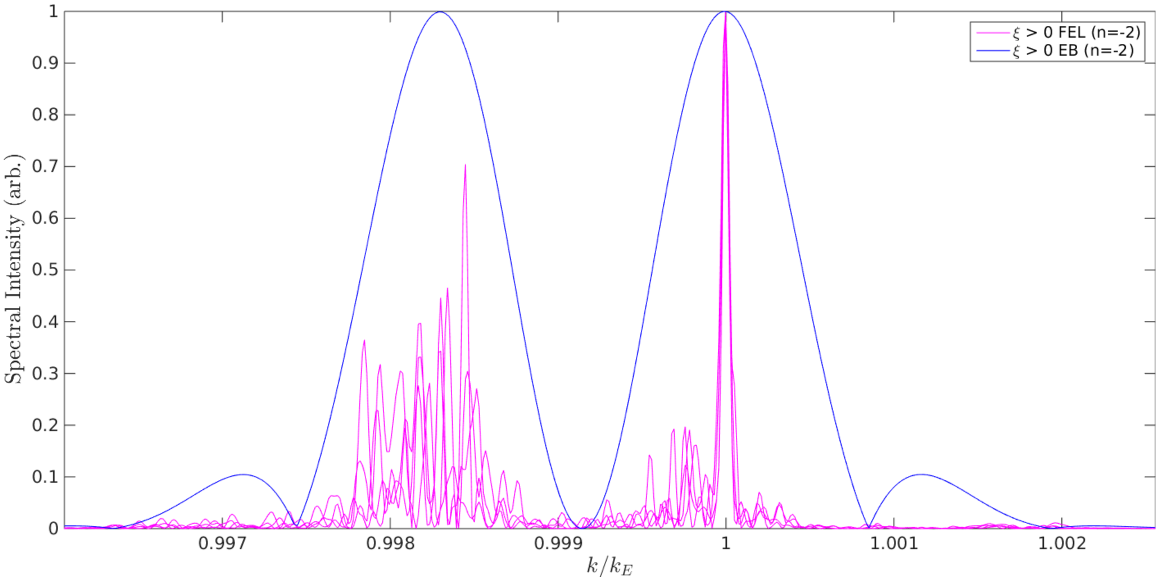


Figure 5.11: Various FEL spectra from the $n = -2$, $\xi > 0$ configuration with microbunching structure. Due to the lower bunching factor at the target harmonic, the contrast between the sidebands and central peak is less extreme than in the $n = -1$ case. Note the axis scale as compared to figure 5.10.

configurations, such as those with higher n number. The spectrum of the same FEL adopting the $n = -2$, $\xi > 0$ configuration with the same microbunching structure is shown in figure 5.11. Since the bunching at the target harmonic is lower in the high $|n|$ configuration, the contrast between the central peak and sideband is reduced considerably. Note, however, that the low-pass nature of the FEL still purifies the blue end of the spectrum. We conclude that in the presence of significant microbunching structure, choosing the $\xi < 0$ configuration is highly beneficial, especially when operating in a higher $|n|$ mode.

This interplay between the EEHG manipulation and FEL mechanics presents the possibility of minimizing negative effects due to the microbunching instability. Given a well-characterized UBI, one can selectively tune the EEHG configuration through study of equation 2.128 to suppress amplification at the peak location of the UBI. These sidebands, however, may actually be beneficial to certain user experiments, and experiments at FERMI have intentionally used them as a seed to generate multicolor FEL pulses [129]. The extended ability of EEHG to manipulate how the UBI presents as bunching structure on the beam, then, is another avenue to control and manipulate in detail the bunching structure that radiates in the undulator.

5.3 Start-to-End NLCTA Simulations

As previously mentioned, start-to-end simulations consist of three separate parts. The first is the low energy, space charge relevant section in which ASTRA is used to model the electron gun through the first accelerating cavity. The second section employs ELEGANT to track the electrons through the entire linac including the EEHG modulators and chicanes. Finally, PUFFIN is used to simulate the FEL interaction (or in the case of NLCTA, just coherent emission) in the VISA undulator.

Simulations of the NLCTA electron gun have already been described in section 3.3, so we briefly revisit them here. One difference between the optimization routine described previously is the focus on peak current. Having abandoned the possibility of significant gain in the VISA undulator, the more critical parameters become the slice energy spread and emittance of the electron bunch. The electron bunch length, and hence peak current, is important only to ensure the relative uniformity of the electron bunch after passing through the X-band acceleration cavities.

Due to the betatron smearing effect described in section 4.1.1, a low emittance is critical in order to preserve bunching at high harmonics. To illustrate this effect, we created several Gaussian electron bunches with different rms bunch lengths and emittances and pass each through an identical transport line. The transport line is optimized for EEHG with a zero emittance bunch and no energy structure (one which has no curvature due to the RF accelerators), and thus comes directly from the optimization of equation 2.124. The resulting bunching at the 75th harmonic of the 2400nm seed laser as a function of emittance and bunch length is shown in figure 5.12. The quoted emittance is the slice emittance at the end of the first linac section, rather than the measured projected emittance. The bunching is measured immediately prior to entry into the VISA undulator, and so does not contain the additional smearing incurred due to that focusing lattice.

These simulations were performed with 10^6 particles, which puts an effective floor

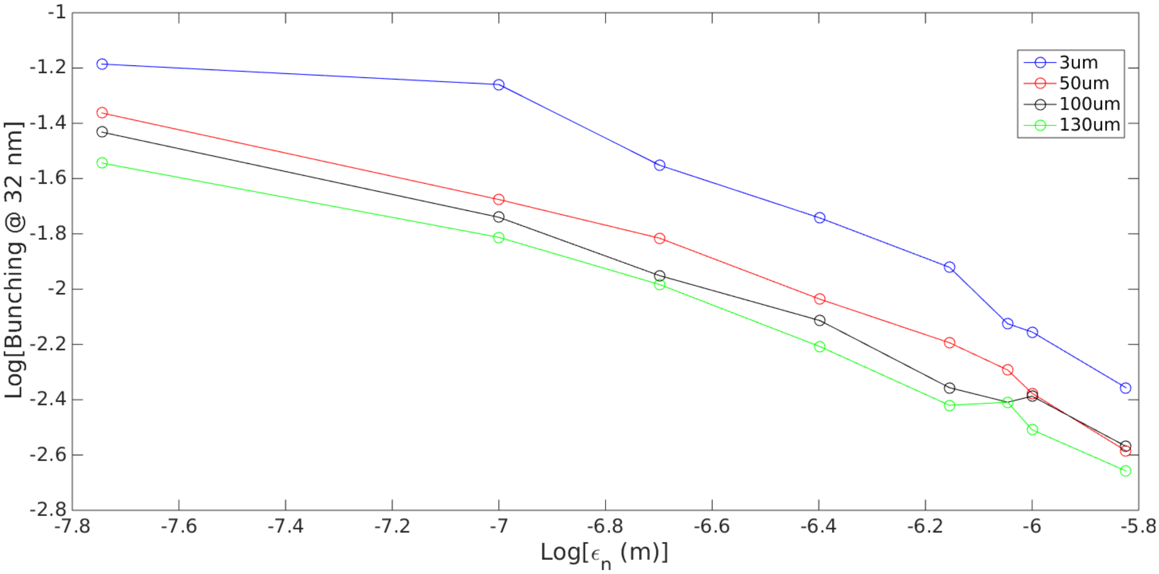


Figure 5.12: The EEHG-induced bunching at 32 nm as a function of emittance and bunch length. The different rms bunch lengths are shown as different colors, with 130 um approximating the true bunch length.

on the bunching around $1/\sqrt{N_p} = 10^{-3}$, which is approached by the longest bunches as $\epsilon_n \rightarrow 10^{-6}$ m. As the slice emittance could not be accurately measured with the NLCTA diagnostics, figure 5.12 provides a clue as to the true slice emittance. Values larger than ≈ 1 µm show suppression of the bunching, and we thus infer that the true slice emittance is below this level. Based on this observation, the following simulations assume a beam with initial slice emittance of $\epsilon_n \approx 0.7$ µm , rather than the measured projected value of ≈ 2 µm. In a sense then, the existence of bunching at such high harmonics provides a crude upper limit for the slice emittance of the electron bunch. This idea has recently been expanded upon in the HGHG-seeded FEL FLASH, wherein the slice properties of the beam can be recovered from the efficiency of the seeding process [130].

For the representative start-to-end simulation, the number of macroparticles was increased to 4×10^7 to be able to ultimately be able to resolve the low, high harmonic bunching. For a 50 pC electron bunch, this corresponds to a ratio of 8 electrons per macroparticle. These macroparticles are first generated by the Astra program `generator`, and are used for each subsequent leg of the simulation.

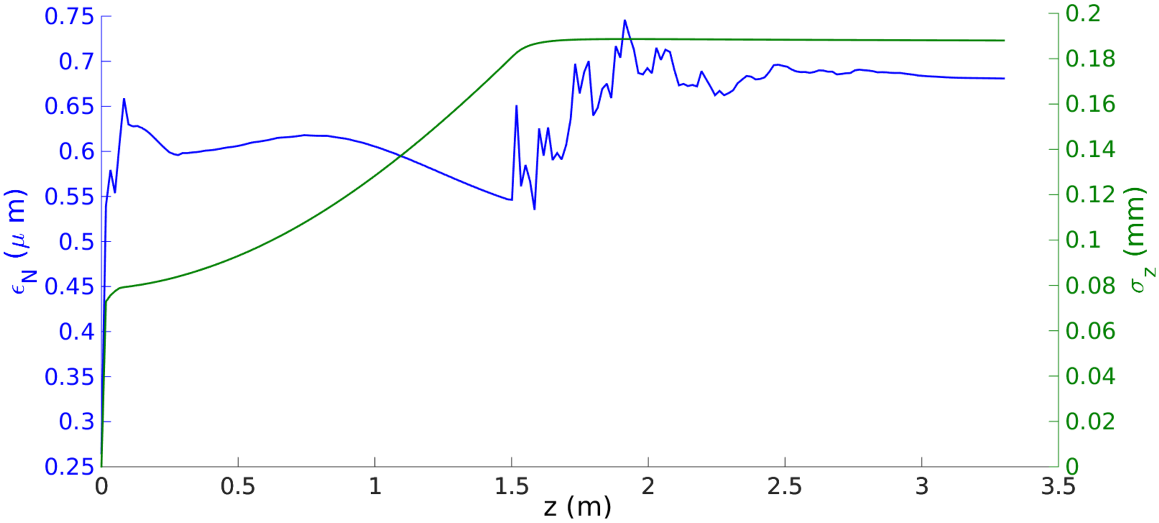


Figure 5.13: The evolution of the normalized transverse slice emittance and rms beam length in the NLCTA electron gun from Astra.

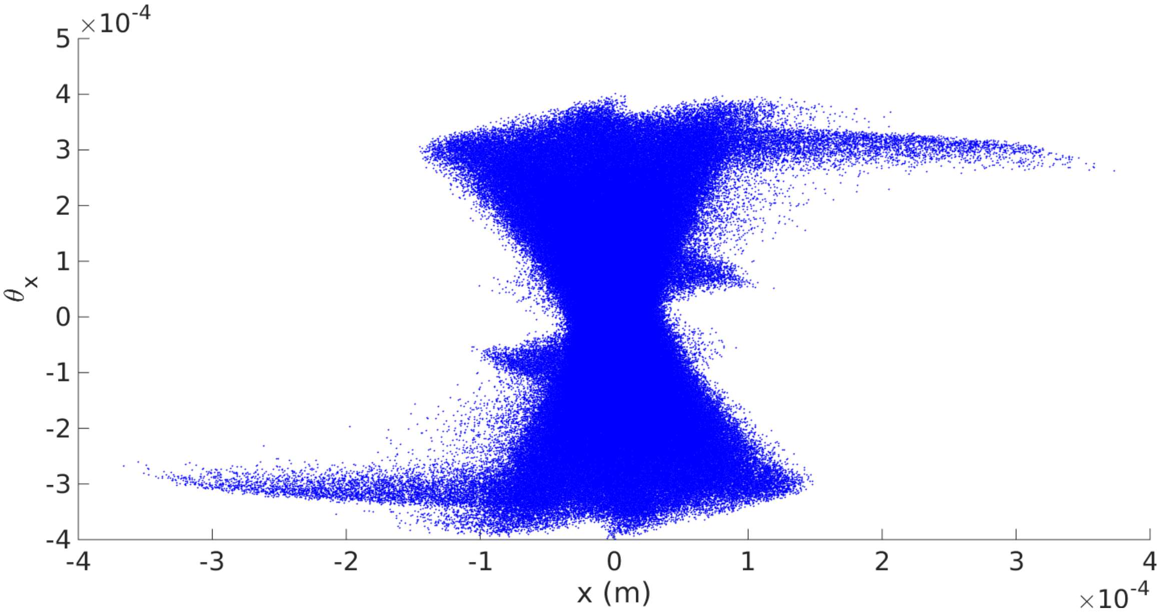


Figure 5.14: The transverse phase space distribution at the end of the NLCTA injector line. Plotted is a 100:1 down-sampled distribution. The vertical dimension is similar.

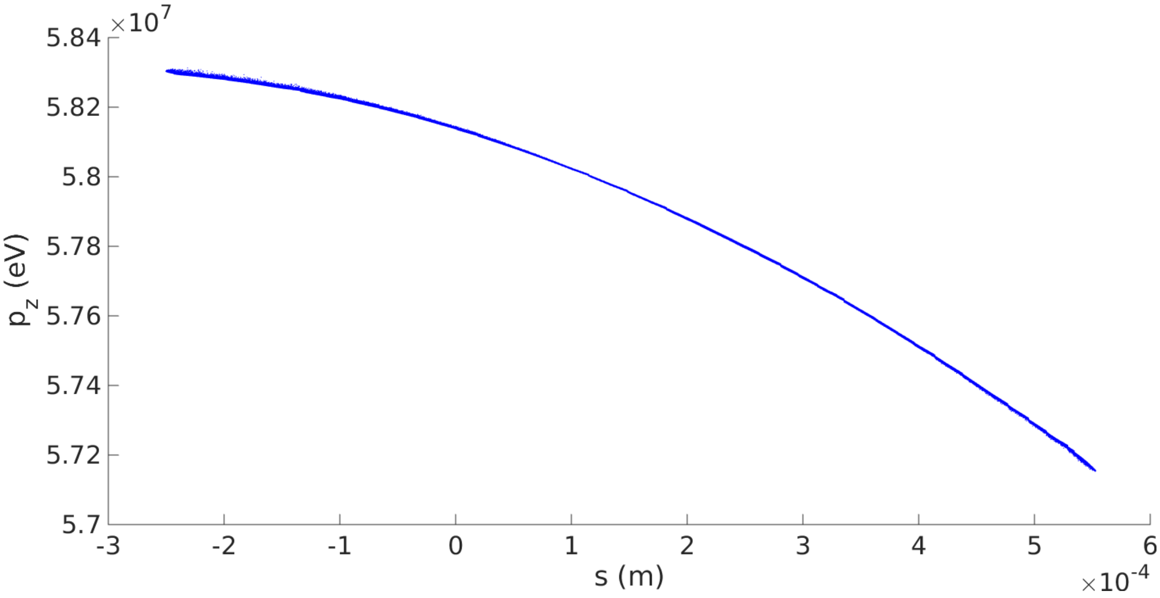


Figure 5.15: The longitudinal phase space distribution at the end of the NLCTA injector line. Plotted is a 100:1 down-sampled distribution. The off-crest acceleration in the final X-band structure results in a substantial energy chirp.

The electron beam of 50 pC is generated at the cathode with 1 ps FWHM time spread (0.4 ps rms) from a 1 mm spot. The transverse momenta are drawn from a Fermi-Dirac distribution with a work function of 4.4 eV and photon energy of 4.7 eV [131]. The S-band gun has a maximum accelerating voltage of 110 MV/m and is dephased by 28.65° forward of crest, such that the accelerating voltage grows with time as the electrons are emitted from the cathode. The first X-band cavity, which is also simulated in Astra, is also de-phased 9° off crest to provide some compression to the electron bunch. The evolution of the beam emittance and rms bunch length through the electron injector for this particular run is shown in figure 5.13. The final phase space distribution is shown for the transverse plane in figure 5.14 and in the longitudinal plane in figure 5.15.

