The physics of x-ray free-electron lasers

C. Pellegrini, A. Marinelli, and S. Reiche

The physics of x-ray free-electron lasers

C. Pellegrini
Department of Physics and Astronomy, University of California at Los Angeles, Los Angeles, CA 90095, USA and SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

A. Marinelli
SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA

S. Reiche
Paul Scherrer Institute, 5232 Villigen PSI, Switzerland

Outline

Abstract

X-ray Free-electron Lasers (X-ray FELs) give us for the first time the possibility to explore structures and dynamical processes of atomic and molecular systems at the Ångstrom-femtosecond space and time scales. They generate coherent photon pulses with time duration of a few to 100 femtosecond, peak power of 10 to 100 GW, over a wavelength range extending from about 100 nm to less than 1 Å. Using these novel and unique capabilities new scientific results are being obtained in atomic and molecular sciences, in areas of physics, chemistry and biology. In this paper we review the physical principles, the theoretical models and the numerical codes on which X-ray FELs are based, starting from a single electron spontaneous undulator radiation to the FEL collective instability of a high density electron beam, strongly enhancing the electromagnetic radiation field intensity and its coherence properties. We present also a short review of the main experimental properties of X-ray FELs, and discuss the results of the most recent research to improve their longitudinal coherence properties, increase the peak power and generate multicolor spectra.

Introduction

Following the initial development of lasers in the infrared and visible spectral region in the 1960s (Bertolotti, 2005; Zinth, Laubereau and Kaiser, 2011), there has been a continued effort to extend the generation of coherent electromagnetic radiation to shorter and shorter wavelengths, with the ultimate goal of reaching the X-ray region. However the conventional atom-based population inversion approach cannot be easily extended to X-rays because of the very short lifetime of excited atom-core quantum energy levels. In addition a larger energy is required to excite electrons in the inner core levels. George Chapline and Lowell Wood (Chapline and Wood, 1975), of the Lawrence Livermore National Laboratory, estimated that the radiative lifetime of an

X-ray laser transition would be about 1 fs times the square of the wavelength in Ångstroms. Satisfying these conditions to obtain population inversion requires a very large amount of pumping power.

Building low loss optical cavities for X-ray laser oscillators is also difficult, in fact beyond the present state of the art. This led Livermore scientists to propose the use of a nuclear weapon to drive an X-ray laser. They tried the scheme in the Dauphin experiment, apparently with success, in 1980. The idea was part of the Star Wars Defense Initiative: generate an X-ray beam in space to kill incoming missiles by exploding an atomic bomb. When Star War was terminated this program ended (Hecht, 2008).

The development of high peak power, short pulse, visible light lasers made possible another approach: pumping cylindrical plasmas, in some cases also confining the plasma with magnetic fields. These experiments led to X-ray lasing around 18 nm with gain of about 100 in 1985 at Livermore and Princeton (Matthews *et al.*, 1985; Suckewer *et al.*, 1985). More work has been done from that time and lasing has been demonstrated at several wavelengths in the soft X-ray region, however with limited peak power and tunability. A review of the most recent work and developments with this approach is given in (Suckewer and Jaeglé, 2009).

An alternative to atomic transition with population inversion X-ray lasers is the Self Amplified Spontaneous Emission (SASE) X-ray free-electron laser (X-ray FEL) (Kondradenko and Saldin, 1980; Bonifacio, Pellegrini and Narducci, 1984; Murphy and Pellegrini, 1985; Pellegrini, 1988). Claudio Pellegrini proposed in 1992 to build a SASE X-ray FEL in the wavelength range 0.1 to 4 nm (Pellegrini 1992; Pellegrini 2012) using one third of the 40 GeV, two miles long, linear accelerator of the SLAC National Accelerator Laboratory (Neal, *et al.,* 1967). The proposal is based on the emission of radiation from relativistic electron beams in a periodically alternating magnetic field: an undulator magnet. It has been shown (Friedman *et al.*, 1988) that an accelerated electron beam constitutes an inverted population medium. When the beam traverses a wiggler or a slow-wave structure it radiates electromagnetic waves and the radiative transitions in the FEL can be described in terms of conventional laser physics. However in the case of interest to us the SASE X-ray FEL operates in the classical regime where a large number of photons are emitted before the electron energy changes significantly.

The proposal led to the design and construction, at the SLAC National Accelerator Laboratory, of LCLS, that successfully lased at a record short wavelength of l.5 Å in 2009 (Emma *et al.*, 2010). Another X-ray FEL, SACLA, has been recently successfully commissioned in Japan (Ishikawa *et al.*, 2012). Three more hard X-ray FELs are under construction in Korea (Kim and Yoon, 2009), Switzerland (Patterson *et al.*, 2010) and the European Union (Altarelli, *et al.* eds., 2006). Two soft X-ray FELs, FLASH and Fermi (Ackermann *et al.*, 2007; Allaria *et al.*, 2012), are also in operation at DESY, in Germany, and Sincrotrone Trieste, in Italy. LCLS is being upgraded to LCLS-II,

covering both the soft and hard X-rays regions, from 10 nm to less than 1Å (Galayda, 2014).

LCLS and SACLA X-ray pulses have very high intensity and brightness. The peak power is tens of GW and the pulse duration can vary between a few and about 100 femtosecond. The radiation is spatially coherent, nearly diffraction limited. The number of photons in a coherent volume of the six-dimensional radiation phase-space is 10^9 or larger, compared to less than one in a conventional synchrotron undulator radiation. The radiation pulse of a SASE FEL is not transform limited, the typical relative line width is about 1/1000, larger than the minimum line width corresponding to the pulse duration. Much work is being done to improve the longitudinal coherence. Recent experimental results demonstrate several schemes to obtain nearly transform limited radiation pulses in the hard and soft X-ray regions at LCLS (Amann *et al.*, 2012; Ratner *et al.*, 2015), Fermi (Allaria *et al.*, 2012; Allaria *et al.*, 2013a) and the SPARC FEL in Frascati (Labat, 2011) in the UV region. Other recent developments have shown how to obtain high power, near transform limited multicolor pulses (Allaria et al., 2013b; Marinelli *et al.*, 2015).

These combined characteristics make the X-ray FEL a unique tool to explore atomic and molecular science at the length and time scale, 1 Å and 1fs, characterizing these phenomena, obtaining information on structures and dynamical processes not accessible until now. A review of the scientific results obtained in the broad fields of application of LCLS -physics, chemistry, biology and material sciences- during its first 5 years of operations is published concurrently in this edition of Review of Modern Physics (Bostedt et al., 2015).

The continued development of X-ray FELs, now being pursued in many laboratories, is following lines analogous to those of conventional lasers. The resulting advances will make X-ray FELs even more "spectrally bright" and flexible, and will make it the X-ray source of choice, probably the only source, for use in even more challenging experiments on imaging of non-periodic structures, non linear science, dynamical processes at femtosecond, and even attosecond, scale.

Free-electron lasers combine the physics and technology of particle accelerators and lasers to generate electromagnetic radiation with very high brightness, larger than any other existing source. John Madey (Madey, 1971) introduced the initial FEL concept in 1971. An FEL consists basically of an electron accelerator and an undulator magnet, as shown in FIG. 1. Optical elements and input electromagnetic waves to be amplified may be added, as we will see later. Many types of accelerators have been and are used to produce the electron beams: microtrons, electron storage rings, electrostatic accelerators, room temperature and superconducting radio frequency linear accelerator. They cover an energy range from a few MeV to about 20 GeV. Electron linear accelerators (linacs), at room temperature or superconducting, with an energy from about 1 to 15 GeV are being used for X-ray FELs.

The undulators are mainly of two types (Elleaume, 1990). One is a helical undulator, with a periodic transverse helical magnetic field produced on the axis of a double-helix-wound bifilar magnet with equal and opposite currents in each helix (Kincaid, 1997). The electron trajectory is also a helix around the system axis. The other is a planar undulator (Motz, 1951), produced by alternating dipole magnets, with the field changing in a plane like a sinusoid of period λ_U -usually a few centimeters- and amplitude B_0, typically about 1 Tesla. In this field the electron moves along a sinusoidal, oscillating trajectory in a plane perpendicular to the magnetic field direction. In both cases the electron emits a wave train with a number of periods equal to the number of undulator periods, N_U. The radiation wavelength is equal to the undulator period reduced by a relativistic contraction factor, proportional to the square of the electron energy. This quadratic dependence makes it easy to

change the wavelength from the centimeter or millimeter range for electrons with energy of a few MeV, to about 1Å for energy of about 10 GeV or larger.
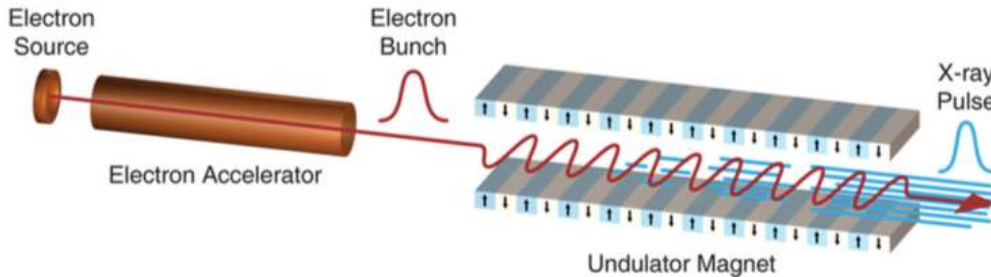

FIG. 1. Schematic representation of a free-electron laser.