The Astra model ends after the first accelerating structure and the particles are transferred to Elegant. The rest of the NLCTA beamline, up until the VISA undulator, is modeled in Elegant. The laser modulators are modeled using the `LSRMDLTR` element which integrates through the undulator trajectory with a co-propagating laser

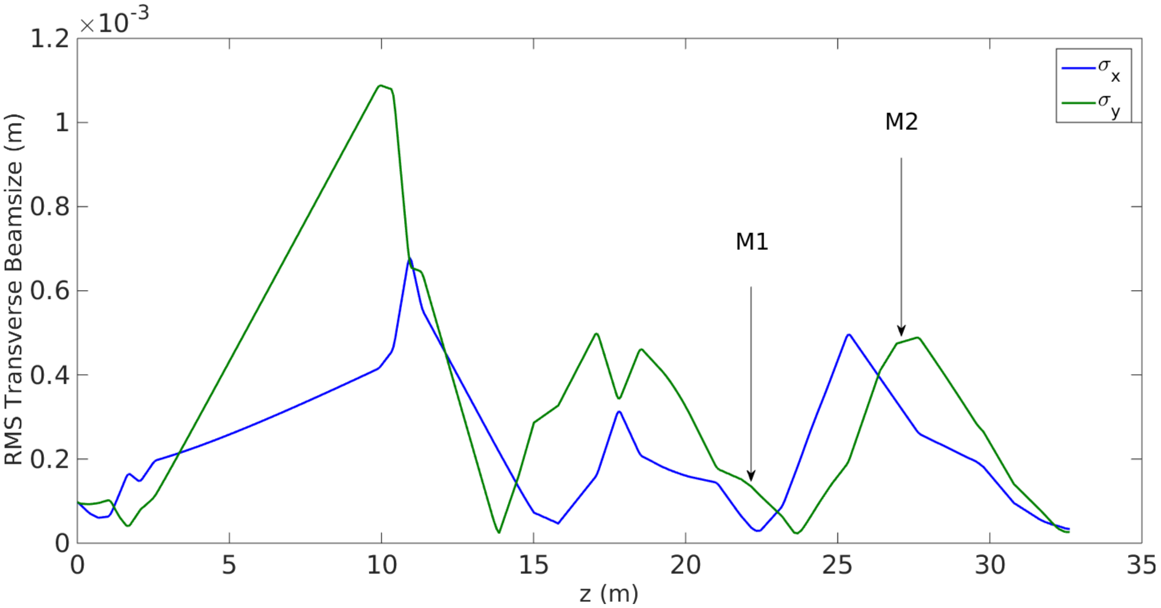


Figure 5.16: The simulated RMS beam sizes along the NLCTA beamline. The positions of the first (M1) and second (M2) modulator are indicated.

beam. The laser spot sizes used, 2mm, are significantly larger than the electron beam sizes and the resulting modulation is almost perfectly sinusoidal.

The full collection of NLCTA quadrupoles are scanned in order to find an adequate optics solution which satisfies three criteria. The first is that the electron beam size remains relatively small, less than 0.5mm rms, in each of the two modulators. The second criterion is that the final Twiss parameters are adequately matched into the VISA undulator at the end of the beamline. Due to the small matched beta functions for the VISA, this leads to a sharp squeeze at the end of the beamline. The final, and most tricky criterion, can be stated as the minimization of the quantity $\int \gamma_T dz$ over the EEHG portion of the beamline. This is done to minimize the emittance smearing effect discussed in section 4.1.1. Unfortunately, the built-in Elegant optimizer is limited, and $\int \gamma_T dz$ could not be optimized on directly. A useful proxy was to minimize the number of waists in the EEHG section of the beamline. Since waists contain small β and large α, and thus large γ_T, this provides a crude optimization. Indeed, without a careful choice of beam optics through the EEHG section any produced bunching can be destroyed before even reaching the VISA undulator. The resulting beam sizes

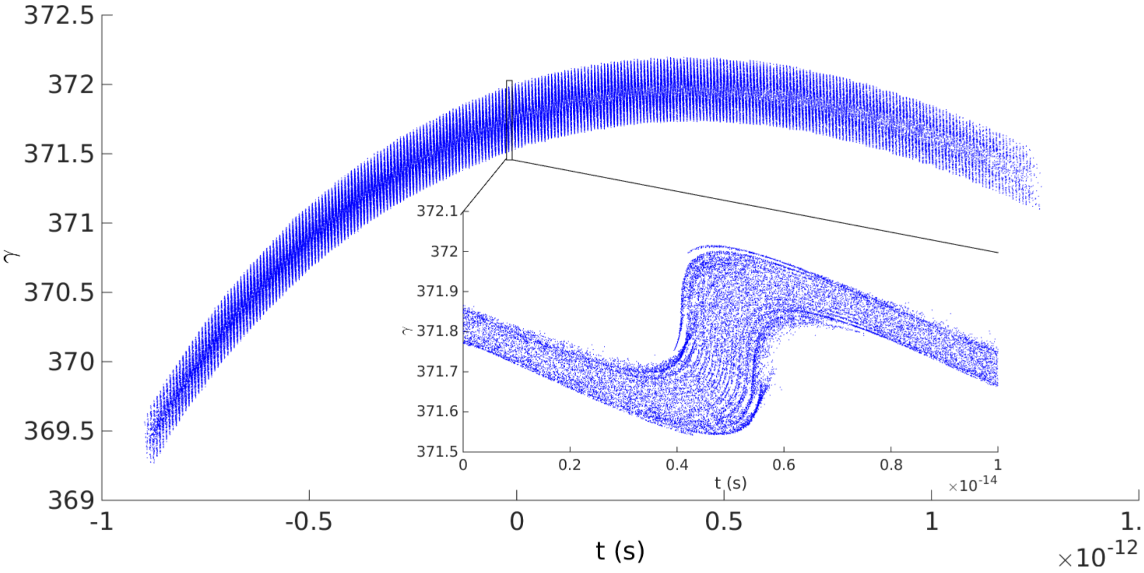


Figure 5.17: The simulated longitudinal EEHG phase space immediately prior to the VISA undulator. The inset shows the region near $t = 0$ wherein the detailed EEHG structure is resolved.

along the lattice are shown in figure 5.16.

The longitudinal phase space immediately prior to the VISA undulator is shown in figure 5.17. The overall curvature due to the X-band accelerating cavities is visible, although the large chirp which was introduced by the first cavity (X1, figure 5.15) has been mostly removed by the second accelerating structure (X2). The inset shows the detailed EEHG structure for a small region near the center of the beam.

At this point the particle distribution is handed off to PUFFIN to compute the undulator radiation from the VISA. The PUFFIN simulation contains 15 longitudinal grid nodes per radiation wavelength (≈ 114 nm) and performs 20 integration steps per undulator period. The transverse grid mesh used for the radiation is 45×45 nodes and extends to $\pm 9 \times \sigma_{x,y}$. The radiation is diffracted after the integration through every undulator period.

The resulting radiation spectrum, integrated over the transverse plane, is shown in figure 5.18. The spontaneous radiation spectrum obtained by an otherwise identical simulation with the modulating lasers turned off was subtracted from the lasers-on spectrum.

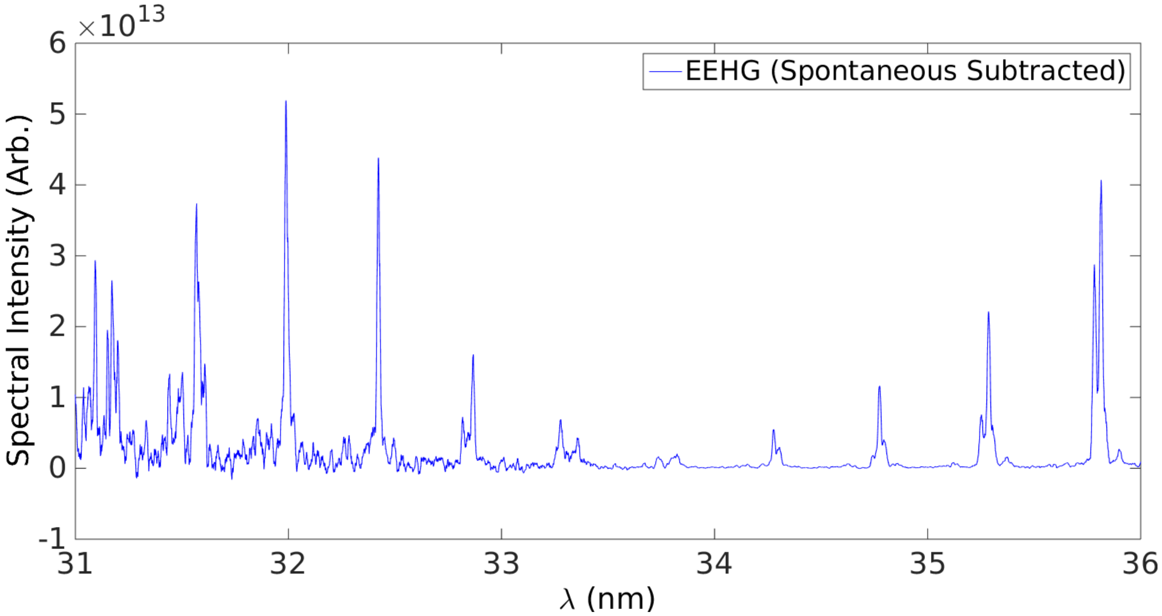


Figure 5.18: The VISA radiation spectrum in the vicinity of 32 nm as simulated in PUFFIN. The spontaneous background radiation (with no laser modulations) has been subtracted off.

The simulated harmonic envelope of the EEHG peaks in figure 5.18 is slightly different from the one measured in the experiment (figure 4.5). This is likely caused by slightly differing chicane and laser parameters which were allowed to vary in a narrow range for the start-to-end simulation. It is also possible that a slightly different electron beam energy chirp and curvature contributed to the different optimal configurations.

Nevertheless, the start-to-end simulation provides a representative spectrum for those obtained in the EEHG experiments at NLCTA. In particular, the ratio of coherent emission to spontaneous emission (not shown) for the highest harmonics is only a factor of a few in both the simulation and experiment. This initially surprising experimental result is confirmed by the simulations. The time-of-flight emittance effect, discussed in section 4.1.1, likely causes large portions of the beam to de-bunch prematurely. This de-bunching happens both in the focusing lattice leading up to the VISA and in the channel of the VISA itself, and seems to ultimately limit the attainable high harmonics at NLCTA.

It is clear from these simulations that the production of 32 nm coherent radiation at the current NLCTA facility pushes the envelope of what is possible with that setup. A more detailed study, or an effort to extend to even lower wavelengths, would require modification of one of the factors which enter into equation 4.10. Perhaps the easiest knob to tune is the focusing lattice, embodied in the factor $\langle \gamma_T \rangle$. A detailed study of the focusing lattice for the final Echo-75 data was not performed however, so it is unknown how close to the optimal configuration it may have been. Nevertheless, the VISA undulator focusing is the one aspect of the lattice which cannot be substantially altered. Similarly, barring the introduction of a new, state-of-the-art electron gun, it is unlikely that the slice emittance could be lowered substantially.

The final tool, therefore, is to use the adiabatic damping of the beam emittance with energy to compensate. An increase in the beam energy also has the added effect of radiating on a lower undulator harmonic. While there are several practical barriers to implementing a beam energy increase at the NLCTA, it seems clear from both simulations and the experiments that this would be the most promising avenue to extend the capability to produce high harmonic radiation.

Chapter 6

Stochastic Coherent Synchrotron Radiation

> If you don't know where you're going, you'll end up somewhere else.
>
> — Yogi Berra

Coherent synchrotron radiation (CSR) is an important effect which needs to be considered when designing modern FEL accelerators. The simple one-dimensional, steady state CSR result was discussed in section 2.3.1, whereby an electron traversing a bend creates a wakefield which can change the energy of other particles copropagating with it. This one-dimensional theory is an excellent approximation in most cases to the effect in modern machines, with only small corrections in the shape of the wakefield arising for bunches with full 3D structure.

In practice, the full CSR-field is often computed in a semi-analytical fashion, by convolving the numerically computed line density of the electron bunch with the analytically derived one-dimensional wakefunction [132] [133]. This approach accurately produces the one-dimensional effect, but due to the analytic nature of the wakefunction employed, typically does not capture all physical effects. Various numerical schemes introduce more complexity [134] [135] or numerical efficiency [136], but generally proceed along similar lines and make various approximations to keep the computation manageable.

Recently, however, advances in computing power have made it possible to brute force solve the Liénard-Wiechert equations for a realistic number of electrons ($\approx 10^9$) tranversing a bend magnet [137] [138]. These simulations make no assumptions on the nature of the wake, and can be performed with one with one macroparticle per electron in order to properly simulate any physical shot noise effects. Curiously, it was discovered that the longitudinal electric field due to CSR became noisy at high energies [137], with its mean value being given by the simple 1D expression and a standard deviation about this value which seems to grow quickly with γ. Since the simulations are one-to-one, this noise is not numerical, and instead is due to the granular nature of the particle distribution. In addition, the growth with γ, hypothesized in [137] to be as γ^2, is alarming as it has the potential to become a dominant source of scatter for high energy electron beams. We therefore endeavored to uncover the source for this noisy CSR effect, and understand any role it may play in limiting future accelerators.

Before delving into the mathematics, we first briefly give the form of the following derivation. By carefully examining the 3-dimensional CSR field, we identify a narrow cone of radiation whose dimensions shrink rapidly with γ, while the intensity of this radiation cone grows with γ. By carefully considering the volume of this region, we derive the mean number radiation cones a centrally-located electron interacts with. Due to simple counting statistics, there is then a fluctuation about this mean value given by the square root of the number of expected interactions. These counting statistics fluctuations are found to be ultimately responsible for the noise observed in [137]. After describing this field variation, we then extend the analysis to a diffusive model in which, as an electron traverses the bend, it receives many longitudinal 'kicks' from this field variation, and therefore performs a diffusion in energy. These are the essential physical mechanisms which we will now describe in detail.

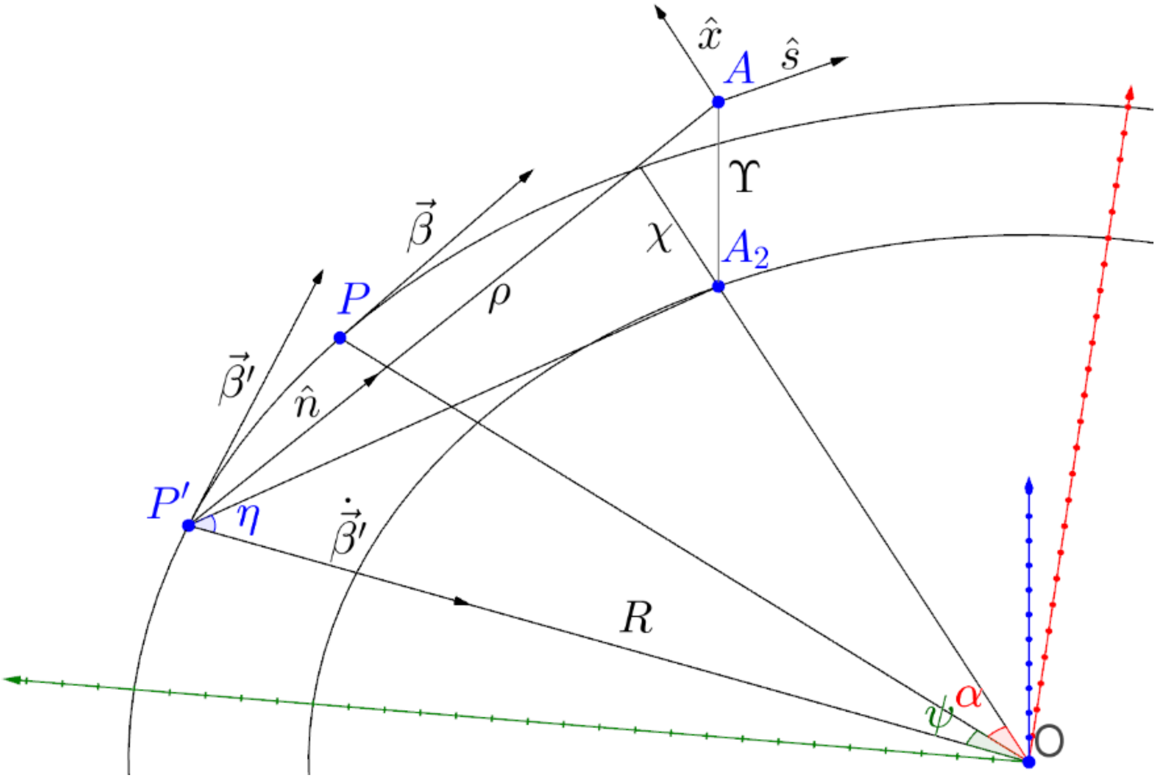


Figure 6.1: The 3D coordinate system used to describe the CSR radiation. The radiation is emitted at retarded position P' and received at the observation point A (which has its 2D projection A_2). The observation point is offset from the nominal circular trajectory by horizontal coordinate χ and vertical coordinate Υ. The vectors $\hat{s}$ and $\hat{x}$ respectively represent the tangent and perpendicular to the observation point trajectory.

6.1 A Three-Dimensional CSR Model

We begin with a three-dimensional model for the CSR effect based on the Liénard-Wiechert fields and extends the two-dimensional model by Huang, Kwan, and Carlsten [135]. A similar model was also developed recently to handle off-momentum particles by Cai [139]. The model involves electrons moving on perfectly circular tracks created by an ideal dipole magnet. In this model, there is no momentum in either the transverse or vertical direction. The coordinate system used to describe this motion, and subsequent radiation, is shown in figure 6.1. In these coordinates radiation is emitted by the particle $\mathcal{P}$ at the retarded position P' at time t', and received by particle $\mathcal{A}$ at position A and present time t, at which moment $\mathcal{P}$ is now at P. The position of particle $\mathcal{A}$ is described then by the coordinates (α, x, y), where $x \equiv \chi/R$ and $y \equiv \Upsilon/R$.

The radiation field can be computed via the Liénard-Wiechert fields at location $\vec{x} = (\alpha, x, y)$,

$$\vec{E}_{\text{rad}}(\vec{x}, t') = \frac{1}{4\pi\epsilon_0}\frac{q}{c}\left[\frac{\hat{n} \times \left(\left(\hat{n} - \vec{\beta}\right) \times \dot{\vec{\beta}}\right)}{\rho\left(1 - \vec{\beta} \cdot \hat{n}\right)^3}\right]_{\text{ret}} . \tag{6.1}$$

This is only the radiation component of the field, as we are not concerned with the velocity component which is only a short-ranged effect. The vector arithmetic is facilitated by the following geometric relations,

$$\hat{n} \cdot \dot{\vec{\beta}}' = \frac{\beta^2 c}{R}\left(\frac{\beta}{2\psi}\left[1 + \frac{\psi^2}{\beta^2} - y^2 - (1+x)^2\right]\right), \tag{6.2}$$

$$\hat{n} \cdot \vec{\beta}' = \beta\sqrt{1 - \left(\frac{y\beta}{\psi}\right)^2}\sin(\eta), \tag{6.3}$$

$$\hat{n} \cdot \hat{s} = \sin(\alpha + \psi + \eta) + \sqrt{1 - \left(\frac{y\beta}{\psi}\right)^2} - 1, \tag{6.4}$$

$$\hat{n} \cdot \hat{x} = -\sqrt{1 - \left(\frac{y\beta}{\psi}\right)^2}\cos(\alpha + \eta + \psi), \tag{6.5}$$

$$\hat{n} \cdot \hat{y} = \frac{y\beta}{\psi}, \tag{6.6}$$

$$\vec{\beta}' \cdot \hat{s} = \beta\cos(\alpha + \psi), \tag{6.7}$$

$$\vec{\beta}' \cdot \hat{x} = \beta\sin(\alpha + \psi), \tag{6.8}$$

$$\dot{\vec{\beta}}' \cdot \hat{s} = \frac{\beta^2 c}{R} \sin(\alpha + \psi), \tag{6.9}$$

$$\dot{\vec{\beta}}' \cdot \hat{x} = -\frac{\beta^2 c}{R} \cos(\alpha + \psi), \tag{6.10}$$

where we also have the following geometric relations for η,

$$\cos(\eta) = \frac{(1+x)^2 + y^2 - 1 - \frac{\psi^2}{\beta^2}}{-2\sqrt{\frac{\psi^2}{\beta^2} - y^2}}, \tag{6.11}$$

$$\sin(\eta) = \sin(\alpha + \psi) \frac{1+x}{\sqrt{\frac{\psi^2}{\beta^2} - y^2}}. \tag{6.12}$$

Finally, the angle ψ represents the retarded time condition, and follows the relation $\psi = \beta\rho/R$. It is related to the other physical coordinates (α, x, y) through the following transcendental equation,

$$\frac{\psi^2}{\beta^2} = 1 + y^2 + (1+x)^2 - 2(1+x)\cos(\alpha + \psi). \tag{6.13}$$

It is this transcendental equation which creates the bulk of the difficulty in CSR calculations, as it cannot be solved analytically. Various approximations, such as expanding out the cosine, yield solutions with small domains of applicability, but this procedure is increasingly less helpful for a three-dimensional beam. We therefore leave the angle ψ as implicitly defined by equation 6.13.

The various electric field components can then be computed as,

$$
\begin{aligned}
E_s =& \frac{q\beta^3}{4\pi\epsilon_0 R^2}\frac{1}{\psi\left(\frac{\beta^2(x+1)\sin(\alpha+\psi)}{\psi}-1\right)^3}\Bigg[\sin(\alpha+\psi)\left(1-\frac{\beta^2(x+1)\sin(\alpha+\psi)}{\psi}\right)\\
&-\frac{\beta}{2\psi}\left(\frac{\psi^2}{\beta^2}-(x+1)^2-y^2+1\right)\left(-\beta\cos(\alpha+\psi)-\frac{\beta\sin(\alpha+\psi)\left(-\frac{\psi^2}{\beta^2}+x^2+2x+y^2\right)}{2\sqrt{\psi^2-\beta^2y^2}}\right.\\
&\left.+\frac{\beta(x+1)\sin(\alpha+\psi)\cos(\alpha+\psi)}{\sqrt{\psi^2-\beta^2y^2}}+\sqrt{1-\frac{\beta^2y^2}{\psi^2}}-1\right)\Bigg],
\end{aligned}
\tag{6.14}
$$

$$
\begin{aligned}
E_x =& \frac{q\beta^3}{4\pi\epsilon_0 R^2}\frac{1}{\psi\left(\frac{\beta^2(x+1)\sin(\alpha+\psi)}{\psi}-1\right)^3}\Bigg[\cos(\alpha+\psi)\left(\frac{\beta^2(x+1)\sin(\alpha+\psi)}{\psi}-1\right)\\
&+\frac{1}{4\beta^2\psi^2}\left(\beta^2\left(x(x+2)+y^2\right)-\psi^2\right)\Big(2\beta^2\sin(\alpha+\psi)\left((x+1)\sin(\alpha+\psi)-\psi\right)\\
&+\cos(\alpha+\psi)\left(\beta^2\left(x(x+2)+y^2\right)-\psi^2\right)\Big)\Bigg],
\end{aligned}
\tag{6.15}
$$

$$
E_y = \frac{q\beta^3}{4\pi\epsilon_0 R^2}\frac{y\left(\psi^2-\beta^2\left(x(x+2)+y^2\right)\right)}{2\left(\psi-\beta^2(x+1)\sin(\alpha+\psi)\right)^3}, \tag{6.16}
$$

where $E_s = \vec{E}\cdot\hat{s}$, $E_x = \vec{E}\cdot\hat{x}$ and $E_y = \vec{E}\cdot\hat{y}$. While messy, these expressions allow a reasonable computation of the CSR fields to be performed.

A noteworthy aspect of these fields, first identified in Ref. [135], is the denominator which gives rise to a sharp, trough-like feature primarily in the $+x, -\alpha$ direction. The two-dimensional cut of this trough feature for the longitudinal field E_s is shown in figure 6.2 in scaled coordinates where $\tilde{E} = E/(q/(4\pi^2\epsilon_0\gamma^4))$, $\tilde{x} = x/\gamma^2$, and $\tilde{\alpha} = \alpha/\gamma^3$. Due to the presence of the same denominator factor, a similar but not identical trough feature exists in the transverse x and y fields as well. Introducing the scaled vertical coordinate $\tilde{y} = y/\gamma^2$, we visualize the full nature of the 'tube' by plotting the $\tilde{E}_s = -2$

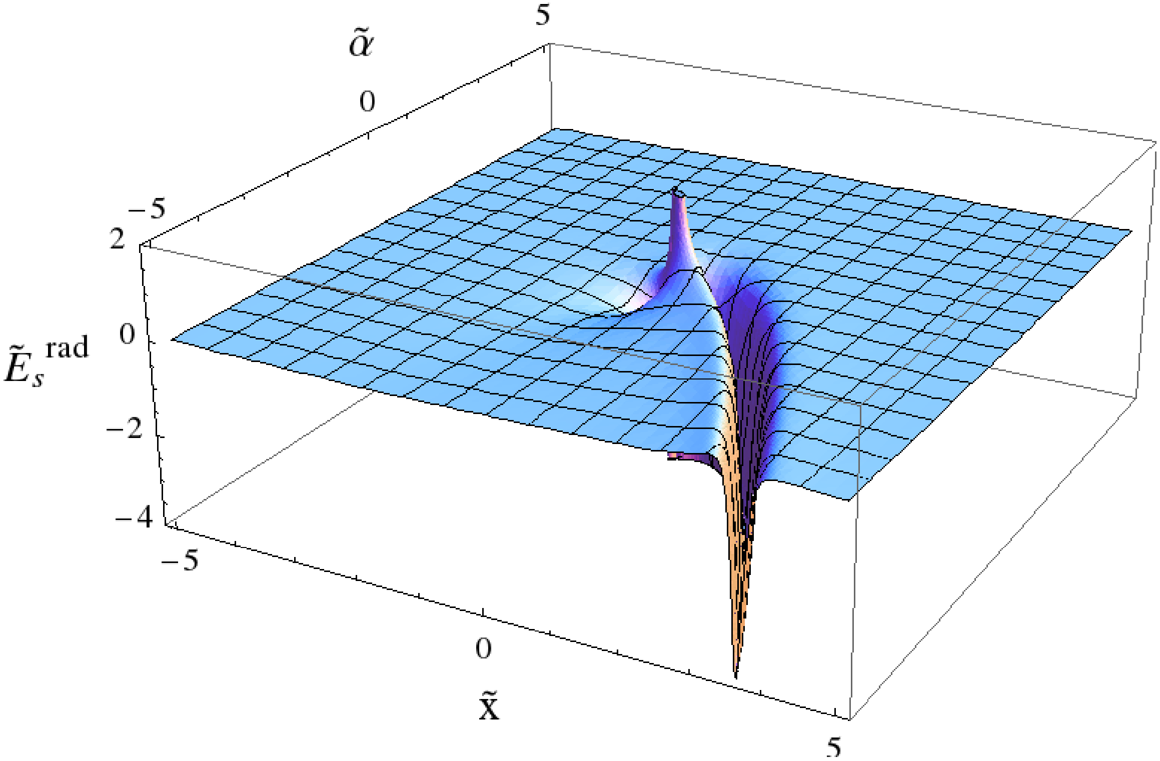


Figure 6.2: The trough feature viewed in the $(\tilde{x}, \tilde{\alpha})$ plane in scaled coordinates. The scaled minimum is at $\tilde{E}_s^{\text{rad}} = -4$. Note that there are significant, close-range fields near $\tilde{x} = \tilde{\alpha} \approx 0$ which we do not discuss in detail.

contour in figure 6.3 (as seen in equation 6.18, the low point is $\tilde{E}_s = -4$). Already we can see the opening in the vertical dimension as well as the constant width in the α dimension.

As this trough feature is an interesting long ranged feature of the CSR field, we study it in more detail. To begin with, its location is easily found in ψ to be given by,

$$\psi_{\text{T}} = \beta\sqrt{x(x+2)}, \tag{6.17}$$

And the value of the field at this trough $(y = 0)$ is found to be,

$$E_s^{\text{T}} \approx \frac{-q\beta^2\gamma^4}{\pi\epsilon_0 R^2 (1+x)}. \tag{6.18}$$

Since in general $x \ll 1$, the trough can be considered infinite in extent and with nearly

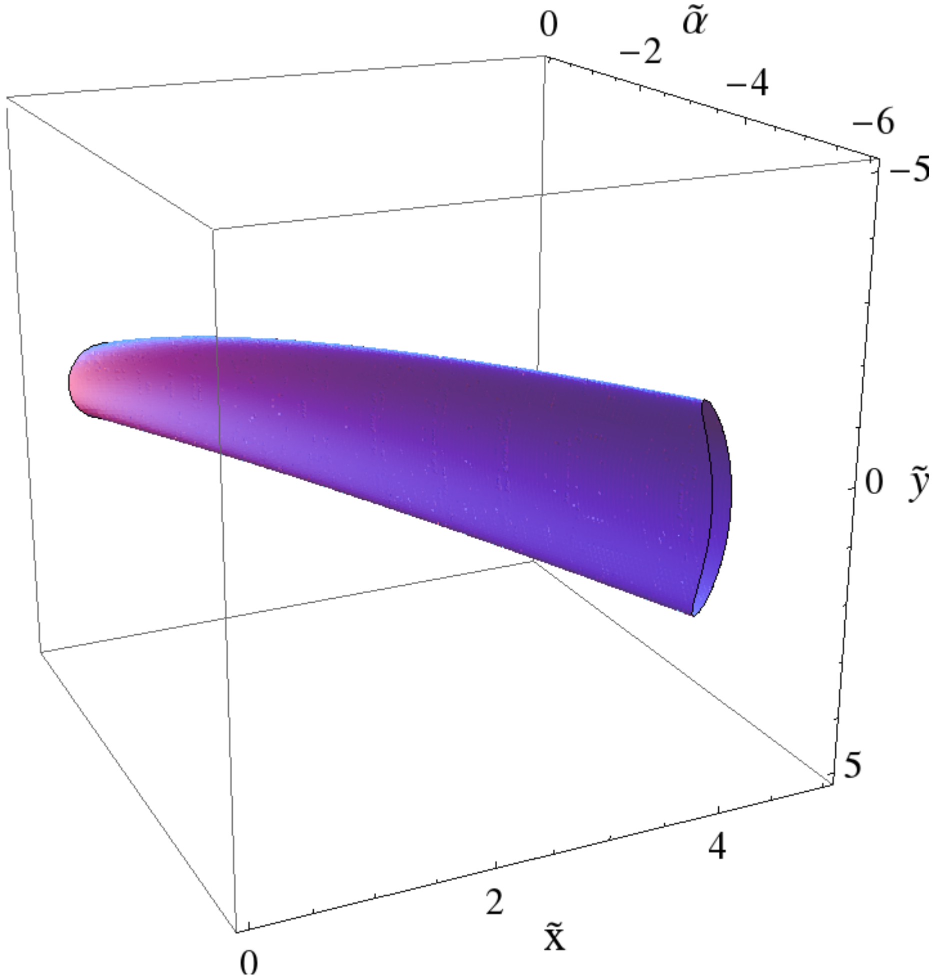


Figure 6.3: A three-dimensional contour ($\tilde{E}_s = -2$) of the radiation tube in scaled coordinates.

constant amplitude which scales as γ^4. This simple observation yields a potential path towards explaining the field variations observed in numerical simulations. If there is a region in field space with finite volume which scales inversely with γ to some power, as γ is increased fewer and fewer particles populate this region. Due to simple counting statistics, there will be a variation in the realized field due to the finite number of particles N_p sampling the region, which will scale as $\sqrt{N_p}$. Depending on the exact scaling of the volume and amplitude of this region, it is possible to get a noise term which grows rapidly with γ simply due to particle granularity.

We take this tack to explain this phenomenon. What we would like to compute is,

$$\int_{\text{Trough}} E_s. \tag{6.19}$$

However, due to the complicated form of E_s, analytically computing this integral is out of the question. Instead, we approximate the value of E_s as constant in the trough at E_s^{T} from equation 6.18. There is of course a geometrical factor accounting for the real non-constancy which we refer to as g. The task is then reduced to finding the volume of the trough region in 3D. By examining the bounding countours of the trough volume, we will compute its volume and hence the expected number of particles in this region, thereby deriving the variance as described above.

As a preliminary, we examine the net contribution of the trough to the electric field, we can rewrite in terms of the coordinate η and compute, at fixed $x_0 > 0$,

$$\int_{-\infty}^{\infty} \tilde{E}_s(\alpha, x_0, 0)d\alpha = \int_{\pi/4+x_0/2}^{3\pi/4-x_0/2} \tilde{E}_s(\eta, x_0, 0)\frac{d\alpha}{d\eta}d\eta. \tag{6.20}$$

Note that the limits of $\pm\infty$ should not be taken literally due to the 2π periodicity in α – in practice, all contributions have decayed by $\alpha \sim \pm 1$. The transformation from α to η variables allows us to bypass the transcendental condition on ψ and compute this integral directly. It turns out that the inner trough portion of this integral from $\eta = [\frac{\pi}{2} - \frac{1}{\gamma}, \frac{\pi}{2} + \frac{1}{\gamma}]$ contributes -2γ, while the wings as $\alpha \to \pm\infty$ each contribute γ and the entire integral vanishes. Therefore, for a uniform beam, the trough feature does not contribute to the radiation field. This result is complimentary to the well-known 1D result that $\int_{-\infty}^{\infty} E_s(\alpha, 0, 0)d\alpha = 0$, that a uniform beam does not radiate (this is trivially true for the 1D case as evidenced by the total derivative in equation 2.102). Even for a Gaussian beam, if the longitudinal extent is much larger than the longitudinal scale $\sim \gamma^{-3}$, the net contribution is small, and the primary function of the trough is therefore to contribute to the variance around the deterministic CSR solution.