Madey and coworkers successfully built in the 1970s the first two FELs, operating at infrared wavelengths. The first (Deacon *et al.*, 1977), shown in FIG. 2, amplified the radiation from a CO_2 laser at a wavelength of 10.6µm. It used a 24 MeV electron beam from a superconducting linear accelerator at Stanford, with current of 5 to 70 mA. The undulator was of the helical type obtained with a superconducting, right-hand double helix, a period of 3.2 cm and a length of 5.2 m. The single pass gain was as large as 7%.

The second experiment (Elias *et al.*, 1976) was an FEL oscillator, operating at a wavelength of $3.4\,\mu m$, beam energy of 43 MeV, the same helical undulator and an optical cavity 12.7 m long, as shown in FIG. 3. The cavity length is chosen so that the back and forth travel time of the light pulse in the cavity is equal to the time separation between electron bunches from the linac. The oscillator starts from the spontaneous radiation generated initially by the electron beam.
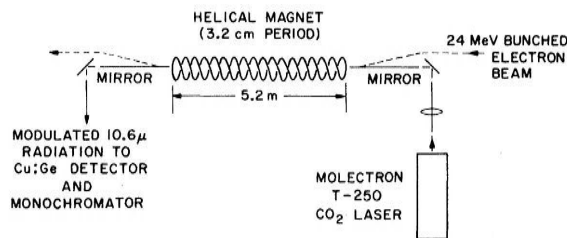
FIG. 2. Madey's amplifier experiment. The undulator is a bifilar superconducting coil. From Deacon *et al.*, 1977.

The theoretical framework for this work was Madey's low gain theory (Madey, 1971), using quantum theory to describe the FEL process as stimulated bremsstrahlung in a periodic magnetic field. A remarkable result was that, even starting from quantum theory, the final gain did not depend on Planck's constant. The FEL gain is a classical effect.

The works of Motz (Motz 1953) on undulator radiation, Pantell, Soncini and Puthoff (Pantell, Soncini and Puthoff, 1968) on stimulated Compton scattering, Phillips on the Ubitron (Phillips, 1960), are a precursor to that of Madey. Palmer (1972), Robinson (Robinson 1985), whose paper was published posthumous, and Csonka (Csonka 1978) explored similar ideas.

The experiments by Madey and co-workers stimulated the interest of many scientists, and led to the construction of more FELs in the infrared and visible region. However the scaling with wavelength of the low signal gain was not favorable for extending the FEL operation to shorter wavelengths.

The next important step was the development of the high gain theory (Kroll and McMullin, 1978; Sprangle and Smith, 1980; Gover and Sprangle, 1981; Dattoli, Marino, Renieri and Romanelli, 1981; Bonifacio, Casagrande and Casati, 1982; Bonifacio, Pellegrini and Narducci, 1984; Gea-Banacloche, Moore and Scully, 1984; Sprangle, Tang and Roberson, 1985; Jerby and Gover, 1985; Kim 1986a; Wang and Yu, 1986; Bonifacio, Casagrande and Pellegrini, 1987) and three-dimensional theory (Kondradenko and Saldin, 1980; Moore 1984; Moore 1985; Scharlemann, Sessler and Wurtele 1985; Kim1986b; Krinsky and Yu 1987; Yu, Krinsky and Gluckstern, 1990) including diffraction effects.

In the one-dimensional high gain theory the FEL process can be seen as a collective instability of the electron beam-electromagnetic wave system described by a single quantity summarizing the electron and undulator characteristics: the FEL parameter (Bonifacio, Pellegrini and Narducci, 1984). The instability can start from noise at the radiation wavelength in the electron beam longitudinal density distribution, taking the beam from a disordered initial state to one with electrons organized in micro-bunches separated by one radiation wavelength, a kind of relativistic one dimensional electron crystal. This transition, starting from noise, characterizes a SASE FEL. While the radiation intensity from an electron beam with a random longitudinal distribution is proportional to the electron number, that from an ordered beam, with all electrons emitting in phase, can be proportional to the square of the electron number, a very large increase.

The validity of the SASE FEL theory was initially demonstrated by an experiment in the microwave region, at a frequency of about 30 GHz, by a Livermore-Berkeley group (Orzechowski *et al*.1985), and later in two experiments, at infrared wavelength, by a UCLA-Kurchatov group (Hogan *et al*. 1998a) and a UCLA-Los Alamos-Kurchatov group (Hogan *et al*., 1998b). The last experiment, at a wavelength of 12 μm, gave a gain larger than 3×10^5, as shown in FIG. 4. Since the Bonifacio, Pellegrini and Narducci, 1984, SASE theory of is independent of wavelength, this result gave confidence that an X-ray FEL based on this concept is feasible.
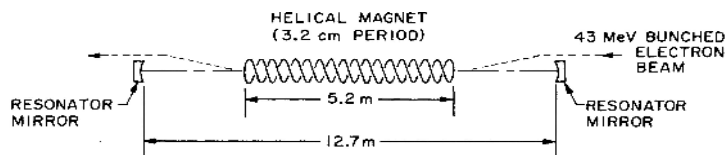
FIG. 3. Madey's FEL oscillator configuration. From Elias *et al*. 1976.

The scaling laws with wavelength of a high gain SASE FEL are much more favorable that those derived from the small signal gain case (Pellegrini, 1988). Starting from noise and amplifying to saturation in a single undulator pass eliminates the need for low loss optical cavities needed for an oscillator configuration. These elements were the key to successfully design and construction of an X-ray FEL. A much more detailed discussion of how LCLS, the first Ångstrom wavelength X-ray FEL, successfully reached a fully operational status in 2009, from the first initial proposal (Pellegrini, 1992), is given in Pellegrini, 2012.
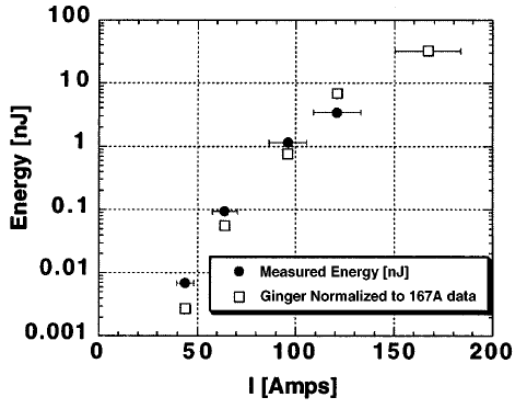
FIG. 4. Average FEL output energy, in nJ, as a function of electron beam peak current, compared with Ginger simulations for the UCLA-Los Alamos-Kurchatov FEL. The strong nonlinear dependence is in good agreement with SASE FEL theory. From Hogan *et al.*, 1998b.

In our discussion of FEL physical properties and theoretical models we proceed in a series of steps, each using some approximations. Some of these approximations are removed later in the paper or in the simulation codes, in other cases we maintain that they are justified and have a negligible effect on the final results. Most of our work is based on a classical description of the FEL. We will discuss the limits of this approximation, when it fails, and refer to other papers on this subject.

The first step, in section I, is to evaluate the electron trajectories in undulator magnets and evaluate the spontaneous radiation emitted by one or many electrons traversing an undulator. In section II we derive the FEL equations describing the interaction between the electron beam and the electromagnetic radiation field. In section III we discuss a simple physical picture of the FEL interaction and write the FEL equations in a universally scaled form depending only on one quantity, the FEL parameter. In this model all FEL main characteristics, like gain length and saturation power, are described by a single parameter, a function of the electron beam energy, its density and of the undulator magnet period and magnetic field. This form of the theory is useful because its validity can be proved experimentally at any wavelength. In section IV we develop the 1-dimensional theory of the FEL in the small gain and high gain regimes, and obtain the properties of a SASE FEL. In section V we consider three-dimensional and diffraction effects and discuss the radiation eigenmodes.

In section VI we discuss high efficiency FELs using longitudinally varying undulator magnetic field strength, a tapered undulator. In section VII we consider methods to seed the FELs, improving its longitudinal coherence properties respect to a SASE FEL. Numerical codes developed to simulate FELs, in a SASE, seeded or high efficiency configuration, are discussed in Section VIII. These codes have been very important to design X-ray FELs and to analyze experimental results, including a realistic description of the electron beam phase space distribution and system errors.