Turning now to the question of the volume of this region, in the (α, x) plane this is easily found [135] where $\sin\eta = \beta$. One finds that the width in α , $\Delta\alpha_{\mathrm{T}} = 4/(3\gamma^3)$, and we can consider the trough infinite in x extent. To investigate the vertical behavior, we restrict to the ‘bottom’ of the trough by setting $\alpha = \alpha_{\mathrm{T}} = \arccos\left[\frac{1}{1+x}\right] - \beta\sqrt{x(x+2)}$, and note that it is bounded in α by $\alpha_{\pm}(x) = \alpha_{\mathrm{T}} \pm \frac{2}{3\gamma^3}$.

At this point, we make the qualitative distinction between a ‘large’ and ‘small’ beam in vertical extent. For a large beam, we expand out the electric field in the vicinity of the trough to second order in y gives the following approximate bound on the vertical extent of the trough,

$$y_{\pm}(x) \approx \pm 2\sqrt{\frac{2}{3}}\frac{\sqrt{x}}{\gamma}. \tag{6.21}$$

Thus, the vertical extent scales as γ^{-1} for a large beam. By contrast, a ‘small’ beam is not large enough to sufficiently sample this vertical extent, and it is as though the trough is constant in the vertical dimension. The precise meaning of this qualitative distinction will now be made clear. In the context of classical synchrotron radiation, this scaling is simply a different way of seeing the $1/\gamma$ opening angle of the radiation [25].

We now transition the discussion from the electric field due to a single electron to the contribution from the entire electron bunch. The electric field given above is the electric field due to an electron centered at the origin, and the total contribution at any point can be computed as,

$$\begin{aligned} E_s^{TOT}(\alpha, x, y) = &\sum_{i=1}^{N_p} E_s(\alpha_i - \alpha, x_i - x, y_i - y) \\ &\xrightarrow{\text{Continuum}} \int \rho(\alpha', x', y') E_s(\alpha' - \alpha, x' - x, y' - y) d\alpha' dx' dy', \end{aligned} \tag{6.22}$$

where in passing to the continuum limit we introduce the density $\rho(\alpha, x, y)$. For a symmetric particle distribution, however, $\rho(-\alpha, -x, -y) = \rho(\alpha, x, y)$. The above integral can therefore be understood by integrating the particle distribution over the single-electron electric field. In this language, we therefore refer to particles being ‘contained within the trough’, even though in more physical terms, what is meant is ‘particles whose trough intersects the point of interest’. This symmetry argument allows for the much simpler computation of the total fraction of electrons contained within the trough region.

Assuming a Gaussian distribution defined by the scaled rms quantities $\sigma_x = \sigma_{x,R}/R$ and likewise for y and α, where $\sigma_{x,R}$ is the physical rms width, we have the fraction computed as,

$$f = \int_0^\infty dx \int_{y_-(x)}^{y_+(x)} dy \int_{\alpha_-(x)}^{\alpha_+(x)} d\alpha \rho(x, y, \alpha), \tag{6.23}$$

where for the Gaussian distribution, in terms of the scaled variables, centered at $x = y = \alpha = 0$,

$$\rho(x, y, \alpha) = \frac{1}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_\alpha} e^{-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}-\frac{\alpha^2}{2\sigma_\alpha^2}}. \tag{6.24}$$

For large γ, the α integral can be simplified by noticing that $\alpha_\pm(x) \approx -\frac{2}{3}\sqrt{2}x^{3/2} \pm \frac{2}{3\gamma^3}$. Remarkably, this integral admits an analytic, exact solution in terms of Bessel and hypergeometric functions with only this mild assumption on $\alpha_\pm(x)$, and is given by,

$$\begin{aligned} f = \frac{1}{\sqrt{2\pi}\sigma_x} \Bigg\{ & \frac{2\sqrt[4]{2}\Xi^{-1/4}\sigma_x \mathrm{erf}\left(\frac{\sqrt{2}}{3\gamma^3\sigma_\alpha}\right)}{9\sqrt{3\pi}} \Bigg[9\sqrt{2}\Gamma\left(\frac{3}{4}\right) {}_2F_2\left(\frac{1}{4}, \frac{3}{4}; \frac{1}{2}, \frac{5}{4}; \frac{8\Xi^{-1}}{9}\right) \\ & - 8\Xi^{-1/2}\Gamma\left(\frac{5}{4}\right) {}_2F_2\left(\frac{3}{4}, \frac{5}{4}; \frac{3}{2}, \frac{7}{4}; \frac{8\Xi^{-1}}{9}\right) \Bigg] \\ & - \frac{8\sqrt{2}\Xi^{-1}\sigma_x^3\sigma_y^2}{729\gamma\sigma_\alpha^3} \Bigg[-(16\Xi^{-1} + 63) I_{\frac{1}{4}}\left(\frac{4\Xi^{-1}}{9}\right) + 16\Xi^{-1}\left(I_{\frac{3}{4}}\left(\frac{4\Xi^{-1}}{9}\right) - I_{\frac{5}{4}}\left(\frac{4\Xi^{-1}}{9}\right)\right) \\ & + (16\Xi^{-1} + 45) I_{-\frac{1}{4}}\left(\frac{4\Xi^{-1}}{9}\right) \Bigg] e^{\frac{4\Xi^{-1}}{9} - \frac{2}{9\gamma^6\sigma_\alpha^2}} \Bigg\} \end{aligned} \tag{6.25}$$

Where we have introduced the (dimensionless, as all σ are scaled variables) parameter,

$$\Xi \equiv \frac{\gamma^4\sigma_y^4}{\sigma_x^2}. \tag{6.26}$$

In light of equation 6.21, the case of $\Xi \ll 1$ is essentially a 2-D beam, where the vertical extent of the radiation trough plays no role over the entire beam. By contrast, for $\Xi \gg 1$, the beam's vertical size is much larger than the radiation trough – as far as the radiation trough is concerned, this is a beam which is infinite and uniform in y extent. In these limits of large and small Ξ we have the simple solutions for $\gamma \gg 1$,

$$f = \begin{cases} \frac{16\sqrt[4]{2}\sqrt{\sigma_x}\left(27\sqrt{2}\sigma_\alpha^2\Gamma\left(\frac{7}{4}\right)-16\sigma_x^3\Gamma\left(\frac{13}{4}\right)\right)}{243\sqrt{3}\pi^{3/2}\gamma^4\sigma_\alpha^3\sigma_y} & \Xi \gg 1 \\ \frac{1}{2}\mathrm{erf}\left(\frac{1}{6\sqrt{2}\gamma^3\sigma_\alpha}\right) & \Xi \ll 1 \end{cases}. \tag{6.27}$$

The incredibly complicated expression of equation 6.25 is reduced to this extremely simple form in these Ξ limits. Of particular note is the fact that $f(\Xi \gg 1) \sim \gamma^{-4}$ while $f(\Xi \ll 1) \sim \gamma^{-3}$. Given the interpretation of Ξ as the effective dimensionality of the beam, these scaling results agree with the preliminary trough volume study.

Having described how the volume of the trough is populated by the electrons, we now can state that (again, due to simple counting statistics) there will be a variance in the field due to the number of electrons populating it. In particular, the variance in the longitudinal electric field can be expressed as,

$$\sigma_{E_s} = gE_s^{\mathrm{T}}\sqrt{fN_p}, \tag{6.28}$$

where N_p is the total number of electrons, E_s^{T} is the value of the E-field at the bottom of the trough (equation 6.18), and g is a geometrical factor to account for its non-constancy over the trough ($g = 1/4$ in the assumption that the trough is a 2-D pyramid, $g = 4/9$ for a parabola, and $g \approx 0.73$ for a Gaussian to $\pm 1\sigma$, etc.).

We briefly recapitulate how we arrived at equation 6.28. By assuming the electric field is constant across the trough region, we reduce the problem to finding the fraction of electrons contained within the trough: f. Having computed this for a Gaussian distribution, there is a counting variance proportional to the strength of the field and the square root of the number of particles $\sqrt{N_p f}$ in the trough. Thus we are able to immediately write down equation 6.28.

With equation 6.28 in hand, and since $E_s^{\mathrm{T}} \sim \gamma^4$, we can immediately write down

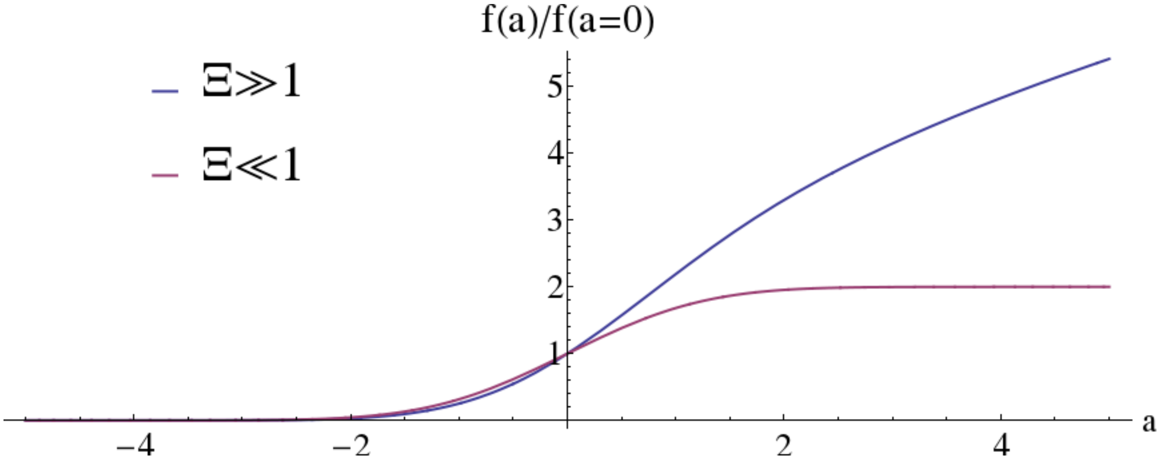


Figure 6.4: The normalized electron trough fraction as a function of $a = x_0/\sigma_x$ for the cases $\Xi \gg 1$ and $\Xi \ll 1$.

the scaling with energy in the different Ξ regimes,

$$\sigma_{E_s} \sim \begin{cases} \gamma^2 & \Xi \gg 1 \\ \gamma^{2.5} & \Xi \ll 1 \end{cases}. \tag{6.29}$$

This is in agreement with the results of [137], wherein it was computationally shown that $\sigma_{E_S} \sim \gamma^2$ with simulations in the $\Xi \gg 1$ regime. This preliminary result supports the conclusion that this variance could potentially become a problem at high energies.

These results, due to the form of the integral 6.23, have only applied to the center of the electron bunch. By shifting $\rho(x, y, \alpha) \to \rho(x - x_0, y - y_0, \alpha - \alpha_0)$ we can compute the quantity f for off-center particles. In practice, we are most interested in those with $x_0 \neq 0$, so we restrict the analysis to $\rho(x - x_0, y, \alpha)$. As before, analytic solutions are available in the large and small Ξ limit, and can be expressed in terms of the ratio $a \equiv x_0/\sigma_x$. As with equation 6.25 however, their form is not particularly illuminating. Particularly useful solutions are available in the realistic limit that $\sigma_\alpha^2 \gg \sigma_x^3$ as,

$$\frac{f(a)}{f(a=0)} = \begin{cases} e^{-\frac{a^2}{2}} \left({}_1F_1\left(\frac{3}{4}; \frac{1}{2}; \frac{a^2}{2}\right) + \frac{\sqrt{2} a \Gamma\left(\frac{5}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} \, {}_1F_1\left(\frac{5}{4}; \frac{3}{2}; \frac{a^2}{2}\right) \right) & \Xi \gg 1 \\ \operatorname{erf}\left(\frac{a}{\sqrt{2}}\right) + 1 & \Xi \ll 1 \end{cases}. \tag{6.30}$$

Due to the complicated nature of the hypergeometric functions, the normalized electron fraction is plotted in figure 6.4. The general trend is that $f(a)$ is larger for a positive, and quickly goes to zero for a negative, implying that the noise contribution to the CSR field is larger for electrons on the outside of the bend than those on the inside. Intuitively, given the field of figure 6.2, since the trough feature extends only to positive x, those electrons on the outside are exposed to more of these features, thus leading to a greater noise contribution. Notably, the noise contribution grows much quicker in the $\Xi \gg 1$ limit, whereas for $\Xi \ll 1$, the noise levels off around $a \approx 2$.

This distinction is in accord with intuition based on the different Ξ limits and the opening angle of the radiation cone. For $\Xi \ll 1$, all electrons are essentially in the $y = 0$ plane, and therefore the opening of the radiation in the y plane has no effect and only increasing the number of electrons visible through an increase in x matters. By contrast, for $\Xi \gg 1$, there are $y \neq 0$ electrons which sit below or above the receiving particle. One must therefore go far enough out in x that their $1/\gamma$ opening angles intersect the $y = 0$ plane for their influence to be felt. It is this distinction that gives rise the steady increase in $f(a)$ with a for the $\Xi \gg 1$ case.

6.2 Fluctuations in Particle Energy and Emittance Growth

Having fully described the variance of the longitudinal field, the question now arises as to how this field variance affects the electron distribution. In particular, equation 6.28 gives the variance due to a particular electron configuration, but does not include any information on how this configuration evolves as the entire electron bunch traverses a bend magnet.

Recall that the coordinate system of 6.1 is a co-moving coordinate system with the electrons as they traverse a bend of constant radius. In general, only the particle on the reference trajectory will remain at constant coordinates as it traverses the bend. We model the electron beam as monoenergetic, so all electrons travel on tracks of the same radius R, although their center-points are offset from one another by their initial coordinates. Due to these offset centers, their relative displacements from the reference trajectory change as the bend is traversed. In the picture of 6.2, electrons at initial coordinates (α_0, x_0) move on distorted tracks as the bend is traversed. The exact form of the tracks is computed from the simple offset-circle trigonometry. Note that we also consider a beam with zero emittance, such that the particles do not move in y.

The interpretation of this motion, however, is that electrons which originally populate the trough region will, after some distance through the bend, leave the region and be replaced by a new set of particles. The frequency with which this happens is referred to as the 'refresh rate' for the CSR trough. Intuitively, particles closer to the central particle will refresh much slower than those further away, and an average needs to be taken. The particles drift primarily in the α dimension with a 'velocity' given by simple geometry as,

$$\frac{d\alpha}{dz} \approx \frac{x_0}{R}, \tag{6.31}$$

where recall that x_0 is already scaled to the radius R, and we have omitted a small term quadratic in the offset α_0. The trough is of constant α extent $\Delta\alpha = \frac{4}{3\gamma^3}$, so we define the distance z_c it takes for a particle at x_0 to cross the trough as,

$$z_c = \frac{4R}{3\gamma^3 x_0}. \tag{6.32}$$

We now seek to define an average crossing distance over the entire electron bunch: $\bar{z} = \langle z_c \rangle$. The naive expectation for this average,

$$\bar{z} \overset{?}{=} \frac{\sqrt{2}}{\sigma_x\sqrt{\pi}} \int_0^\infty \frac{4R}{3\gamma^3 x_0} e^{-x_0^2/(2\sigma_x^2)} dx_0, \tag{6.33}$$

is formally divergent for small x. Fortunately, nature provides a lower bound cutoff for this integral. As the electron bunch moves through a bend magnet for only a finite angle Θ, and hence a finite time, those electrons with small x_0 will not have enough time to traverse the trough region. Therefore, we can define a minimum coordinate b such that through an evolution of angle Θ this particle will have enough time to traverse the trough as,

$$b = \frac{4}{3\gamma^3\Theta}. \tag{6.34}$$

This provides a lower-limit to the integral of equation 6.33 and allows its evaluation to a solution for small ratio b/σ_x,

$$\bar{z} = \frac{4R}{\sqrt{2\pi}3\gamma^3\sigma_x}\left(-\gamma_E + \log 2 - \log\frac{b}{\sigma_x}\right), \tag{6.35}$$

where $\gamma_E \approx 0.577$ is the Euler-Mascheroni constant.

This is the average arclength it takes for the electron distribution in the CSR trough to refresh. A physical meaning of $\bar{z}$ is that if we observe the longitudinal electric field at the center of the bunch as it traverses the bend, the fluctuations in that field will have a characteristic length scale $\bar{z}$.