In section IX we give a short review of the present status and main physical properties - wavelength range, pulse intensity, duration, transverse and longitudinal coherence- of high gain, X-ray FELs, in operation and under construction. We also discuss some recent work to improve and control the temporal coherence and generate multicolor spectra.

7

I. Undulator magnets and spontaneous undulator radiation: Intensity, angular and frequency distributionEquation Section 1

The fundamental FEL process is the generation of electromagnetic waves by a relativistic electron moving in a periodic magnetic field, called undulator radiation. We consider it here in the classical limit, neglecting quantum and recoil effects. After a general review of the spectral and angular properties of the radiation we evaluate the trajectories of electrons in undulator magnets. From these we derive the expressions for the intensity, angular and spectral properties for the radiation emitted by one electron and generalize the results to the case of many electrons. When discussing electron trajectories we also introduce some important quantities characterizing electron beams and their statistical phase space distribution function.

1. Lienard-Wiechert fields, total radiated power, frequency and angular distribution

In the classical limit our system is described by Maxwell equations for the electromagnetic field and Lorentz equation for the electron motion. In this section, we summarize the equations describing the electromagnetic field generated by a moving charge. Here and in rest of the paper we use Gaussian units.

The electromagnetic field generated by one electron of charge e, position $\vec{r}_e(t)$ and velocity $c\vec{\beta}_e(t)$, where c is the light velocity, in an inertial reference frame, is described by the Lienard-Wiechert scalar and vector potentials, (Jackson, 1998a)

$$
\begin{aligned}
\Phi(\vec{r},t) &= e/(R - \vec{\beta}_e \cdot \vec{R}), \\
\vec{A}(\vec{r},t) &= e\vec{\beta}_e/(R - \vec{\beta}_e \cdot \vec{R}),
\end{aligned}
\tag{I.1}
$$

where the position and velocity vectors are evaluated at the retarded time t', $\vec{R}$ is the vector from the electron position at time t' to that of the observer at time t, as shown in FIG. 5. The observer and retarded times are related by

$$
\vec{R}(t') = \vec{r} - \vec{r}_e(t'),
\tag{I.2}
$$

$$
t' = t - R(t')/c.
\tag{I.3}
$$

The electric field at the observer position and time is obtained from the potentials and is given by (Jackson 1998b)

$$
\vec{E}(\vec{r},t) = \frac{e(\vec{n} - \vec{\beta}_e)}{\gamma^2 R^2(1 - \vec{n} \cdot \vec{\beta}_e)^3} + \frac{e\vec{n} \times [(\vec{n} - \vec{\beta}_e) \times \dot{\vec{\beta}}_e]}{cR(1 - \vec{n} \cdot \vec{\beta}_e)^3}
\tag{I.4}
$$

where $\vec{n} = \vec{R}/R$ and the electron position, velocity and acceleration are evaluated again at the retarded time.
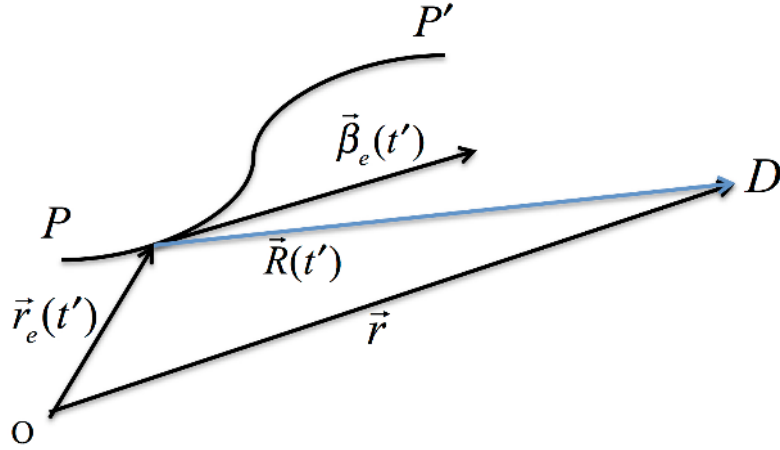
FIG. 5. The electron moves on the trajectory PP', the observer is at point D and measures, at time t, the fields generated by the electron at time t'. The point O is the origin of the reference frame.

The magnetic field is given by

$$\vec{B}(\vec{r},t) = \vec{n} \times \vec{E}(\vec{r},t) \qquad (I.5)$$

The first term in (I.4) is the velocity field, giving the Coulomb field in the limit of velocities smaller than the light velocity, $\beta_e \ll 1$. The second term is the radiation term, proportional to the acceleration, and orthogonal to $\vec{R}$ and $\vec{B}$.

The total power and the angular and frequency distributions of the electromagnetic fields generated by the moving charge are evaluated from (I.4), (I.5). The results are conveniently expressed using the velocity and acceleration evaluated at the particle time. The total energy loss is given by the Larmor-Lienard formula (Jackson, 1998c)

$$P = \frac{2r_e}{3c} mc^2 \gamma^6 [\dot{\beta}^2 - (\vec{\beta} \times \dot{\vec{\beta}})^2], \qquad (I.6)$$

where we eliminated the sub-script e from the electron velocity and acceleration, $r_e = 2.82 \times 10^{-15} m$ is the classical electron radius, $mc^2 = 5.11 \times 10^5 eV$ the electron rest energy, γ the electron energy in rest energy units.

The far field power radiated per unit solid angle, in terms of the observer time, is (Ellaume, 1983; Jackson, 1998d)

$$\frac{dP}{d\Omega} = \frac{e^2}{4\pi c} \frac{\left| \vec{n} \times [(\vec{n} - \vec{\beta}) \times \dot{\vec{\beta}}] \right|^2}{(1 - \vec{n} \cdot \vec{\beta})^5}. \qquad (I.7)$$

9

For highly relativistic particles, $\gamma \gg 1$, the power angular distribution depends strongly on the emission direction respect to the particle velocity. When observing along the velocity direction the denominator is very small, $1 - \beta = 1 / (1 + \beta)\gamma^2$. If θ is the angle between the direction of observation and the particle velocity the denominator can be approximated, for small angles, as $1 - \vec{n} \cdot \vec{\beta} = 1 - \beta \cos\theta \approx (1 + \gamma^2 \theta^2) / 2\gamma^2$ and the radiation is mostly in a cone of aperture $\theta \approx 1 / \gamma$.

The intensity angular-frequency distribution is given by (Jackson, 1998e)

$$\frac{dI}{d\omega d\Omega} = \frac{e^2 \omega^2}{4\pi^2 c} \left| \int_{-\infty}^{\infty} \vec{n} \times (\vec{n} \times \vec{\beta}) e^{i\omega(t - \vec{n}\cdot\vec{r}(t)/c)} \, dt \right|^2$$

$$\equiv \frac{e^2 \omega^2}{4\pi^2 c} \left| \vec{M}(\omega) \right|^2 .$$

(I.8)

The vector $\vec{M}(\omega)$ depends on the observation direction and on the electron trajectory evaluated at the particle time. In the last formula the electron acceleration does not appear explicitly, and the integration is extended over all times. If the acceleration is non zero only over a finite interval of time, say from $t = 0$ to $t = T$, one has to add and subtract the integrals over the time when the velocity is constant, removing any ambiguity by inserting a convergence factor $\exp(-\varepsilon |t|)$ and taking the limit $\varepsilon \to 0$ after evaluating the integral.

2. Magnetic fields in helical and planar undulators

The two most common types of periodic undulator magnets are helical and planar undulators with constant period and peak magnetic field along the axis (Elleaume, 1990). Electrons moving in a helical undulator produce circularly polarized radiation, while for planar undulator the radiation is plane polarized in the plane of the electron oscillations. Undulators that can vary from a helical to a planar configuration have also been built (Sasaki, Miyata and Takada, 2002) and are used at synchrotron radiation sources and FELs. Other undulators that have been studied and used are tapered field undulators (Kroll, Morton and Rosenbluth, 1979), with variable field and/or period along the axis, magnetic field transverse gradient undulators (Smith *et al.,* 1979), and a combination of undulators, as for instance in the case of the optical klystron (Vinokurov and Skrinsky, 1982). In this section we discuss electron trajectories and radiation properties for the helical and planar cases.

The simplest way to build a helical, constant period undulator is to use two helical coils with equal and opposite currents. Permanent magnets can also be used to generate a helical undulator. The helical undulator magnetic field, using cylindrical coordinates $x = r\cos\theta$, $y = r\sin\theta$, z along the undulator axis, is (Kincaid, 1997)

$$B_r = B_0 [I_0(k_U r) + I_2(k_U r)] \sin(\theta - K_U z),$$

(I.9)

$$B_\theta = B_0 [I_0(k_U r) - I_2(k_U r)] \cos(\theta - K_U z),$$

(I.10)

$$B_z = -2B_0 I_1(k_U r) \cos(\theta - K_U z), \qquad (I.11)$$

where λ_U is the undulator period, $k_U = 2\pi / \lambda_U$ and B_0 is the magnetic field on axis. The functions I_0, I_2 are modified Bessel functions. The value of the field on axis depends on the technology used to generate the field. For a double helix bifilar magnet with equal and opposite currents one has

$$B_0 = \frac{4k_U I}{10}\Big[k_U a K_0(k_U a) + K_1(k_U a)\Big] \qquad (I.12)$$

with the field in Gauss, the current I in A, the radius of the helix a and the undulator period in cm, and K_0, K_1 modified Bessel functions (Smythe, 1950).