To tie this result back to the variance of the longitudinal field for a single distribution, we approximate the dynamical motion of the electrons into and out of the trough region as a binary process. At arc position z_0, the distribution is in state $\mathfrak{S}_0$, which pulls one particular sample from the variance of equation 6.28. At position $z_0 + \bar{z}$, the distribution is now in state $\mathfrak{S}_1$, which is assumed to be independent of $\mathfrak{S}_0$, and pulls a new, uncorrelated sample from equation 6.28. Of course, the distribution $\mathfrak{S}_1$ is not independent of $\mathfrak{S}_0$, and is instead the direct temporal evolution of it. However, since we are only concerned with a small volume in space, we can treat these two as independent since those particles which have not entered or have already left give approximately zero contribution.

One caveat here is that the form of the distribution $\mathfrak{S}$ is assumed to be independent of arc position z. For example, if $\mathfrak{S}$ is a Gaussian in x and α, its standard deviations σ_x and σ_α are assumed independent of z. In reality, this is not true given the geometric model described. Given that each particle moves on circular trajectories with merely offset centers, not different radii, an initially Gaussian bunch will deform under transport. In effect, the bunch does not rotate along with the circular trajectory, and so becomes skewed as the two dimensions mix.

This situation is illustrated in figure 6.5. In this example, $\sigma_\alpha = 10^{-6}$, while

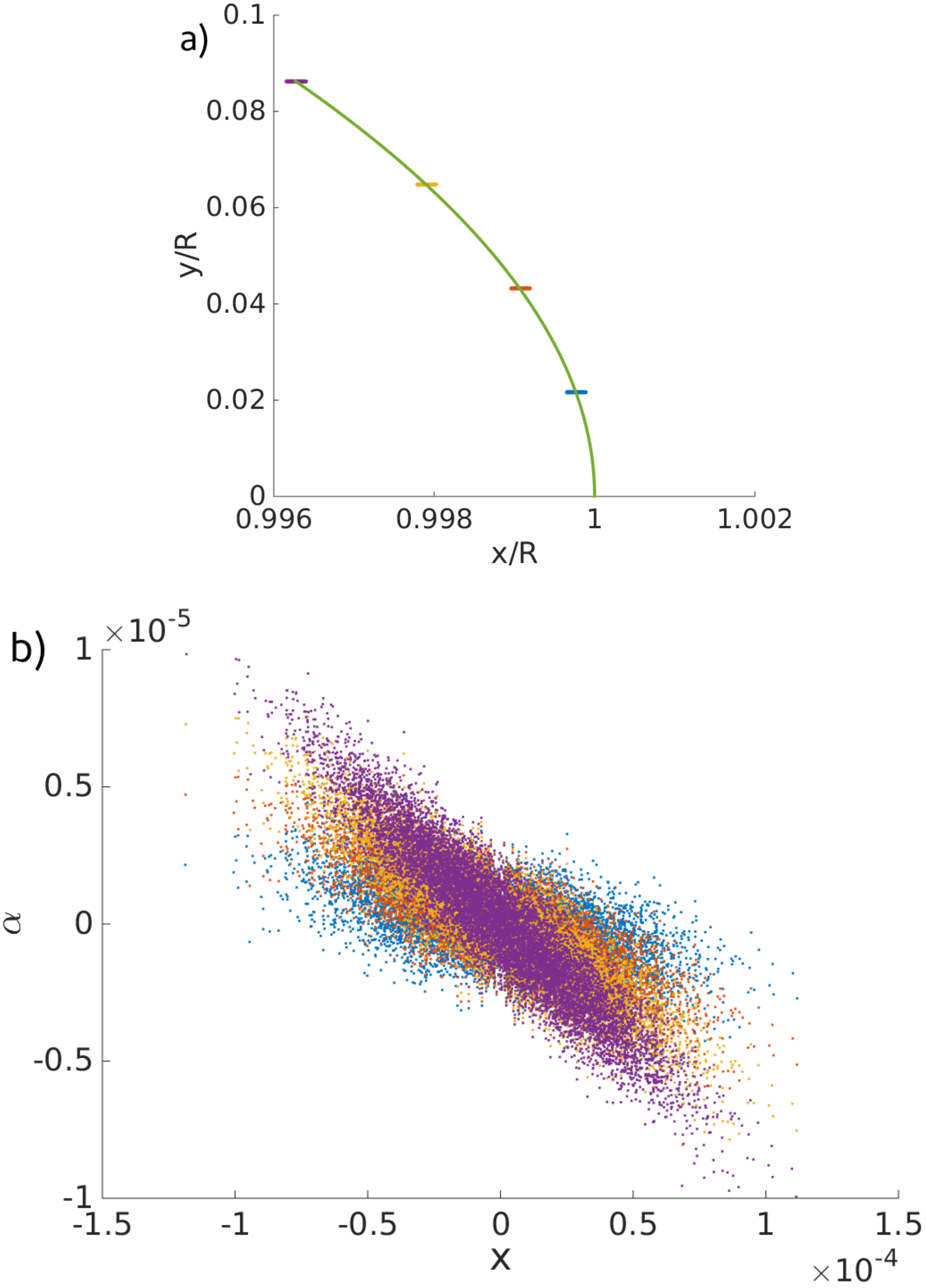


Figure 6.5: An uneven electron beam is geometrically skewed in the bend magnet. Subfigure a) shows the mean electron trajectory in green, with the superimposed electron distributions at four separate points. As the electron beam does not rotate with the track, its dimensions in the (x, α) plane distort, seen in subfigure b).

$\sigma_x = 3 \times 10^{-5}$. As the distribution distorts, in particular, σ_α becomes significantly larger, which alters the mean 1-dimensional CSR field.

This complication can be avoided if this geometric effect is kept small compared to the rms beam size. The requirement on the rms distribution can be represented by the expression,

$$\sigma_\alpha \gg \frac{\sigma_x \sin(\Theta)}{\sqrt{1 + \sigma_x^2 + 2\sigma_x \cos(\Theta)}}, \tag{6.36}$$

for a total evolution angle Θ. The geometric effect is also irrelevant under the stronger condition $\sigma_x \leq \sigma_\alpha$, and we restrict our analysis to situations in which this holds.

This model is then essentially that of a random walk with discrete step size $\bar{z}$ and step amplitude σ_{E_s}, from which we can directly write down the diffusion in energy $\mathcal{E}$ experienced by the central particle as,

$$\sigma_{\mathcal{E}} = q\sigma_{E_s}\sqrt{z\bar{z}}. \tag{6.37}$$

Again, in the language of the random walk, here $z/\bar{z}$ is the number of steps taken, while the kick per step is given by $\sigma_{E_s}\bar{z}$ (the variance is approximated as constant over the refresh distance). We briefly note that the $\bar{z}$ which defines the random walk step size is not necessarily identical to the trough-crossing $\bar{z}$ derived in equation 6.35. There may be a small discrepancy constant between the two, although equation 6.35 captures the fundamental scaling of this parameter with both σ_x and γ.

The diffusive character of this effect is evidenced by the fact that it is proportional to $\sqrt{z}$, which is of course by construction. Due to the step size from $\bar{z}$, however, the γ dependence is altered from the field variance of equation 6.28. Although the effect seemed to grow quickly with γ when observing the field fluctuations, the important physical effect, the energy change, grows only between $\gamma^{1/2}$ and γ depending on the dimensionality (Ξ) of the beam. This observation casts doubt on the initial hypothesis that this stochastic effect may become competitive with other diffusive effects such as ISR (eq. 4.1) at high energies.

Nevertheless, this fluctuation in particle energy causes an increase in beam emittance as the bend is traversed. Unlike the average CSR field effect, given in the 1D approximation by the wake 2.16, this effect cannot be removed due to clever choice of optics [52]. Assuming the emittance growth is small compared to the original emittance, this growth can be written as [16],

$$\Delta\epsilon = \frac{1}{2}\left(\frac{\sigma_{\mathcal{E}}}{\mathcal{E}}\right)^2 \langle\mathcal{H}\rangle, \tag{6.38}$$

where $\mathcal{H}$ is the well-known dispersion invariant,

$$\mathcal{H} \equiv \gamma\eta^2 + 2\alpha\eta\eta' + \beta\eta'^2, \tag{6.39}$$

where α, β, γ are all Twiss parameters, not relativistic parameters. The function $\mathcal{H}$ comes up often when discussing emittance growth in dispersive systems, and indeed the steady-state beam sizes and energy spreads in a storage ring are proportional to it [16]. Despite its relative importance, it is still known simply as 'curly-H'.

The final conclusion is therefore that the emittance growth through a bend scales somewhere between γ and γ^2 for the stochastic CSR term. By contrast, the ISR effect (eq 4.1) scales as γ^4. As before, this suggests that this new stochastic term is not of dominant importance at the high energy range for bunch compressors in X-FEL systems. However, the emittance growth will also be proportional to the beam density through the various beam sizes $\sigma_{x,y,\alpha}$ which are packed into equation 6.37. This causes the stochastic CSR effect to have more in common with Intra-Beam scattering (discussed in section 6.4) than ISR.

6.3 Comparison with Liénard-Wiechert Simulations

Having developed the analytical theory to explain the original surprising simulation results of [137], we now turn to a comparison between these results. The simulation code we use is the same as was used in the studies of [137] and [138], which is based on a massively parallel solution the Liénard-Wiechert equations. This parallelization enables the code to run efficiently on large distributed clusters, and thus model a physical number of electrons for realistic beams. The main difficulty, finding the retarded time corresponding to a given observation point, is solved by first performing a crude bisection search to approximately find the retarded time and then employing a secondary root-finding step (Newton's method) to achieve a relative accuracy in the retarded time of 10^{-10}.

In this model, the particle trajectories are all known beforehand as perfectly circular trajectories due to the bending magnetic field. Thus, the equations of motion are not solved for the electrons. Rather, their position is simply advanced along this predefined track. In this way, there is no reaction to the electromagnetic field by the electron bunch, and its distribution can be considered 'frozen' as far as analytical calculations are concerned (although, the geometric distortion effect shown in figure 6.5 is still present). The program computes the electromagnetic field resolved into velocity and radiation components on a cartesian grid at an arbitrary location. For the time being, however, we are primarily interested in the longitudinal radiation electric field for comparison with the above model.

The first point of comparison is in the variation in longitudinal electric field, which is computed from equation 6.28. We also adjust the electron bunch to live in either Ξ extreme limit to facilitate the comparison. The electron beam used has a charge of 10 pC, and is therefore composed of 6.25×10^7 electron macroparticles. The bend radius is taken to be one meter. For this study, we sample the longitudinal electric field at the bunch center only. The results of this simulation and comparison to the analytic expression from equation 6.28 are shown in figure 6.6.

The simulation with $\Xi \ll 1$ has $\sigma_x = 1 \times 10^{-5}$, $\sigma_y = 5 \times 10^{-7}$, and $\sigma_\alpha = 1 \times 10^{-5}$, while the $\Xi \gg 1$ has $\sigma_x = 5 \times 10^{-7}$, $\sigma_y = 1 \times 10^{-4}$, and $\sigma_\alpha = 1 \times 10^{-4}$. For each γ, 500 realizations of the electron distribution were used to compute the standard deviation in longitudinal electric field. The error bars in the simulation data points represent this finite sample size. The analytical curves are based on the full expression from equation 6.25 rather than the limiting expressions, and a value of $g = 0.33$ is assumed for the each calculation. Due to the extreme dimensions in the $\Xi \gg 1$ beam, simulations above $\gamma \sim 2000$ become significantly noisy and therefore the curve was only computed up to this point.

The analytical result is in extremely good agreement with the simulation data, although the arbitrary nature of g should be understood to essentially provide a scaling adjustment on these curves. Nevertheless, the fact that $g = 0.33$ is used for

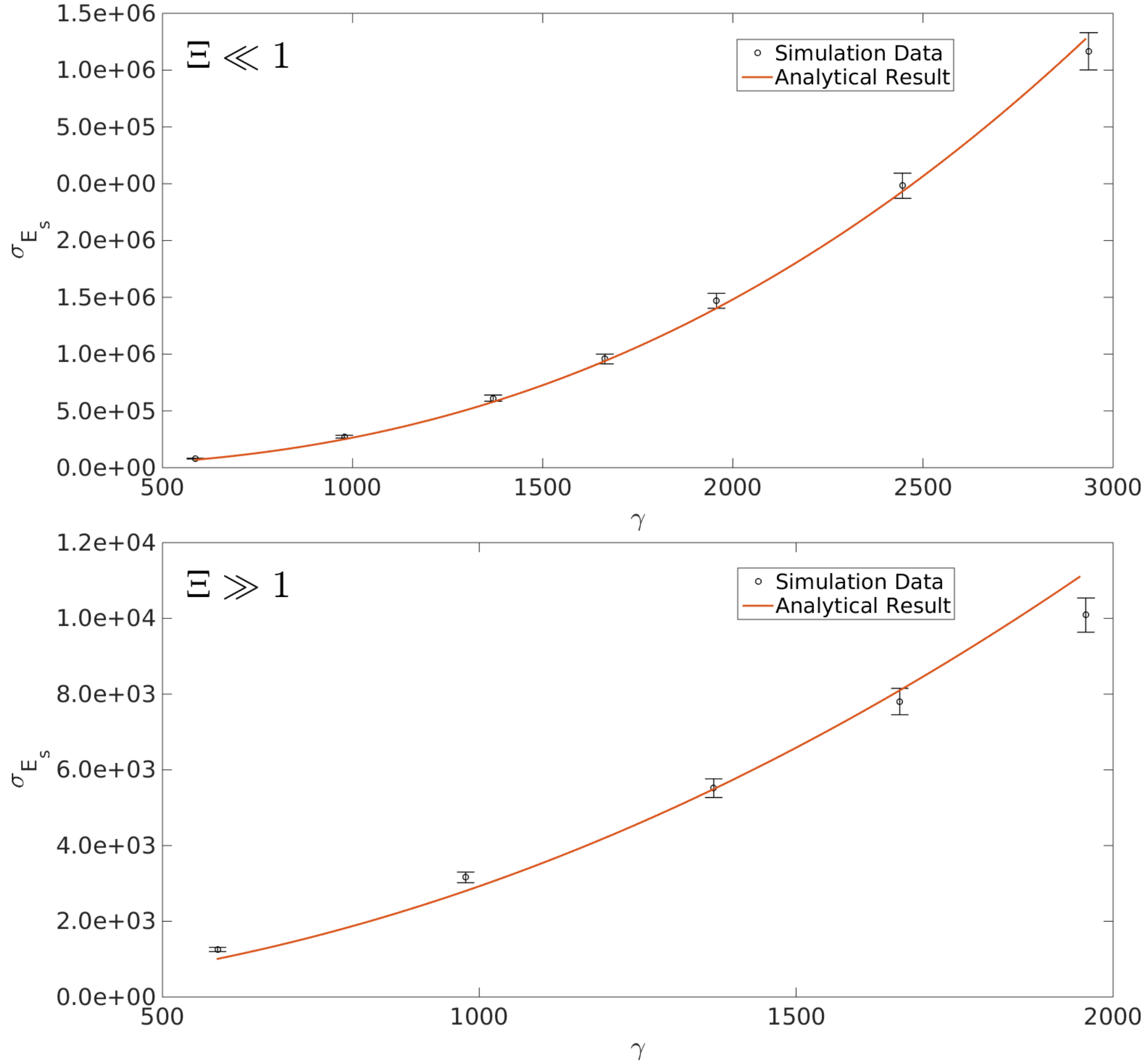


Figure 6.6: Comparison between theory and simulation for the longitudinal CSR field in both Ξ limits as a function electron beam energy γ. All other parameters are fixed. The error bars are uncertainty estimates due to the finite number of simulations (500).

both cases demonstrates that it is not a free parameter. Note, however, that the magnitude of the fluctuations for $\Xi \ll 1$ is approximately two orders of magnitude lower than for $\Xi \gg 1$, which is generally attributed to the extra factor of $\gamma^{1/2}$.

We also simulate the field in off-axis locations to test the radial dependence of equation 6.30. The results are shown in figure 6.7 for both a $\Xi \gg 1$ and $\Xi \ll 1$ case.

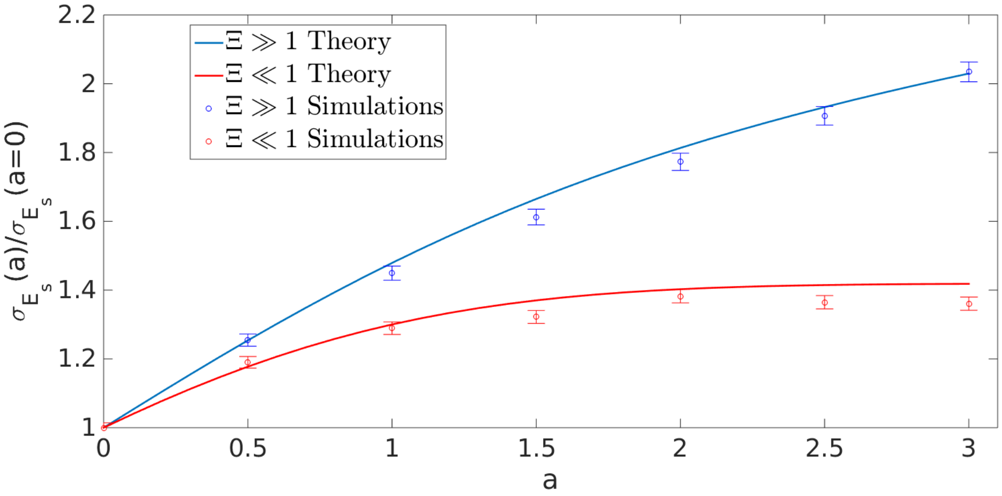


Figure 6.7: A comparison of the off-axis field variation $\sigma_{E_s}(a)$, normalized to its central value $\sigma_{E_s}(a = 0)$ for both $\Xi \gg 1$ and $\Xi \ll 1$ bunches. The error bars represent a $1/\sqrt{N_{\text{sim}}}$ variation in the results of 5000 independent simulations.

Both simulations have $E = 500$ MeV and a bend radius $R = 1$ m with 6.24×10^6 particles and $\sigma_x = \sigma_r = 10^{-5}$. To achieve the different Ξ limits, the $\Xi \ll 1$ simulation has $\sigma_y = 5 \times 10^{-7}$ while $\Xi \gg 1$ has $\sigma_y = 4 \times 10^{-5}$. The agreement between simulation and theory is excellent as much of the theoretical uncertainty, contained in the g factor of equation 6.28, is normalized out.

Having verified the analytical model for the static field case, we turn to the evolution of the field through the bend. For these simulations, we take the electron beam to have an energy 200-350 MeV and low charge 0.5 pC, which then evolves through three degrees of a 1 m bend magnet. The evolution proceeds in discrete steps of size $\delta z \approx 25$ µm, which is sufficient to resolve the noise structure at these low energies. From equation 6.35, even the highest energy simulation with $E = 350$ MeV has $\bar{z} \approx 800$ µm. The electron beam for this study has $\sigma_\alpha = \sigma_x = 1 \times 10^{-5}$, and $\sigma_y = 1 \times 10^{-6}$. For each beam energy, 100 separate runs are performed to develop statistics for resulting normalized diffusion in energy, and are compared with the analytical results in figure 6.8.

This diffusion is normalized to the mean energy of the electron bunch, which highlights the fact that it scales as predicted with γ. The analytical curve is produced

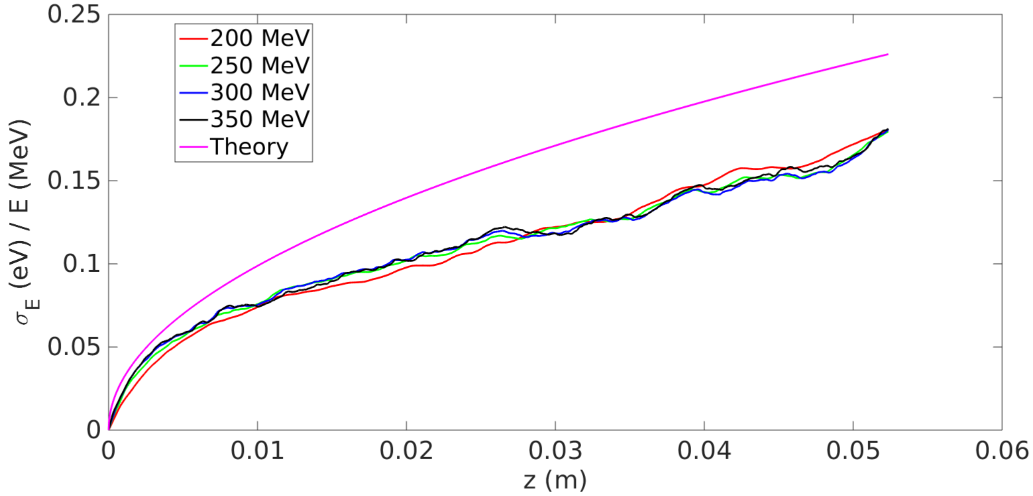


Figure 6.8: The cumulative diffusion in energy, normalized to the mean electron energy, due to the longitudinal field through the bend. The four averaged simulation curves are shown for each energy are each computed from 100 separate realizations of the electron beam. The 200 MeV analytic curve from equation 6.37 is shown in magenta.

by choosing the ambiguous g factor as before to be $g = 0.33$ which produces good agreement in the calculation of simply σ_{E_s}. The analytical result thus differs from the simulated curves likely in the discrepancy between the $\bar{z}$ definition of equation 6.35 and the fixed-time random walk definition of $\bar{z}$ which is referred to in equation 6.37. Although the two should be (and are clearly) related, there is likely an $\mathcal{O}(1)$ factor to convert between the two that needs to be established empirically.

Nonetheless, the curves are clearly proportional to $\sqrt{z}$, which provides evidence for the random-walk like character assumed in the theoretical model. Note that this case has $\sigma_\alpha = \sigma_x$ which was explicitly chosen to remove the complication of a z-varying Gaussian distribution (figure 6.5).

We can also test the σ_x dependence of the solution by varying this parameter. In order to ensure a z-invariant Gaussian, $\sigma_x < \sigma_\alpha$ should be maintained however. Several different evolution curves for various values of σ_x are plotted in figure 6.9. These simulations have $E = 200$ MeV, $\sigma_\alpha = 5 \times 10^{-5}$, and $\sigma_y = 10^{-6}$, and 6.24×10^6 particles, or 1 pC of charge. The various computational curves are the deviations

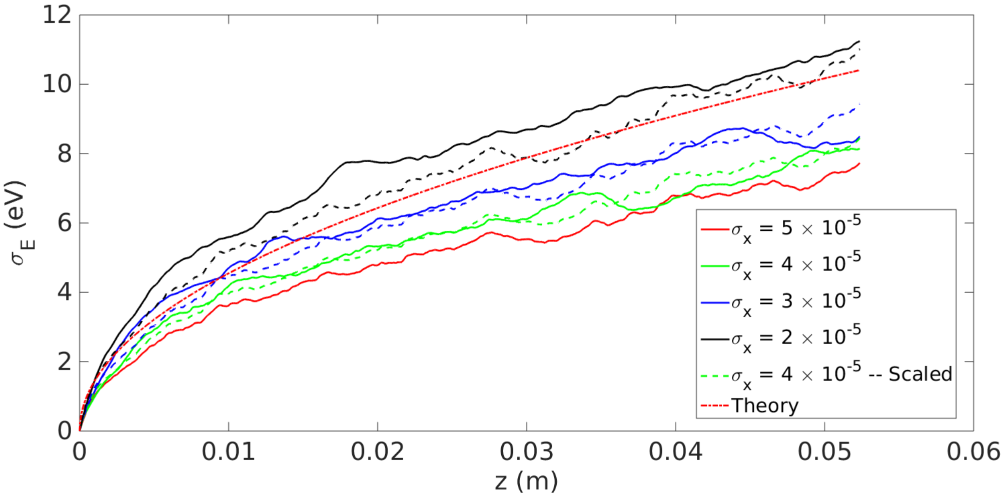


Figure 6.9: The cumulative diffusion in energy for various values of σ_x. Each computational curve is the result of 100 separate simulations. The dashed scaled curves are taken from the $\sigma_x = 5 \times 10^{-5}$ and scaled according to equation 6.35 and 6.37. The theory curve is the pure result from equation 6.37.

produced by 100 separate runs as a function of the transport distance z. The scaled curves represent the $\sigma_x = 5 \times 10^{-5}$ curve scaled according to the theoretical prediction for $\bar{z}$ of equation 6.35. These scaled curves accurately reproduce the σ_x dependence, confirming that the the full σ_x dependence, including the logarithm, of equation 6.35 is correct.

Finally, the pure theory curve is presented for the $\sigma_x = 5 \times 10^{-5}$ case. As mentioned before, due to some ambiguity regarding a constant factor with $\bar{z}$, this curve is relatively close but does not exactly match the computational situation. Given the myriad approximations that enter into this curve, however, and the verification of the scaling of this curve with σ_x and γ, this is reasonable agreement. In comparison with simulations, it appears that $\bar{z} \to \bar{z}/2$ accurately reproduces the computational curves. This suggests that the correct transition from a full-width trough crossing value of $\bar{z}$ to a random-walk $\bar{z}$ differ by approximately a factor of two. This empirical fact could be employed to obtain more accurate results from equation 6.37.

6.4 Implications for Current and Future Facilities

Having established the theoretical model and compared against full simulations, we seek now to take stock of the importance of this effect. From the numerical examples presented in figures 6.8 and 6.9, the absolute magnitude of this effect seems to be quite small. This is especially true when compared with another source of stochastic energy spread in high-energy bend magnets: ISR. We should therefore consider this new effect in comparison to other deleterious effects.

In addition to ISR, the closely spaced electrons in the bunch can perform multiple small-angle Coulomb scatters in which emittance can grow. This process is known as Intra-Beam Scattering (IBS), and can also lead to emittance growth at high energy. The theory behind this effect has been worked out for some time, and numerous formulations exist [140] [141] [17]. The exact formula are complicated, but the effect is generally characterized by a rise time T such that,

$$\frac{1}{T} = \frac{1}{\sqrt{\epsilon}} \frac{d\sqrt{\epsilon}}{dt}. \tag{6.40}$$

The fact that the effect is usually characterized by a rise time is evidence that it is typically important in ring machines only after many turns. A useful approximation for high energy beams has been given by Bane [142] as,

$$\frac{1}{T_{x,y}} = \frac{r_e^2 c N_e \Lambda}{16\gamma^3 \epsilon_x^{3/4} \epsilon_y^{3/4} \sigma_s \sigma_p \epsilon_{x,y}} \langle \mathcal{H}_{x,y} \sigma_H g_{\text{bane}}\left(\frac{a}{b}\right) (\beta_x \beta_y)^{-1/4} \rangle, \tag{6.41}$$

where Λ is the Coulomb logarithm, σ_p is the relative energy spread, $\sigma_H^{-2} = \sigma_p^{-2} + \mathcal{H}_x/\epsilon_x + \mathcal{H}_y/\epsilon_y$, $a = \sigma_H \gamma^{-1}\sqrt{\beta_x/\epsilon_x}$, $b = \sigma_H \gamma^{-1}\sqrt{\beta_y/\epsilon_y}$, and we have the function definition,

$$g_{\text{bane}}(x) = \frac{2\sqrt{x}}{\pi} \int_0^\infty \frac{du}{\sqrt{1+u^2}\sqrt{x^2+u^2}}. \tag{6.42}$$

A similar timescale for dilution of the slice energy spread can be worked out with

$1/T_p = \sigma_p^{-1} d(\sigma_p)/dt$, and,

$$\frac{1}{T_p} = \frac{r_e^2 c N_e \Lambda}{16\gamma^3 \epsilon_x^{3/4} \epsilon_y^{3/4} \sigma_s \sigma_p^3} \langle \sigma_H g_{\text{bane}} \left(\frac{a}{b}\right) (\beta_x \beta_y)^{-1/4} \rangle. \tag{6.43}$$

Despite all this complexity, the γ^3 dependence of the prefactor is evident, suggesting that IBS can be dominant for dense beams in moderate energy environments. A simplified model of the IBS effect for a cylindrical beam yields the more useful energy spread growth rate [**?**] [106],

$$\frac{d\sigma_p^2}{dz} = \frac{\sqrt{\pi}}{2} \Lambda \frac{I}{I_A} \frac{r_e}{\gamma^2 \sigma_{x,y} \epsilon_N}. \tag{6.44}$$

In contrast to ISR, but with similarity to the effect of this chapter, is the dependence on the beam density. In a dense beam limit, there may therefore be competition with IBS even at relatively moderate beam energies.

To investigate, consider the variation with the radial beamsize, which from equation 6.44 is clearly $\sigma_x^{-1/2}$. In the small beam limit ($\Xi \ll 1$), f is independent of σ_x and thus the only contribution comes from $\bar{z}$, which varies like $\log(\sigma_x)/\sigma_x$. This implies that the diffusion due to the stochastic CSR effect scales as $1/\sigma_x^{1/2 \sim 1}$, depending on the details of the logarithm. This compares favorably to the density dependence of the IBS. In the longitudinal dimension, again, IBS is proportional to $\sigma_\alpha^{-1/2}$, while for the stochastic CSR effect we have in both Ξ limits (assuming $\sigma_\alpha^2 \gg \sigma_x^3$), $\sigma_{\mathcal{E}} \sim \sigma_\alpha^{-1/2}$. Finally, we note that the IBS effect actually falls with energy (or is constant, depending on whether one observes relative changes or not), while the stochastic CSR term grows with $\gamma^{1/2 \sim 1}$. Thus, the stochastic CSR effect combines the density dependence found in IBS with some of the energy dependence found in ISR.

In the comparison to IBS, one sees that there is actually quite a similarity between the two effects. IBS is caused by multiple small-angle scatters off the Coulomb field, which was immediately dropped from the Liénard-Wiechert expression in our calculations. The effect described in this chapter, by contrast, is essentially due to the scattering off multiple synchrotron radiation cones from the radiation portion of the field. In this sense, these effects can be viewed as two sides of the same coin:

stochastic based scattering of the electromagnetic field.

As a concrete example, consider the first bunch compressor for LCLS-II. We assume a bunch energy $E = 250$ MeV, $\epsilon_N = 0.1$ μm, $\sigma_z = 100$ μm, and $\sigma_{x,y} \approx 50$μm. For the final bunch compression magnet, we take $\rho = 2.0$ m with a length $L_M = 0.2$ m. For these parameters, the stochastic CSR effect produces an energy variation $\sigma_{\mathcal{E}} \approx 130$ eV, while ISR produces an energy spread of $\sigma_{\mathcal{E}}^{ISR} \approx 8$ eV. Although IBS is a cumulative effect that occurs even outside bend magnets, in the final magnet alone it produces $\sigma_{\mathcal{E}} \approx 2000$eV.

This suggests a hierarchy of effects which rise and fall in importance along the length of the accelerator as both energy and beam density change. For example, due to the rapid scaling of ISR with energy, by the second bunch compressor ISR dominates over the other two effects in an LCLS-II like scenario. Nevertheless, this numerical example shows that the stochastic CSR term may not be irrelevant in all cases. Whether or not there exists an existing or planned setup in which it is the dominant effect, however, is unknown at this time. One possible candidate, though, is the dense, energetic beams which are often produced via plasma acceleration [143]. These beams can have extremely small emittances, high currents, and substantial energies. Therefore, there is potentially a zone in which the energy is not so high that ISR dominates, but high enough that the stochastic CSR term overwhelms IBS. Future high density plasma-derived beams may therefore be limited by the stochastic CSR effect when using dipole magnets to manipulate the beam.

Finally, the arguments which led to the derivation of the stochastic CSR effect apply equally well to several related problems. The most obvious is to the horizontal and vertical fields in a perfect dipole bend, which should share a similar variance. This transverse field noise would directly couple to emittance growth, and could potentially dominate over the longitudinal effect. We also acknowledge that a similar diffusion effect should exist in undulator magnets. The finite radiation cone of undulator radiation should provide a similar particle granularity effect, resulting in a noisy electromagnetic field. Analysis of this situation is complicated by the more complex trajectory which may render the geometric approach employed in this chapter fruitless. In retrospect, it is only through the simple circular geometry and clever coordinate choice that workable analytical results were able to be obtained even with all the simplifications employed. One method of attack which may prove fruitful is to describe the effect as Compton scattering off a finite number of photons, rather than the classical picture employed here. This should capture the same essential physics through a different lens and result in agreement between the two approaches.

Chapter 7

Conclusion

> I think this is the beginning of a beautiful friendship.
>
> —Rick Blaine, Casablanca

This thesis has dealt with the production of, and issues surrounding, high brightness electron beams. The majority of this work has centered around the production of an EEHG beam containing high harmonic content at the NLCTA facility. Through enabling instrumentation and technologies developed in this thesis, we were able to demonstrate the upconversion of a 2400 nm conventional laser to electron bunching at 32 nm. This represents a harmonic factor increase of 75 in a single seeding stage, which offers tantalizing prospects to seeding a full-blown FEL in the soft X-ray.

As conventional laser pulses at 266 nm, produced from a frequency tripled 800 nm Ti:Sa laser, are readily available, a factor of 75 in harmonic increase puts the target bunching wavelength at 3.5 nm. This wavelength sits solidly in the soft X-ray regime, and in contrast to the cascaded setup employed at FERMI [71], can be achieved in a single stage. The utilization of a single stage allows an EEHG signal which is more robust to electron beam imperfections than cascaded HGHG [76], making it an attractive alternative for generating stable, fully coherent pulses.