Near the axis the field can be approximated to lowest order with $I_o = 1$, $I_1 = k_u r / 2$, $I_2 = 0$, giving, in Cartesian coordinates,

$$\begin{aligned}
B_x &= -B_0 \sin(k_U z), \\
B_y &= B_0 \cos(k_U z), \\
B_z &= -B_0 k_U [x\cos(k_U z) + y\sin(k_U z)].
\end{aligned} \qquad (I.13)$$

The planar undulator was the first built and used to produce radiation from high-energy electrons by Motz (1951), at Stanford University. It is built with an array of electromagnets or permanent magnets alternating North and South poles. The field is given by

$$\begin{aligned}
B_x &= 0, \\
B_y &= B_0 \cosh(k_U y)\cos(k_U z), \\
B_z &= -B_0 \sinh(k_U y)\sin(k_U z),
\end{aligned} \qquad (I.14)$$

and can be approximated near the axis by

$$\begin{aligned}
B_x &= 0, \\
B_y &= B_0 \cos(k_U z), \\
B_z &= -B_0 k_U y \sin(k_U z).
\end{aligned} \qquad (I.15)$$

The strength of the undulator field increases as one moves off-axis. In the helical case this is true in both transverse directions. In the planar case, the field only increases along the main direction of the field, the *y*-direction with our choice of coordinates. As we discuss in the following section this change in field when moving off-axis gives a transverse focusing force on the electron beam. However at the high electron beam energy of X-ray FEL, a few to 10-20 GeV, this focusing is too weak and the undulator too long to transport the beam maintaining a small beam transverse area and avoid losses. For example the LCLS undulator is about 120m long. Such a long undulator is built in sections a few meters long, 3.4m for LCLS, separated by a section where vacuum pumps, beam and radiation diagnostics, and quadrupole

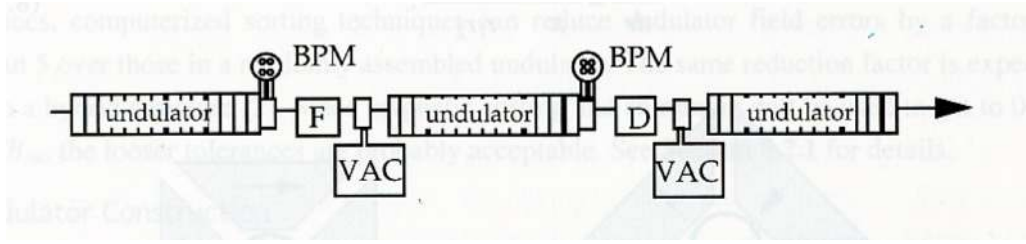magnets are installed (LCLS Design Study Group, 1998), as shown in FIG. 6.

FIG. 6. LCLS undulator schematic view, showing its various components: undulator sections, focusing, F, and defocusing, D, quadrupoles, vacuum pumps and beam position monitors. From LCLS Design Study Group, 1998.

The external focusing field generated by the array of focusing, F, and defocusing, D, quadrupoles is used to focus the beam through a long undulator and optimize the FEL gain. The trajectory of an off axis electron through the quadrupole system is given by (Courant and Snyder, 1958)

$$x(z) = \sqrt{2J_x \beta_{B,x}(z)} \cos(\Phi_x(z)),$$
$$y(z) = \sqrt{2J_y \beta_{B,y}(z)} \cos(\Phi_y(z)),$$

$$(I.16)$$

where the two functions $\beta_{B,x}(z), \beta_{B,y}(z)$, the betatron focusing functions, are periodic with the array periodicity, L_Q . The quantity $\Phi_{x,y}(z) = \int^z dz' / \beta_{B,x,y}(z') + \phi_{x,y}$ is the betatron oscillation phase. $J_{x,y}$, $\phi_{x,y}$ are obtained from the initial displacement and angle respect to the axis and are constants of the motion. For equal strength in the D and F quadrupole the external focusing is the same in the horizontal and vertical directions.

For applications to X-ray FELs the quadrupole array is designed to have a small change in the oscillation amplitude in the undulator and between the minimum and maximum values of the betatron focusing functions. This condition requires a small phase advance per period, and an average beta focusing function much larger than the distance separating the quadrupoles, $\overline{\beta}_{x,y} \gg L_Q$. In this case we can use the smooth focusing approximation (Rosenzweig, 2003), using for the beta functions its average value, assuming constant focusing in both directions, and writing the phases in (I.16) as $\Phi_{x,y} = z / \overline{\beta}_{B,x,y}$.

3. Electron trajectories in a helical undulator

The electron trajectories and the characteristics of the emitted radiation have been discussed in many papers in the 1970s and 1980s. See, for example, Alferov, Bashmakov and Bessonov, (1974), Colson (1977a), Murphy and Pellegrini (1990). Neglecting the energy loss due to the emission of radiation the energy of an electron moving in an undulator is constant. The effect of energy exchange between the electrons and an electromagnetic radiation field will be considered later. We consider only the motion near the axis.

A system of localized focusing and defocusing quadrupole doublets is added to the undulator magnetic field to provide focusing for the electron beam transport through the undulator, as shown in FIG. 6. To simplify the treatment we assume that the betatron focusing functions in (I.16) are much larger than the undulator period and larger than the separation between quadrupoles. The betatron focusing functions can then be considered to be constants, equal to their average value. With these assumptions the quadrupole array system is equivalent to a constant focusing channel in both transverse directions. Let $E = mc^2\gamma$ be the electron energy and $\vec{\beta}$ its velocity measured in units of the light velocity, c. The time independent Hamiltonian for a particle moving in the static magnetic undulator field is a constant of the motion, equal to the electron energy

$$H = c\sqrt{\left(\vec{P} - \frac{e}{c}\vec{A}\right)^2 + m^2 c^2} = mc^2\gamma, \tag{I.17}$$

where $\vec{A}$, $\vec{P}$ are the vector potential and the canonical momentum. The mechanical momentum is related to the canonical momentum and the vector potential by

$$\vec{P} - \frac{e}{c}\vec{A} = mc\vec{\beta}\gamma. \tag{I.18}$$

The magnetic field is the sum of two terms, one describing the undulator magnetic field given in (I.13), and one describing the focusing force due to the quadrupole array. The undulator vector potential is

$$A_{U,x} = (B_0 / k_U)\left[\left(1 + k_U^2(x^2 + 3y^2)/8\right)\sin(k_u z) + k_U^2(xy/4)\cos(k_u z)\right],$$
$$A_{U,y} = -(B_0 / k_U)\left[\left(1 + k_U^2(3x^2 + y^2)/8\right)\cos(k_u z) + k_U^2(xy/4)\sin(k_u z)\right], \tag{I.19}$$
$$A_{U,z} = 0.$$

Using the approximation of constant focusing, the external force is described by an effective potential

$$A_{F,x} = 0, \; A_{F,y} = 0, \; A_{F,z} = -G_F(x^2 - y^2)/2, \tag{I.20}$$

where the term G_F is the strength of the focusing channel in the transverse direction.

Neglecting second order terms the transverse equations of motion are

$$m\gamma\ddot{x} = -eB_0\beta_z\cos(k_U z) - e\beta_z G_F x - eB_0 k_U \beta_y \{x\cos(k_U z) + y\sin(k_U z)\}, \tag{I.21}$$

$$m\gamma\ddot{y} = -eB_0\beta_z\sin(k_U z) + e\beta_z G_F y + eB_0 k_U \beta_x \{x\cos(k_U z) + y\sin(k_U z)\}. \tag{I.22}$$

To solve these equations we consider highly relativistic particles, moving near the undulator axis, with longitudinal velocity near the light velocity, small transverse velocity

$$\beta_z \simeq 1, \ \beta_{x,y} << 1 \tag{I.23}$$

and use a perturbation technique. The zero order solution is obtained neglecting the terms proportional to the transverse displacements and velocities. We then have

$$\beta_{x0}(z) = \dot{x}_0 / c = -(K/\gamma)\sin(k_U z) + \beta_{x0}(0), \tag{I.24}$$

$$\beta_{y0}(z) = \dot{y}/c = (K/\gamma)(\cos(k_U z) - 1) + \beta_{y0}(0), \tag{I.25}$$

$$z_0 = \beta_z ct + z_0(0), \tag{I.26}$$

with

$$K = eB_0 / mc^2 k_U . \tag{I.27}$$

The quantity K, called the undulator parameter, is the normalized undulator vector potential amplitude and has an important role in FEL theory. In most case the undulator period is a few centimeters, the field can be as large as a Tesla and the undulator parameter is of the order of one. We assume that the initial velocities are $\beta_{x0} = 0$, $\beta_{y0} = K/\gamma$, and use this solution to evaluate the last terms in the equation of motion, describing the focusing effect of the external focusing system and of the undulator magnetic field.

Using our initial conditions we have

$$\beta_{x0}^2 + \beta_{y0}^2 = K^2 / \gamma^2 . \tag{I.28}$$

The transverse velocity has constant magnitude. The longitudinal velocity

$$\beta_{z0} = \sqrt{1 - (1 + K^2)/\gamma^2} \simeq 1 - (!+ K^2)/2\gamma^2 \tag{I.29}$$

is also constant. Integrating (I.24), (I.25) we obtain the zero order electron trajectories

$$x_0 = a\cos(k_U z), \ y_0 = a\sin(k_U z) \tag{I.30}$$

with the z-coordinate given by (I.26). These are the equations of a helix with period λ_U and radius

$$a_{helix} = K / k_U \beta_{z0} \gamma . \tag{I.31}$$

The acceleration is

$$a_{x0} = -(cK/a_{helix}\gamma)^2 x_0, \ a_{y0} = -(cK/a_{helix}\gamma)^2 y_0, \ a_{z0} = 0 . \tag{I.32}$$

The electrons will oscillate around this trajectory if the initial conditions are different from our previous assumption, $\beta_{x0} = 0$, $\beta_{y0} = K/\gamma$. Using equations (I.21), (I.22), averaging over the undulator period - approximation valid because of the smallness of these terms- we obtain for the displacements $\xi = x - x_0$, $\zeta = y - y_0$ from the zero order solutions the equations

$$\frac{d^2\xi}{c^2 dt^2} = -(\Omega_{ext}^2 + \Omega_{und}^2)\xi ,$$

(I.33)

$$\frac{d^2\varsigma}{c^2 dt^2} = -(\Omega_{ext}^2 - \Omega_{und}^2)\varsigma .$$

(I.34)

The focusing strengths are

$$\Omega_{ext}^2 = eG_F \beta_z / mc^2 \gamma ,$$

(I.35)

$$\Omega_{und}^2 = eB_0 K k_U / 2m\gamma^2 c^2 .$$

(I.36)

The undulator focusing strength can be rewritten as

$$\Omega_{und} = K k_U / \sqrt{2}\beta_{z0}\gamma .$$

(I.37)

and is a periodic function as discussed before. For multi-GeV X-ray FELs the undulator focusing term is weak and to a good approximation the external focusing dominates. The solutions for the betatron equations can then be written in the form (I.16).

The oscillations described by (I.33), (I.34) are called betatron oscillations. Their wave number,

$$\beta_{B,F} = 1/\sqrt{\Omega_{ext}^2 + \Omega_{und}^2} ,$$

(I.38)

is used to measure the strength of the beam focusing in the undulator.

For an ensemble of particles executing betatron oscillations, and neglecting effects like particle-particle scattering and dissipative forces, the volume in the 6-dimensional phase-space of coordinates (x, y, p_x, p_y, z, p_z) is conserved. In our approximation of small transverse momenta we can write, using the angle of the momentum respect to the undulator axis, $p_{x,y} = mc\gamma\theta_{x,y}$. For uncoupled motion in the three coordinates and constant beam energy, there are three separate invariants, proportional to the areas in the planes $(x, \gamma\theta_x)$ or $(y, \gamma\theta_y)$, for the transverse coordinates and (z, p_z). These invariants are called the transverse and longitudinal normalized beam emittance, $\varepsilon_{N,x,y}$, $\varepsilon_{N,L}$, defined as

$$\varepsilon_{N,x,y} = \sigma_{x,y}\sigma_{\theta_x,\theta_y} ,$$

(I.39)

$$\varepsilon_{N,L} = \sigma_z\sigma_p ,$$

(I.40)

where the σ s characterize the beam transverse size, angular spread, bunch length and longitudinal momentum spread.

Other quantities used to characterize the electron beam are the geometric emittances, usually simply called the emittance, equal to the normalized emittance divided by the relativistic factor γ for the transverse case, and by the momentum for the longitudinal case. For a complete and recent discussion of electron beam transport, focusing, collective effects, phase space properties and characterization, see the recent review by Di Mitri and Cornacchia (2014).

4. Electron trajectories in a planar undulator

The motion in a planar undulator can be discussed in the same way we used for a helical one. In this case the magnetic field is given by (I.15). Expanding to second order terms and with the addition of a focusing term we obtain

$$B_x = -G_F y, \; B_y = B_0 \cos(k_U z) + G_F x, \; B_z = -B_0 k_U y \sin(k_U z) \; \text{(I.41)}$$

The equations of motion are

$$m\gamma \ddot{x} = -eB_0 \beta_z \cos(k_U z) - e\beta_z G_F x - eB_0 k_U y \beta_x \sin(k_U z) \quad \text{(I.42)}$$
$$m\gamma \ddot{y} = e\beta_z G_F y + eB_0 k_U y \beta_x \sin(k_U z) \quad \text{(I.43)}$$

Following the same procedure and with the same approximations used before we obtain, to lowest order

$$\beta_{x0} = -(K/\gamma)\sin(k_U z), \quad \text{(I.44)}$$
$$\beta_{y0} = 0, \quad \text{(I.45)}$$
$$\beta_{z0} = 1 - \{1 + (K^2/2)[1 - \cos(2k_U z)]\}/2\gamma^2. \quad \text{(I.46)}$$

The main difference between the helical and the planar undulator case appears in the longitudinal velocity, a constant in the first case, and modulated at twice the undulator periodicity in the second case. This modulation introduces in the radiation spectrum on axis all odd harmonics of the fundamental frequency, absent in the helical case, as we will see in Sections I-6, I-8.

To obtain the position along the undulator as a function of time we integrate (I.46) and obtain an approximate solution, to order $1/\gamma^4$,

$$z = \bar{\beta}_o ct - S_p \sin(2k_U \bar{\beta}_o ct), \quad \text{(I.47)}$$

where

$$\bar{\beta}_o = 1 - (1 + K^2/2)/2\gamma^2 \quad \text{(I.48)}$$

is the average longitudinal velocity over an undulator period and $S_p = K^2/(8k_U \bar{\beta}_o \gamma^2)$.

16

The oscillations around the zero order trajectory due to a mismatch in the initial conditions are described by equations similar to (I.33), (I.34) with the elastic force coefficients now given again by (I.35) for the external focusing case and by

$$\Omega_{und,x} = 0, \ \Omega_{und,y} = K k_U / \sqrt{2}\gamma \tag{I.49}$$

for the undulator focusing. There is no undulator focusing in the x-direction and in the y-direction there is the same focusing as in the helical case. By shaping the magnetic poles some of the undulator focusing can be shifted between the horizontal and vertical planes (Scharlemann, 1985). For X-ray FELs, operating at electron beam energies of a few to 20 GeV, the undulator focusing is not large enough to provide a good transport thorough the undulator, external focusing is used and is dominant.

The phase space distribution for the electron beam can be represented as in the helical case using the beam transverse and longitudinal emittances.

5. Spontaneous emission and retardation effects

The electromagnetic field generated by a charged particle moving in an undulator magnet is obtained from the Lienard-Wiechert field (I.4). The electron trajectory consists of three parts. The first and last, the motion outside the undulator, have constant velocity directed along the undulator axis. The second is a periodic, helical or sinusoidal, motion around the same axis.

For an observer, near the undulator axis, at a distance larger than the undulator length, the far field term is generated in the undulator region, where the particle acceleration is different from zero. If the motion in this region is periodic, so is the electromagnetic field.

The most important characteristics of the radiation field can be obtained in a simple way from retardation effects. As an example we consider a helical undulator, but the same results apply to the planar case.

Let $z = 0$, $z = L_U$ be the entrance and exit points of the undulator, with its axis along the z coordinate. Let $z = R_0$, $x = R_0\theta$ be the detector position, assumed to be at a small angle $\theta \ll 1$ respect to the undulator axis. We also assume the electron transverse displacement from the z-axis, the radius of the helix, to be small and neglect it to first order. The electron motion along z in the undulator is given by

$$z = \beta_z ct'. \tag{I.50}$$

Solving (I.3) we have

$$ct' = \frac{ct - \beta_z R_0 - \sqrt{(ct - \beta_z R_0)^2 - (1 - \beta_z^2)(c^2 t^2 - R_0^2 - R_0^2\theta^2)}}{(1 - \beta_z^2)}. \tag{I.51}$$

When the detector is on axis the result is $ct' \simeq (ct - R_0)/(1 - \beta_z)$.