The experiments reported on in this thesis not only confirm that EEHG bunching can be produced at high harmonics, but also serve as a scaled experiment for larger machines. The energy modulations employed, $\Delta E \approx 100$ keV, are similar in magnitude to those which would be used at an FEL facility. Due to the relative coldness of the NLCTA electron beam, $\sigma_E \approx 1$ keV, the scaled EEHG parameters (A, B) differ dramatically from those which would be used at a larger, more energetic accelerator. Nevertheless, the absolute values of the energy modulations and chicane R_{56} values are relatively similar to those being considered in, for example, the FERMI study of possible EEHG configurations [126]. In this sense the NLCTA results provide a representative proof of concept for EEHG at high harmonics.

Along the way to ultimately obtaining radiation at 32 nm, emittance related smearing effects were observed to severely limit the harmonic upconversion efficiency. Fortunately, this main obstacle does not seem to be a limiting factor at larger facilities. Although the target radiation wavelength, say 3.5 nm, is smaller by a factor of 9, both the beam energy and averaged beta function are significantly larger at full-scale facilities. From equation 4.5, and recalling the extraordinarily small β of the VISA undulator, the scaling with facility size seems favorable to overcome this emittance effect. Furthermore, in a facility with significant gain, the length L in equation 4.5 need only be of order the gain length as the FEL interaction will compensate for small amounts debunching.

In contrast, energy smearing effects such as IBS, ISR, and the newly derived stochastic CSR, were exceedingly small at the NLCTA due to the low beam energy in the magnetic chicanes. These effects, however, are well known and can be designed around by clever design of the chicanes and beam optics. So while many of the problems that would be faced by a full fledged EEHG based FEL were dealt with in the course of this thesis, there exist some which were unimportant at the NLCTA but would be of concern at high energies and vice versa.

Beyond merely generating radiation at high harmonics, we established several different methods of generating multicolor radiation pulses using EEHG. One method is to generate distinct regions of HGHG and EEHG bunching on the same beam and tune their separation by varying the electron beam energy chirp. The other method involves choosing a degenerate EEHG configuration where multiple $|n|$ modes are present and again adjusting the linear beam chirp to provide wavelength separation.

While we were able to make a cursory verification of these configurations, neither the facility nor the experimental program was designed to provide a thorough investigation.

In particular, a facility with a zero-crossing accelerating cavity immediately prior to the EEHG manipulation sections would allow for a much more exhaustive study of these effects. This would allow a much finer scale control of the wavelength spacing of the produced sub-harmonics. Furthermore, the EUV spectrometer constructed as part of this thesis was not designed to perform high resolution measurements, but rather as a gain-driven device to detect small signals. As its operation resolution was observed to be somewhat large ($\delta\lambda \approx 0.2$ nm), it is not an ideal tool for measuring subharmonic spacing. A dedicated, high resolution spectrometer would allow for the detailed analysis of these schemes, including the possible production of radiation sidebands contained within a representative FEL bandwidth $\approx \rho$. The most likely scenario is that these measurements would be carried out in the VUV, similar to figure 4.11, in order to work with more stable harmonics as well as to facilitate the use of a higher resolution spectrometer.

The NLCTA facility is not large enough to develop a reliable microbunching instability (UBI), so simulations were performed to investigate the response of an EEHG-seeded FEL to this instability as well. These simulations agree well with analytical theory which states that the EEHG parameters can selectively excite or damp certain modulation wavelengths. Given a broad-band source such as the microbunching instability there exists the possibility of tuning the EEHG configuration to excite only the desired multicolor spectrum. This is fundamentally similar to using the laser heater suppression to selectively frequency mix UBI elements into the HGHG-seeded FEL to provide a multicolor seed [129]. In the EEHG configuration, however, the control of two sets of lasers and chicanes provide a greater range of tunability for this type of configuration. As before, a larger facility with a tunable laser heater and EEHG setup would be better equipped to experimentally explore this possibility.

Finally, while the analytical theory and simulations behind the stochastic CSR effect were presented in chapter 6, not all aspects of this phenomenon were discussed. Chief among these are the contributions due to the vertical and horizontal electric fields. Since both fields contain the same denominator term which is responsible for the trough, they too should contain a noise term. These noise terms would act directly on the transverse dimensions of the beam, rather than coupling to the dispersion through the factor $\mathcal{H}$. It is possible, therefore, that they may represent a comparable or larger effect compared to the longitudinal growth term which was thoroughly discussed. Further analysis should determine regimes of dominance for this various effect, and whether they may need to be considered in dense beam facilities.

The larger question of the effect on the beams in realistic facilities remains open as well. Although we have tried to give some accounting of the various scalings, and example magnitudes, a comprehensive analysis of existing electron beams is lacking. From the theoretical point of view, one would like to find a facility in which this effect is substantial enough to be measured and verify both the analytical theory and simulations. This confirmation would lend confidence to efforts to mitigate the effect, through, for example, deforming the beam shape, should the stochastic term ever become a major source of emittance degradation at a future facility.

Bibliography

[1] F. R. Elder, A. M. Gurewitsch, R. V. Langmuir, and H. C. Pollock, “Radiation from electrons in a synchrotron,” *Phys. Rev.*, vol. 71, pp. 829–830, Jun 1947.

[2] “LCLS Homepage.” `http://lcls.slac.stanford.edu/`. Accessed November 9, 2017.

[3] “DESY Homepage.” `http://flash.desy.de`. Accessed November 9, 2017.

[4] “SACLA Homepage.” `http://xfel.riken.jp/eng/`. Accessed November 9, 2017.

[5] “FERMI Homeage.” `http://www.elettra.eu/lightsources/fermi.html`. Accessed November 9, 2017.

[6] “PAL Homepage.” `http://pal.postech.ac.kr/paleng/`. Accessed November 9, 2017.

[7] “XFEL Homepage.” `http://www.xfel.eu/`. Accessed November 9, 2017.

[8] J. Amann, W. Berg, V. Blank, *et al.*, “Demonstration of self-seeding in a hard-X-ray free-electron laser,” *Nature photonics*, vol. 6, no. 10, pp. 693–698, 2012.

[9] A. Kondratenko and E. Saldin, “Generation of coherent radiation by a relativistic electron beam in an ondulator,” *Part. Accel.*, vol. 10, pp. 207–216, 1980.

[10] R. Bonifacio, C. Pellegrini, and L. Narducci, “Collective instabilities and high-gain regime in a free electron laser,” *Optics Communications*, vol. 50, no. 6, pp. 373 – 378, 1984.

[11] User Danh, Wikimedia Commons, "Overview of the wavelengths of commercially available lasers." `https://commons.wikimedia.org/wiki/File:Commercial_laser_lines_with_legend.svg`, 2010. Accessed November 9, 2017. Licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

[12] T. Popmintchev, M.-C. Chen, P. Arpin, M. M. Murnane, and H. C. Kapteyn, "The attosecond nonlinear optics of bright coherent X-ray generation," *Nature Photonics*, vol. 4, no. 12, pp. 822–832, 2010.

[13] E. Courant and H. Snyder, "Theory of the alternating-gradient synchrotron," *Annals of Physics*, vol. 3, no. 1, pp. 1 – 48, 1958.

[14] M. P. Dunning, *Coherent terahertz radiation emitted by a sub-picosecond electron beam in a magnetic chicane.* PhD thesis, University of California, Los Angeles, 2009.

[15] M. Conte and W. W. MacKay, *An Introduction To The Physics Of Particle Accelerators (Second Edition).* World Scientific Publishing Company, 2008.

[16] S.-Y. Lee, *Accelerator physics.* World Scientific Publishing Co Inc, 2004.

[17] A. W. Chao, K. H. Mess, M. Tigner, and F. Zimmermann, eds., *Handbook of accelerator physics and engineering.* Hackensack, USA: World Scientific, 2013.

[18] P. Emma, R. Akre, J. Arthur, *et al.*, "First lasing and operation of an ångstrom-wavelength free-electron laser," *Nature Photonics*, vol. 4, no. 9, pp. 641–647, 2010.

[19] M. Reiser, "Free energy and emittance growth in nonstationary charged particle beams," *Journal of Applied Physics*, vol. 70, no. 4, pp. 1919–1923, 1991.

[20] M. Reiser, *Theory and Design of Charged Particle Beams.* Wiley-VCH, 2008.

[21] I. Hofmann, "Emittance growth of beams close to the space charge limit," *IEEE Transactions on Nuclear Science*, vol. 28, pp. 2399–2401, June 1981.

[22] A. W. Chao, B. Richter, and C. Y. Yao, "Beam Emittance Growth Caused by Transverse Deflecting Fields in a Linear Accelerator," *Nucl. Instrum. Meth.*, vol. 178, p. 1, 1980.

[23] K. P. Wootton, M. J. Boland, and R. P. Rassool, "Measurement of ultralow vertical emittance using a calibrated vertical undulator," *Phys. Rev. ST Accel. Beams*, vol. 17, p. 112802, Nov 2014.

[24] A. Hofmann, *The physics of synchrotron radiation*, vol. 20. Cambridge University Press, 2004.

[25] J. D. Jackson, *Classical electrodynamics.* New York, NY: Wiley, 3rd ed. ed., 1999.

[26] T. Nakazato, M. Oyamada, N. Niimura, *et al.*, "Observation of coherent synchrotron radiation," *Phys. Rev. Lett.*, vol. 63, pp. 1245–1248, Sep 1989.

[27] G. L. Carr, M. C. Martin, W. R. McKinney, *et al.*, "High-power terahertz radiation from relativistic electrons," *Nature*, vol. 420, no. 6912, pp. 153–156, 2002.

[28] Y. Shibata, S. Hasebe, K. Ishi, *et al.*, "Coherent Smith-Purcell radiation in the millimeter-wave region from a short-bunch beam of relativistic electrons," *Phys. Rev. E*, vol. 57, pp. 1061–1074, Jan 1998.

[29] U. Happek, A. J. Sievers, and E. B. Blum, "Observation of coherent transition radiation," *Phys. Rev. Lett.*, vol. 67, pp. 2962–2965, Nov 1991.

[30] Y. Ding, A. Brachmann, F.-J. Decker, *et al.*, "Measurements and simulations of ultralow emittance and ultrashort electron beams in the linac coherent light source," *Phys. Rev. Lett.*, vol. 102, p. 254801, Jun 2009.

[31] C. B. Schroeder, P. B. Lee, J. S. Wurtele, E. Esarey, and W. P. Leemans, "Generation of ultrashort electron bunches by colliding laser pulses," *Phys. Rev. E*, vol. 59, pp. 6037–6047, May 1999.

[32] N. Naumova, I. Sokolov, J. Nees, *et al.*, “Attosecond electron bunches,” *Phys. Rev. Lett.*, vol. 93, p. 195003, Nov 2004.

[33] V. V. Kulagin, V. A. Cherepenin, M. S. Hur, and H. Suk, “Theoretical investigation of controlled generation of a dense attosecond relativistic electron bunch from the interaction of an ultrashort laser pulse with a nanofilm,” *Phys. Rev. Lett.*, vol. 99, p. 124801, Sep 2007.

[34] G. V. Stupakov and M. S. Zolotorev, “Ponderomotive laser acceleration and focusing in vacuum for generation of attosecond electron bunches,” *Phys. Rev. Lett.*, vol. 86, pp. 5274–5277, Jun 2001.

[35] D. G. Bocek, *Generation and characterization of superradiant undulator radiation.* PhD thesis, Stanford University, 1997.

[36] K.-J. Kim, “Characteristics of synchrotron radiation,” *AIP Conference Proceedings*, vol. 184, no. 1, pp. 565–632, 1989.

[37] Z. Huang and K.-J. Kim, “Review of x-ray free-electron laser theory,” *Phys. Rev. ST Accel. Beams*, vol. 10, p. 034801, Mar 2007.

[38] A. Siegman, *Lasers.* University Science Books, 1986.

[39] R. B. Palmer, “Acceleration theorems,” *AIP Conference Proceedings*, vol. 335, no. 1, pp. 90–100, 1995.

[40] E. Hemsing, G. Stupakov, D. Xiang, and A. Zholents, “Beam by design: Laser manipulation of electrons in modern accelerators,” *Rev. Mod. Phys.*, vol. 86, pp. 897–941, Jul 2014.

[41] B. Garcia, E. Hemsing, T. Raubenheimer, L. T. Campbell, and B. W. J. McNeil, “Method to generate a pulse train of few-cycle coherent radiation,” *Phys. Rev. Accel. Beams*, vol. 19, p. 090701, Sep 2016.

[42] P. Schmser, M. Dohlus, and J. Rossbach, *Ultraviolet and Soft X-Ray Free-Electron Lasers: Introduction to Physical Principles, Experimental Results,*

Technological Challenges. Springer Publishing Company, Incorporated, 1st ed., 2008.

[43] K.-J. Kim, "An analysis of self-amplified spontaneous emission," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 250, no. 1, pp. 396 – 403, 1986.

[44] J.-M. Wang and L.-H. Yu, "A transient analysis of a bunched beam free electron laser," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 250, no. 1, pp. 484 – 489, 1986.

[45] M. Xie, "Exact and variational solutions of 3D eigenmodes in high gain FELs," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 445, no. 13, pp. 59 – 66, 2000.

[46] G. Schott, "Electromagnetic radiation," 1912.

[47] J. Schwinger, "On the classical radiation of accelerated electrons," *Physical review*, vol. 75, no. 12, p. 1912, 1949.

[48] J. Murphy, R. Gluckstern, and S. Krinsky, "Longitudinal wake field for an electron moving on a circular orbit," *Part. Accel.*, vol. 57, no. BNL-63090, pp. 9–64, 1996.

[49] E. L. Saldin, E. A. Schneidmiller, and M. Yurkov, "On the coherent radiation of an electron bunch moving in an arc of a circle," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 398, no. 2-3, pp. 373–394, 1997.

[50] R. D. Ryne, B. Carlsten, J. Qiang, and N. Yampolsky, "A Model for One-Dimensional Coherent Synchrotron Radiation including Short-Range Effects," *ArXiv e-prints*, Feb. 2012.

[51] C. Mayes and G. Hoffstaetter, "Exact 1d model for coherent synchrotron radiation with shielding and bunch compression," *Phys. Rev. ST Accel. Beams*, vol. 12, p. 024401, Feb 2009.

[52] P. Emma and R. Brinkmann, "Emittance dilution through coherent energy spread generation in bending systems," in *Particle Accelerator Conference, 1997. Proceedings of the 1997*, vol. 2, pp. 1679–1681, IEEE, 1997.

[53] M. Venturini, "CSR-induced emittance growth in achromats: Linear formalism revisited," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 794, pp. 109 – 112, 2015.

[54] M. Venturini, "Design of a triple-bend isochronous achromat with minimum coherent-synchrotron-radiation-induced emittance growth," *Phys. Rev. Accel. Beams*, vol. 19, p. 064401, Jun 2016.

[55] D. Khan and T. Raubenheimer, "LCLS-II bunch compressor study: 5-bend chicane," in *FEL2014 Conference Proceedings, Basel*, 2014.

[56] M. Borland, Y. Chae, P. Emma, *et al.*, "Start-to-end simulation of self-amplified spontaneous emission free electron lasers from the gun through the undulator," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 483, no. 1, pp. 268 – 272, 2002. Proceedings of the 23rd International Free Electron Laser Confere nce and 8th FEL Users Workshop.

[57] E. Saldin, E. Schneidmiller, and M. Yurkov, "Klystron instability of a relativistic electron beam in a bunch compressor," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 490, no. 12, pp. 1 – 8, 2002.

[58] Z. Huang and K.-J. Kim, "Formulas for coherent synchrotron radiation microbunching in a bunch compressor chicane," *Phys. Rev. ST Accel. Beams*, vol. 5, p. 074401, Jul 2002.

[59] S. Heifets, G. Stupakov, and S. Krinsky, "Coherent synchrotron radiation instability in a bunch compressor," *Phys. Rev. ST Accel. Beams*, vol. 5, p. 064401, Jun 2002.

[60] D. Ratner, C. Behrens, Y. Ding, *et al.*, "Time-resolved imaging of the microbunching instability and energy spread at the linac coherent light source," *Phys. Rev. ST Accel. Beams*, vol. 18, p. 030704, Mar 2015.

[61] J. Rosenzweig, C. Pellegrini, L. Serafini, C. Ternieden, and G. Travish, "Space-charge oscillations in a self-modulated electron beam in multi-undulator free-electron lasers," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 393, no. 1, pp. 376 – 379, 1997.

[62] Z. Huang, M. Borland, P. Emma, *et al.*, "Suppression of microbunching instability in the linac coherent light source," *Phys. Rev. ST Accel. Beams*, vol. 7, p. 074401, Jul 2004.