The difference in arrival time at the observer position of two wave fronts emitted at the undulator entrance and exit is given, using (I.51), by

$$c\Delta T = N_U \lambda_U \left(\frac{1-\beta_z}{\beta_z} + \frac{\theta^2}{2} \frac{R_0}{R_0 - N_U \lambda_U} \right) \qquad \text{(I.52)}$$

In the far field limit, $R_0 \gg L_U$, the difference is simply

$$c\Delta T = N_U \lambda_U \left(\frac{1-\beta_z}{\beta_z} + \frac{\theta^2}{2} \right) \qquad \text{(I.53)}$$

For relativistic particles and small angles the delay is much shorter than the time, $N_U \lambda_U / \beta_z c$ to traverse the undulator. When the condition $c\Delta T = N_U \lambda$ is satisfied the waves emitted along the undulator have positive interference, giving a peak in the radiation intensity, at the wavelength

$$\lambda = \lambda_U \left(\frac{1-\beta_z}{\beta_z} + \frac{\theta^2}{2} \right) \qquad \text{(I.54)}$$

For this wavelength we can see from (I.53) that the photon moves ahead of the electron emitting it by one wavelength per period. For the full undulator a phton emitted at the undulator entrance moves ahead by $S = N_U \lambda$, a quantity called the slippage.

The radiation term in (I.4), proportional to the electron acceleration, oscillates at the undulator period evaluated at the retarded time. The field executes N_U oscillations during the time an electron traverses the undulator. For the observer, the time needed for the oscillations is given by (I.53), and the oscillation wavelength is the same as in (I.54).

The light signal duration is related to its frequency spread by the condition (Goodman, 1985) $\Delta\omega\Delta T \geq 2\pi$, giving for the undulator radiation line width

$$\Delta\omega / \omega \approx 1 / N_U . \qquad \text{(I.55)}$$

We can see from the retardation effects that the light emitted by one electron is a wave train of N_U cycles at the wavelength (I.54), with the line width (I.55).

6. Spontaneous radiation by an electron moving in a helical undulator

The characteristics of the radiation emitted by an electron traversing an undulator are important because a spontaneous emission process can become a stimulated process and generate lasing. In this section we review the spectral and angular properties of the undulator radiation, including harmonics, for the two cases of helical and planar undulators.

The transverse electric far field at the detector position, point D in FIG. 5, evaluated from (I.4) is shown in FIG. 7. Only the first 10 waves of a total of 100 are shown. The velocity field is about 10^3 to 10^4 times smaller than the radiation field, as shown in FIG. 8, and can be neglected, even near to the undulator entrance, when we evaluate the radiation outside the undulator.

$\theta=0$
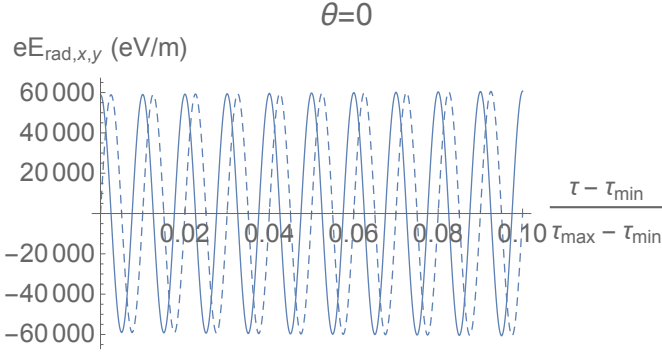
FIG. 7. Radiation electric field *x*, solid line, and *y*, dashed line, components evaluated at a detector on axis at $R_0 = 10m$. The undulator has a 3 cm period, 100 periods and $K = 3$.
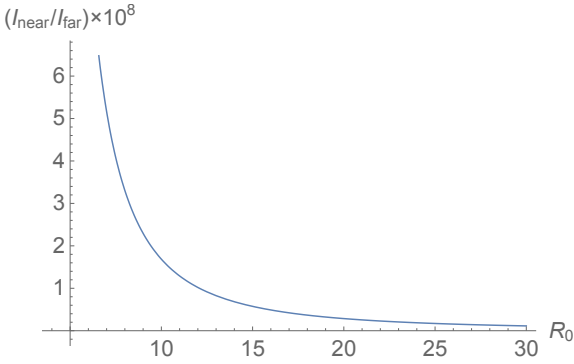
FIG. 8. Ratio of the energy density of the velocity and acceleration fields, for a detector on axis, as a function of the distance R_0 , in m, from the undulator entrance, for a 3 cm period, 100 periods, $K = 3$ undulator. The velocity field intensity is about 10 - 8 times smaller and can be neglected even a a few meters from the undulator exit.

The frequency angular spectrum is evaluated from (I.8). Neglecting the wide band undulator transition radiation that appears because of the sudden change in velocity at the undulator entrance and exit (Kim, 1996), we obtain (Kincaid, 1997)

$$\frac{d^2 I}{d\omega d\Omega} = \frac{e^2 \omega^2 K^2 N_U^2}{c \omega_U^2 \gamma^2} \sum_{n=-\infty}^{\infty} \left[J_n'^2(s) + \left(\frac{\gamma\theta}{K} - \frac{n}{s} \right)^2 J_n^2(s) \right] \left(\frac{\sin\Delta_n(\theta)}{\Delta_n(\theta)} \right)^2 , \qquad (I.56)$$

where

$$\omega_R(\theta) = \frac{2\gamma^2 k_U c}{1 + K^2 + \gamma^2\theta^2}, \ \Delta_n(\theta) = \pi N_U \left(\frac{\omega}{\omega_R(\theta)} - n \right), \quad (I.57)$$

$$s = K\omega\theta / \gamma k_U c = a_{helix}\theta\beta_z\omega / c, \qquad (I.58)$$

19

and a_{helix} is the radius of the helix given by (I.31).

The spectrum is a sum of harmonics of the fundamental. However on axis, $\theta = 0$, only the fundamental is present and is given by

$$\frac{d^2 I}{d\omega d\Omega} = \frac{2r_e mc^2}{c} N_U^2 \gamma^2 \frac{K^2}{(1+K^2)^2} \left(\frac{\sin \Delta}{\Delta} \right)^2 ,$$

(I.59)

where

$$\Delta = \pi N_U (\frac{\omega}{\omega_R} - 1) .$$

(I.60)

A factor ω^2 / ω_R^2 in (I.59) has assumed to be equal to 1, a very good approximation. The FWHM width of the radiation line on axis is

$$\frac{\Delta \omega}{\omega} \simeq \frac{2.8}{\pi N_U} .$$

(I.61)

In FIG. 9 we show the spectrum of the radiation intensity as a function of the relative frequency change respect to the resonant frequency.
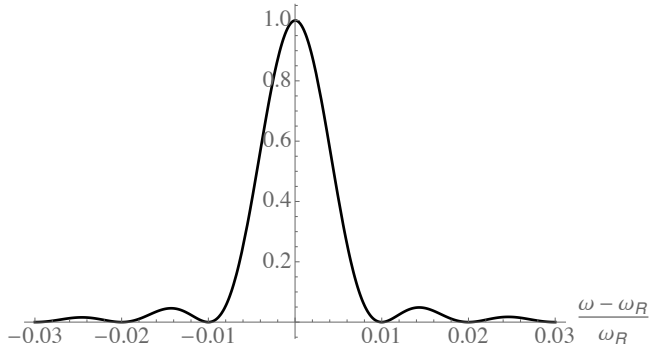
FIG. 9. Spectrum on axis from a 100 periods helical undulator, normalized to the value at the resonant frequency. The intensity is large only near the resonant frequency in a band $\simeq 1/N_U$

The intensity of the off-axis harmonics depends on the angle θ and on the helix radius. It is zero if either of them are zero. For high energy, multi-GeV electrons, as is typically the case for X-ray FELs, the helix radius, (I.31), is of the order of 1 μm, usually small compared to the electron beam radius.

7. *Radiation intensity and coherent photon number in the fundamental line*

We estimate the number of photons emitted in the fundamental line and near the axis, within the line width $1/2N_U$. Since the frequency depends on the emission angle according to (I.57), to remain within this line width the emission angle must be limited to

$$\theta_c = \sqrt{(1+K^2)/2\gamma^2 N_U} = \sqrt{\lambda/\lambda_U N_U} \,, \qquad (I.62)$$

corresponding to a solid angle

$$\Delta\Omega = \pi\theta_c^2 = \pi\lambda/\lambda_U N_U \,. \qquad (I.63)$$

If we consider the transversely coherent photons the effective source radius is

$$\sigma_r = \sqrt{\lambda\lambda_U N_U}/4\pi \,, \qquad (I.64)$$

so that $\sigma_r\theta_c = \lambda/4\pi$, the minimum possible phase space area. Multiplying (I.59) by the solid angle and by the line width we obtain

$$I_c = 2\pi^2 mc^2 \frac{r_e}{\lambda} \frac{K^2}{1+K^2} \,. \qquad (I.65)$$

The corresponding number of photons of energy $E_{ph} = \hbar\omega$ is

$$N_{ph,c} = \pi\alpha \frac{K^2}{1+K^2} \,, \qquad (I.66)$$

where α is the fine structure constant. The number of photons emitted within the fundamental line width and the corresponding solid angle is a few times 10^{-2}. As we will see later, the FEL collective instability can raise this number to about 10^3 at a wavelength of about 1Å, a rather large gain.

8. Spontaneous radiation by an electron moving in a planar undulator

The spectrum for a planar undulator is more complicated than that of a helical undulator because the axial electron velocity, (I.46), and the longitudinal position, (I.47), have a term oscillating at twice the undulator period. The intensity distribution is not azimuthally symmetric and depends on the angle θ respect to the z-axis and the angle ψ in the x,y plane respect to the x-axis. The radiation is polarized in the electron oscillation plane, x,z.

The spectrum is evaluated again using (I.8), neglecting the undulator transition radiation and using the Jacobi-Anger expansion $\exp(i\alpha\sin\beta) = \sum_{m=-\infty}^{\infty} J_m(\alpha)\exp(im\beta)$. The spectrum on axis is a sum of odd harmonics of the fundamental (Alferov, Bashmakov and Bessonov, 1974; Coisson 1981), with a frequency distribution for the n-th harmonic

$$\frac{d^2 I_n}{d\omega d\Omega} = r_e mc N_U^2 \gamma^2 \frac{K^2 n^2}{(1+K^2/2)^2} F_n(K) \left(\frac{\sin \Delta_n}{\Delta_n}\right)^2, \qquad \text{(I.67)}$$

where $\Delta_n = \pi N_U(\omega/\omega_n - 1)$, $\omega_n = 2nck_U\gamma^2/(1+K^2/2) = n\omega_R$,

$$F_n(K) = [J_{(n+1)/2}(\xi_n) - J_{(n-1)/2}(\xi_n)]^2 \qquad \text{(I.68)}$$

and $\xi_n = nK^2/(4+2K^2)$. Contrary to the helical case all the odd harmonics are present on axis and the radiation angular distribution is peaked on axis.

The relative harmonic line-width decreases with the harmonic number,

$$\frac{\Delta\omega_n}{\omega_n} \simeq \frac{1}{nN_U}, \qquad \text{(I.69)}$$

while the absolute one remains the same. The amplitude of the harmonics depends on the undulator parameter, and for $K > 1$ the harmonic intensity can be larger than that of the fundamental as shown in FIG. 10.
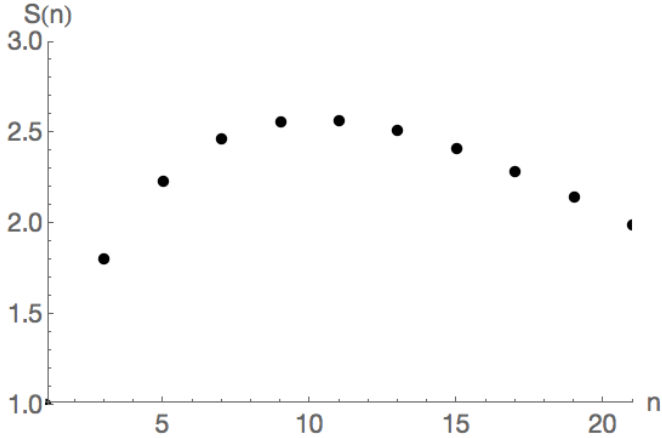
FIG. 10. Intensity at the resonant frequency $\omega = n\omega_R$ normalized to that of the first harmonic, as a function of n. The intensity is evaluated on axis, per unit solid angle and frequency, as given by (I.67). We assume a planar undulator with 100 periods and undulator parameter equal to $K = 3$.

The spectrum off axis is not symmetric around the z-axis, as it is for the helical case. Using polar coordinate around the z-axis, calling θ, Ψ the angles respect to the z-axis and in the x-y plane, taking the limit of large electron energy, $\gamma \gg 1$, and small emission angle θ, we have (Colson, 1981; Ellaume, 1990)

$$\frac{d^2I}{d\omega d\Omega} = 8r_e mc\gamma^2 N_U^2 \times$$

$$\sum \left(\frac{n\xi \sin\nu_n}{\bar{K}\nu_n}\right)^2 \left(\frac{\gamma^2\theta^2}{2\bar{K}^2}A_{0,n}^2 + \frac{\sqrt{2}\gamma\theta}{\bar{K}}A_{0,n}A_{1,n}\cos\Psi + A_{1,n}^2\right), \qquad \text{(I.70)}$$

where $\bar{K} = K/\sqrt{2}$ and

$$A_{m,n} = (-1)^{m+n}\sum(-1)^l J_l(n\xi)[J_{n-m-2l}(nZ)+J_{n+m-2l}(nZ)], \quad \text{(I.71)}$$

$\nu_n = \pi N_U[n-\omega(1+\bar{K}^2+\gamma^2\theta^2)/2\gamma^2\omega_U]$, $Z = \sqrt{2}\chi\cos\Psi$, $\chi = 2\bar{K}/(1+\bar{K}^2+\gamma^2\theta^2)$,
$2\xi = \bar{K}^2/(1+\bar{K}^2+\gamma^2\theta^2)$.

As in the helical case the intensity of the off-axis harmonic depends on the angle θ and on the trajectory amplitude.

9. Radiation from many electrons: effect of position, angular and energy electron distribution

In the previous sections we have discussed the characteristics of the electromagnetic radiation emitted by one electron. We consider now the characteristics of the radiation emitted by many electrons following similar but not exactly equal trajectories.

The electric field generated by an ensemble of electrons is the sum of the fields generated by each one. We assume that the electrons follow similar trajectories and occupy a small volume in the 3-dimensional space, the electron "bunch" volume. We approximate the distance $\vec{R}_k = \vec{R}_0 - \vec{r}_k$ with $\vec{R}_0$, the common distance from the bunch to the observation point. This means that the unit vector defining the observation direction, $\vec{n} = \vec{R}_0/R_0$, does not depend on the particle. The only particle dependence left is that on the velocity and acceleration.

The double differential spectrum is given by

$$\frac{d^2I}{d\omega d\Omega} = \frac{e^2\omega^2}{4\pi^2 c}\left|\sum_{n=1}^{N_e}\int_{-\infty}^{\infty}\vec{n}\times(\vec{n}\times\vec{\beta}_n)e^{i\omega(t-\vec{n}\cdot\vec{r}_n/c)}\,dt\right|, \qquad \text{(I.72)}$$

where N_e is the number of electrons. For a helical undulator and considering only the radiation on axis we have

$$\vec{M}(\omega) = \sum_{k=1}^{N_e}\int_{-\infty}^{\infty}dt\,\vec{\beta}_{T,k}e^{i\omega(t-\vec{n}\cdot\vec{r}_k/c)}. \qquad \text{(I.73)}$$

The subscript T indicates perpendicular to the undulator axis. We assume that the electrons have different longitudinal positions and a small variation in energy

$$z_n = \beta_{z,n}ct + z_{0,n}, \ \gamma_n = \gamma_0(1+\eta_n), \tag{I.74}$$

where γ_0 is the average electron beam energy.

What is the effect on the spontaneous radiation characteristics of the electron spread in longitudinal position and energy? The change in energy gives a variation in the transverse velocity and in the longitudinal velocity

$$\beta_{z,n} = 1 - (1-2\eta_n)(1+K^2)/2\gamma_0^2, \tag{I.75}$$

$$\vec{\beta}_{T,n} = \vec{\beta}_{T,0}(1-\eta_n). \tag{I.76}$$

Introducing $\hat{M} = iM_x + M_y$ and using (I.73) we have

$$\hat{M}(\omega) = \frac{K}{\gamma_0}\sum_{n=1}^{N_e} e^{-i\omega z_{0k}/c}(1-\eta_n)\int_{-\infty}^{\infty} dt e^{ik_U ct\{(\omega/\omega_R - 1)(1-2\eta_n)-2\eta_n\}}, \tag{I.77}$$

where the resonant energy is defined using the average beam energy and zero angle in (I.57).

For a mono-energetic electron beam, $\eta_n = 0$, the dependence on the longitudinal particle distribution function is given only by the bunching factor

$$B(\omega) = \frac{1}{N_e}\sum_{n=1}^{N_e} e^{-i\omega z_{0n}/c}, \tag{I.78}$$

where N_e is the number of electrons in the bunch. The electrons charge per bunch in an X-ray FEL is typically varying from about ten to a few hundred pC and the corresponding number of electrons is of the order of 10^8 to a few times 10^9. If all particles have the same position the bunching factor is simply

$$|B(\omega)| = 1. \tag{I.79}$$

The same is approximately true if they are all grouped together in a distance small compared to the radiation wavelength.

On the contrary $B(\omega) = 0$ if the particles are equally distributed in space between 0 and λ. The spectral intensity, proportional to the square of the Fourier harmonics of the electric field, is proportional to N_e^2 in the first case, and is zero in the second case. In the first case all particles emit radiation in phase as a single super-particle. In the second case for each electron emitting a wave there is another one emitting a wave shifted in phase by π, so they cancel out.