[63] B. Girard, Y. Lapierre, J. M. Ortega, *et al.*, "Optical frequency multiplication by an optical klystron," *Phys. Rev. Lett.*, vol. 53, pp. 2405–2408, Dec 1984.

[64] B. M. Kincaid, R. R. Freeman, C. Pelligrini, *et al.*, "The transverse optical klystron experiment at NSLS: Laser harmonics into the XUV," *AIP Conference Proceedings*, vol. 118, no. 1, pp. 110–124, 1984.

[65] I. Ben-Zvi, L. D. Mauro, S. Krinsky, M. White, and L. Yu, "Proposed UV FEL user facility at BNL," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 304, no. 1, pp. 181 – 186, 1991.

[66] L. H. Yu, "Generation of intense UV radiation by subharmonically seeded single-pass free-electron lasers," *Phys. Rev. A*, vol. 44, pp. 5178–5193, Oct 1991.

[67] E. Hemsing and D. Xiang, "Cascaded modulator-chicane modules for optical manipulation of relativistic electron beams," *Phys. Rev. ST Accel. Beams*, vol. 16, p. 010706, Jan 2013.

[68] D. Ratner and A. Chao, "Seeded radiation sources with sawtooth waveforms," in *Proceedings of the 2011 FEL Conference*, (Shanghai, China), 2011.

[69] D. Xiang and G. Stupakov, "Echo-enabled harmonic generation free electron laser," *Phys. Rev. ST Accel. Beams*, vol. 12, p. 030702, Mar 2009.

[70] E. Allaria, R. Appio, L. Badano, *et al.*, "Highly coherent and stable pulses from the FERMI seeded free-electron laser in the extreme ultraviolet," *Nature Photonics*, vol. 6, no. 10, pp. 699–704, 2012.

[71] E. Allaria, D. Castronovo, P. Cinquegrana, *et al.*, "Two-stage seeded soft-x-ray free-electron laser," *Nature Photonics*, vol. 7, no. 11, pp. 913–918, 2013.

[72] G. Stupakov, "Using the beam-echo effect for generation of short-wavelength radiation," *Phys. Rev. Lett.*, vol. 102, p. 074801, Feb 2009.

[73] E. L. Hahn, "Spin echoes," *Phys. Rev.*, vol. 80, pp. 580–594, Nov 1950.

[74] N. A. Kurnit, I. D. Abella, and S. R. Hartmann, "Observation of a photon echo," *Phys. Rev. Lett.*, vol. 13, pp. 567–568, Nov 1964.

[75] R. M. Hill and D. E. Kaplan, "Cyclotron resonance echo," *Phys. Rev. Lett.*, vol. 14, pp. 1062–1063, Jun 1965.

[76] E. Hemsing, B. Garcia, Z. Huang, T. Raubenheimer, and D. Xiang, "Sensitivity of echo enabled harmonic generation to sinusoidal electron beam energy structure," *Phys. Rev. Accel. Beams*, vol. 20, p. 060702, Jun 2017.

[77] Z. T., WangD., C. H., *et al.*, "First lasing of an echo-enabled harmonic generation free-electron laser," *Nat Photon*, vol. 6, pp. 360–363, Jun 2012.

[78] R. D. Ruth, C. Adolphsen, K. Bane, *et al.*, "The Next Linear Collider Test Accelerator," in *1993 15th IEEE Particle Accelerator Conference*, p. 543, 1994.

[79] C. Adolphsen, T. Lavine, C. Nantista, *et al.*, "Beam loading compensation in the NLCTA," in *Proceedings of the 1997 Particle Accelerator Conference (Cat. No.97CH36167)*, vol. 2, pp. 1676–1678 vol.2, May 1997.

[80] R. Noble, E. Colby, D. Palmer, *et al.*, "The ORION facility," in *Particle Accelerator Conference, 2003. PAC 2003. Proceedings of the*, vol. 3, pp. 1858–1860, IEEE, 2003.

[81] C. M. S. Sears, E. Colby, R. J. England, *et al.*, "Phase stable net acceleration of electrons from a two-stage optical accelerator," *Phys. Rev. ST Accel. Beams*, vol. 11, p. 101301, Oct 2008.

[82] E. Peralta, K. Soong, R. England, *et al.*, "Demonstration of electron acceleration in a laser-driven dielectric microstructure," *Nature*, vol. 503, no. 7474, pp. 91–94, 2013.

[83] E. Hemsing, M. Dunning, C. Hast, *et al.*, "Highly coherent vacuum ultraviolet radiation at the 15th harmonic with echo-enabled harmonic generation technique," *Phys. Rev. ST Accel. Beams*, vol. 17, p. 070702, Jul 2014.

[84] D. T. Palmer, X. J. Wang, R. H. Miller, *et al.*, "Emittance studies of the BNL/SLAC/UCLA 1.6 cell photocathode RF gun," in *Proceedings of the 1997 Particle Accelerator Conference (Cat. No.97CH36167)*, vol. 3, pp. 2687–2689 vol.3, May 1997.

[85] M. Ferrario, D. Alesini, A. Bacci, *et al.*, "Experimental demonstration of emittance compensation with velocity bunching," *Phys. Rev. Lett.*, vol. 104, p. 054801, Feb 2010.

[86] S. Tantawi, M. Shumail, J. Neilson, *et al.*, "Experimental demonstration of a tunable microwave undulator," *Phys. Rev. Lett.*, vol. 112, p. 164802, Apr 2014.

[87] McPherson, "200mm f.l. Vacuum Monochromator." `http://mcphersoninc.com/pdf/234302.pdf`.

[88] D. Xiang, E. Colby, M. Dunning, *et al.*, "Demonstration of the echo-enabled harmonic generation technique for short-wavelength seeded free electron lasers," *Phys. Rev. Lett.*, vol. 105, p. 114801, Sep 2010.

[89] D. Xiang, E. Colby, M. Dunning, *et al.*, "Evidence of high harmonics from echo-enabled harmonic generation for seeding x-ray free electron lasers," *Phys. Rev. Lett.*, vol. 108, p. 024802, Jan 2012.

[90] R. Carr, M. Cornacchia, P. Emma, *et al.*, "Visible-infrared self-amplified spontaneous emission amplifier free electron laser undulator," *Phys. Rev. ST Accel. Beams*, vol. 4, p. 122402, Dec 2001.

[91] K. Halbach, "Physical and optical properties of rare earth cobalt magnets," *Nuclear Instruments and Methods in Physics Research*, vol. 187, pp. 109–117, Aug. 1981.

[92] R. Ruland, D. Arnett, G. Bowden, *et al.*, "VISA undulator fiducialization and alignment," in *Proceedings of the 6th International Workshop on Accelerator Alignment, ESRF*, 1999.

[93] D. Farkas, H. Hogg, G. Loew, and P. Wilson, "SLED: A Method for Doubling SLAC's Energy," in *Proceeding of 9th Int. Conf. on High Energy Accelerators*, Citeseer, 1974.

[94] Center for X-Ray Optics, LBNL, "Solid mirror reflectivity tables." `http://henke.lbl.gov/optical_constants/mirror2.html`.

[95] Hitachi High Technologies, "Hitachi aberration-corrected concave gratings for flat-field spectrographs." `http://www.hitachi-hightech.com/products/images/9797/ana-grating_05.pdf`.

[96] E. Lorek, E. W. Larsen, C. M. Heyl, *et al.*, "High-order harmonic generation using a high-repetition-rate turnkey laser," *Review of Scientific Instruments*, vol. 85, no. 12, p. 123106, 2014.

[97] S. Ghimire, G. Ndabashimiye, and D. A. Reis, "High-order harmonic generation in solid argon," in *2012 Conference on Lasers and Electro-Optics (CLEO)*, pp. 1–2, May 2012.

[98] M. B. Chowdhuri, S. Morita, M. Goto, *et al.*, "Spectroscopic comparison between 1200 grooves/mm ruled and holographic gratings of a flat-field spectrometer and its absolute sensitivity calibration using bremsstrahlung continuum," *Review of Scientific Instruments*, vol. 78, no. 2, p. 023501, 2007.

[99] J. L. Culhane, L. K. Harra, A. M. James, *et al.*, *The EUV Imaging Spectrometer for Hinode*, pp. 69–111. New York, NY: Springer New York, 2008.

[100] C. Palmer and E. G. Loewen, *Diffraction grating handbook.* Newport Corporation New York, 2005.

[101] Z. Huang, D. Ratner, G. Stupakov, and D. Xiang, "Effects of energy chirp on echo-enabled harmonic generation free-electron lasers," in *Proceedings of the 31st International Free Electron Laser Conference (FEL 09), Liverpool, UK*, (STFC Daresbury Laboratory), 2009.

[102] J. Arthur *et al.*, "Linac coherent light source (LCLS) conceptual design report," *SLAC-R-593*, 2002.

[103] K. Flöttmann *et al.*, *Astra: A space charge tracking algorithm.* DESY, Hamburg, 2014. `http://www.desy.de/~mpyflo/`.

[104] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182–197, Apr 2002.

[105] M. G. Minty and F. Zimmermann, *Measurement and control of charged particle beams.* Springer Science & Business Media, 2013.

[106] G. Penn, "EEHG performance and scaling laws," tech. rep., Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), 2013.

[107] Y. Shoji, "Dependence of average path length betatron motion in a storage ring," *Phys. Rev. ST Accel. Beams*, vol. 8, p. 094001, Sep 2005.

[108] E. Hemsing, M. Dunning, B. Garcia, *et al.*, “Echo-enabled harmonics up to the 75th order from precisely tailored electron beams,” *Nature Photonics*, 2016.

[109] G. Penn and M. Reinsch, “Designs and numerical calculations for echo-enabled harmonic generation at very high harmonics,” *Journal of Modern Optics*, vol. 58, pp. 1404–1418, Sept. 2011.

[110] I. Ben-Zvi, K. Yang, and L. Yu, “The fresh-bunch technique in FELs,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 318, no. 1, pp. 726 – 729, 1992.

[111] L. Young and J. Billen, “The particle tracking code PARMELA,” *PAC03 Proceedings, May*, 2003.

[112] M. Borland, *ELEGANT*. Advanced Photon Source LS-287, 2000.

[113] J. Qiang, R. D. Ryne, S. Habib, and V. Decyk, “An Object-Oriented Parallel Particle-in-Cell Code for Beam Dynamics Simulation in Linear Accelerators,” *Journal of Computational Physics*, vol. 163, pp. 434–451, Sept. 2000.

[114] J. Murphy, S. Krinsky, and R. Gluckstern, “Longitudinal wakefield for an electron moving on a circular orbit,” *Particle Accelerators*, vol. 57, pp. 9–64, 1997.

[115] Z. Huang, “Intrabeam scattering in an X-ray FEL driver,” Tech. Rep. SLAC-TN-05-026, LCLS-TN-02-8, 2002.

[116] G. Stupakov, “Effect of coulomb collisions on echo-enabled harmonic generation (EEHG),” *Proceedings of the FEL2011, Shanghai*, pp. 49–52, 2011.

[117] S. Reiche, “GENESIS 1.3: a fully 3d time-dependent FEL simulation code,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 429, no. 13, pp. 243 – 248, 1999.

[118] F. Arecchi and R. Bonifacio, “Theory of optical maser amplifiers,” *IEEE Journal of Quantum Electronics*, vol. 1, pp. 169–178, July 1965.

[119] L. T. Campbell and B. W. J. McNeil, “Puffin: A three dimensional, unaveraged free electron laser simulation code,” *Physics of Plasmas*, vol. 19, no. 9, 2012.

[120] G. Marcus, W. Fawley, M. Yurkov, and S. Reiche, “FEL code comparison for the production of harmonics via harmonic lasing,” in *Proceedings of the 2014 FEL Conference*, (Basel, Switzerland), 2014.

[121] E. Saldin, E. Schneidmiller, and M. Yurkov, “FAST: a three-dimensional time-dependent FEL simulation code,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 429, no. 13, pp. 233 – 237, 1999.

[122] S. Reiche, P. Musumeci, and K. Goldammer, “Recent upgrade to the free-electron laser code GENESIS 1.3,” in *2007 IEEE Particle Accelerator Conference (PAC)*, pp. 1269–1271, June 2007.

[123] W. Fawley, “A user manual for GINGER,” Tech. Rep. Rpt. LBNL-49625-Rev. 1, 2004.

[124] B. McNeil, G. Robb, M. Poole, and N. Thompson, “Harmonic lasing in a free-electron-laser amplifier,” *Physical review letters*, vol. 96, no. 8, p. 084801, 2006.

[125] E. A. Schneidmiller and M. V. Yurkov, “A possible upgrade of FLASH for harmonic lasing down to 1.3 nm,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 717, pp. 37–43, 2013.

[126] P. Rebernik Ribi, E. Roussel, G. Penn, *et al.*, “Echo-enabled harmonic generation studies for the FERMI free-electron laser,” *Photonics*, vol. 4, no. 1, 2017.

[127] D. Ratner, R. Abela, J. Amann, *et al.*, “Experimental demonstration of a soft x-ray self-seeded free-electron laser,” *Phys. Rev. Lett.*, vol. 114, p. 054801, Feb 2015.

[128] Z. Zhang, R. Lindberg, W. M. Fawley, *et al.*, “Microbunching-instability-induced sidebands in a seeded free-electron laser,” *Phys. Rev. Accel. Beams*, vol. 19, p. 050701, May 2016.

[129] E. Roussel, E. Ferrari, E. Allaria, *et al.*, “Multicolor high-gain free-electron laser driven by seeded microbunching instability,” *Phys. Rev. Lett.*, vol. 115, p. 214801, Nov 2015.

[130] T. Plath, C. Lechner, V. Miltchev, *et al.*, “Mapping few-femtosecond slices of ultra-relativistic electron bunches,” *Sci Rep*, vol. 7, p. 2431, May 2017. 2184[PII].

[131] D. H. Dowell and J. F. Schmerge, “Quantum efficiency and thermal emittance of metal photocathodes,” *Phys. Rev. ST Accel. Beams*, vol. 12, p. 074201, Jul 2009.

[132] M. Borland, “Simple method for particle tracking with coherent synchrotron radiation,” *Phys. Rev. ST Accel. Beams*, vol. 4, p. 070701, Jul 2001.

[133] R. D. Ryne, B. Carlsten, J. Qiang, and N. Yampolsky, “A model for one-dimensional coherent synchrotron radiation including short-range effects,” *arXiv preprint arXiv:1202.2409*, 2012.

[134] R. Li, “Curvature-induced bunch self-interaction for an energy-chirped bunch in magnetic bends,” *Phys. Rev. ST Accel. Beams*, vol. 11, p. 024401, Feb 2008.

[135] C. Huang, T. J. T. Kwan, and B. E. Carlsten, “Two dimensional model for coherent synchrotron radiation,” *Phys. Rev. ST Accel. Beams*, vol. 16, p. 010701, Jan 2013.

[136] C. E. Mitchell, J. Qiang, and R. D. Ryne, “A fast method for computing 1-d wakefields due to coherent synchrotron radiation,” *Nuclear Instruments and*

Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 715, pp. 119 – 125, 2013.

[137] R. Ryne, C. Mitchell, J. Qiang, B. Carlsten, and N. Yampolsky, "Large scale simulation of synchrotron radiation using a Lienard-Wiechert approach," *Proc. IPAC*, 2012.

[138] R. Ryne, C. Mitchell, J. Qiang, B. Carlsten, and N. Yampolsky, "Using a Lienard-Wiechert solver to study coherent synchrotron radiation effects," in *Proceedings of the 2013 FEL Conference*, pp. 17–23, 2013.

[139] Y. Cai, "Coherent synchrotron radiation by electrons moving on circular orbits," *Phys. Rev. Accel. Beams*, vol. 20, p. 064402, Jun 2017.

[140] L. R. Evans and B. W. Zotter, "Intrabeam scattering in the SPS," Tech. Rep. CERN-SPS-80-15-DI, 1980.

[141] J. D. Bjorken and S. K. Mtingwa, "Intrabeam Scattering," *Part. Accel.*, vol. 13, pp. 115–143, 1983.

[142] K. L. F. Bane, "A Simplified model of intrabeam scattering," in *Particle accelerator. Proceedings, 8th European Conference, EPAC 2002, Paris, France, June 3-7, 2002*, pp. 1443–1445, 2002.

[143] E. Esarey, C. B. Schroeder, and W. P. Leemans, "Physics of laser-driven plasma-based electron accelerators," *Rev. Mod. Phys.*, vol. 81, pp. 1229–1285, Aug 2009.

Bryant William Garcia

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Tor O. Raubenheimer) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Zhirong Huang)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Roger Blandford)

Approved for the Stanford University Committee on Graduate Studies