For a mono-energetic beam, on axis, the double differential spectrum becomes

$$\frac{d^2 I}{d\omega d\Omega} = \frac{2r_e m c^2}{c} N_U^2 \gamma^2 N_e^2 |B(\omega)|^2 \frac{K^2}{(1+K^2)^2} \left(\frac{\sin\Delta}{\Delta}\right)^2, \quad \text{(I.80)}$$

the product of the single electron intensity on axis per unit frequency and solid angle, given by (I.59), multiplied by the square of the number of electrons and the bunching factor. The quantity $N_e^2 |B(\omega)|^2$ describes how the radiation from all electrons is added.

The bunching factor plays a critical role for the intensity of undulator radiation and, as we will see later, in the FEL gain. At the nanometer or sub-nanometer wavelength being considered in this paper there is no simple way to control, in the electron source, the longitudinal electron position to obtain directly a large bunching factor. In fact the generation of electrons at the cathode is itself a random process and the initial electron distribution is dominated by Schottky noise in the emission time of the electron.

Assuming a random uniform distribution of the electron position at the undulator entrance, the bunching factor is a sum of random phasors, well studied in statistical optics (Goodman, 1985), with zero average value

$$\langle B(\omega) \rangle = 0, \quad \text{(I.81)}$$

and mean square value

$$\langle |B(\omega)|^2 \rangle = 1/N_e. \quad \text{(I.82)}$$

The average value brackets in the last two equations are average over many different electron ensembles.

For large value of N_e the joint density function of the real and imaginary part of the bunching factor is a Gaussian and the distribution of the module of the bunching factor is

$$p(|B|) = \frac{|B|}{\sigma^2} e^{-\frac{|B|^2}{2\sigma^2}}, \quad \text{(I.83)}$$

the Rayleigh density function. The average value of the bunching factor modulus is $\langle |B| \rangle = \sqrt{\pi/2}\,\sigma$, with $\sigma = 1/\sqrt{2N_e}$. These statistical properties are the same of polarized thermal sources (Goodman, 1985).

Consider now the effect of energy spread. We assume that the distribution of the longitudinal position and energy are independent, and that the energy distribution around its average value is given by the function $g(\eta)$. The double differential spectrum on axis is then

$$\frac{d^2 I}{d\omega d\Omega} = \frac{e^2 \omega^2 K^2}{4\pi^2 c k_U^2 \gamma_0^2} N_e^2 |B(\omega)|^2 \left| \int_{-\infty}^{\infty} d\eta g(\eta) \int_0^{2\pi N_U} d\tau\, e^{i\tau[(\omega/\omega_R - 1)(1-2\eta) - 2\eta]} \right|^2. \quad \text{(I.84)}$$

If the width of the energy distribution is small compared to the radiation line-width we expect a small difference respect to the mono-energetic beam case. In the opposite case we expect a strong reduction of the intensity per unit frequency. This result is shown in FIG. 11 for the case of a Gaussian energy distribution.
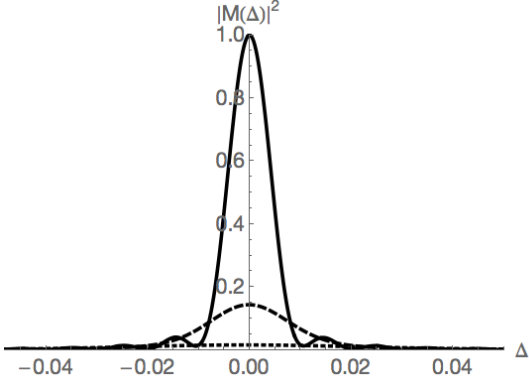
FIG. 11. Relative intensity on axis, near the resonant frequency, as a function of $\Delta = (\omega - \omega_R)/\omega_R$, for an undulator with 100 periods, a Gaussian energy distribution with an rms energy spread of 0.001, 0.003, dashed line, 0.01, dotted line.

10. Qualitative estimate of emittance and energy spread effects

Let us consider an electron moving in a helical undulator and emitting radiation at a central wavelength

$$\lambda = \lambda_U \left(1 + K^2\right)/2\gamma^2 \qquad (I.85)$$

Another electron with different energy, or traversing the undulator at an angle θ_e respect to the axis will have a different central wavelength

$$\lambda' = \frac{\lambda_U}{2(\gamma + \Delta\gamma)^2} \left[1 + \left(K + \Delta K\right)^2 + \gamma^2\theta_e^2\right] \qquad (I.86)$$

The first term in this expression is due to the energy spread. The second term is the change in the undulator parameter, and the last is due to the electron trajectory angle respect to the direction of observation. The change in the undulator parameter can be due to magnetic field errors or to the fact that a particle off axis, executing betatron oscillations, is subject to larger magnetic fields. In fact equation (I.9) shows that the magnetic field in a static undulator is proportional, keeping only the two lowest order terms in the expansion, to $I_0(k_U r) \simeq 1 + (k_U r)^2/2$.

The relative change in wavelength is

$$\frac{\Delta\lambda}{\lambda} = -2\frac{\Delta\gamma}{\gamma} + \frac{2K^2}{1+K^2}\frac{\Delta K}{K} + \frac{\gamma^2\theta_e^2}{1+K^2} \qquad (I.87)$$

The condition for the line width and peak intensity to be changed by a small amount is

$$\frac{\Delta\lambda}{\lambda} < \frac{1}{N_U} \qquad (I.88)$$

The last two terms in (I.87) can be considered equivalent to an energy spread due to the beam distribution in transverse position and angle. Let us consider them in more detail. As in any optical system we can focus the beam to reduce the transverse size, or to decrease the angles. However we cannot reduce the two together because of the conservation of the transverse phase space area. Let the focusing for the electron beam oscillations around the zero order trajectory, (I.33), (I.34), be characterized by the betatron wavelength λ_B, or the inverse betatron wave number $\beta_B = \lambda_B / 2\pi$. The beam transverse phase space area is given by the beam transverse emittance (I.39).

Since the emittance is invariant the transverse beam size and angular spread are related. If the beam is focused to a smaller spot size the angular spread increases or vice versa. The relation between the oscillation amplitude and angle, is given by the inverse betatron wave number, $\sigma_{x,y}^2 = \varepsilon_{x,y}\beta_{B,x,y}$, $\sigma_{\theta_x,\theta_y}^2 = \varepsilon_{x,y} / \beta_{B,x,y}$ (Di Mitri and Cornacchia, 2014). Using this relationship, and assuming for simplicity the same emittance and focusing in the two transverse directions, we can rewrite the last two terms in the expression for the wavelength spread as an effective energy spread (Luccio and Pellegrini, 1980)

$$\left.\frac{\Delta\gamma}{\gamma}\right|_{eff} = \frac{\varepsilon}{1+K^2}(K^2 k_U^2 \beta_{B,F} + \frac{\gamma^2}{\beta_{B,F}}). \qquad (I.89)$$

The minimum value of the effective energy spread as a function of the focusing is obtained when

$$\beta_{B,F} = \frac{\gamma}{k_U K} \qquad (I.90)$$

The previous discussion is only valid in the approximation of a constant external focusing channel, but it can be generalized to the case of a periodic focusing transport channel, as discussed in section V. However (I.89) gives a good first order estimate of the effect.

The effect of beam emittance on the radiation wavelength spread can also be reduced by beam "conditioning" (Sessler, Whittum and Yu, 1992), introducing a correlation between the betatron amplitude and energy to cancel the first two terms in (I.87). One way to introduce the correlation is to change the particle energy using a transverse radio frequency field.

11. Photon brightness

In the previous section we discussed the intensity, spectral and angular properties of the radiation generated by one or many electrons. We introduce now a convenient quantity to characterize the photon beam, the brightness, defined using its six dimensional phase-space volume. Since the phase space volume is invariant under optical transformation (Kim, 1986c) the brightness is a good way to characterize a radiation source. It is also possible to describe the brightness as the Fourier transform of a mutual coherence function of the radiation electric field, describing its transverse coherence properties (Kim, 1986c; Geloni, Kocharyan and Saldin, 2014).

Here we follow the phase-space approach to define the brightness.

The minimum transverse phase space volume for a diffraction limited photon beam is defined, for photons of wavelength λ , by $\Delta x \Delta \theta_x \geq \lambda / 4\pi$, $\Delta y \Delta \theta_y \geq \lambda / 4\pi$ (Siegman, 1986). Considering a Gaussian photon beam and defining the transverse phase space as $2\pi\sigma_{x,y}\sigma_{x',y'} = \lambda / 2$, the expression for brightness for a spatially coherent radiation source is

$$B_{ph} = \frac{N_{ph}}{(\lambda/2)^2 2\pi\sigma_t\sigma_\omega / \omega} \ , \tag{I.91}$$

where N_{ph} is the number of photons in the pulse, with an rms duration and frequency spread σ_t, σ_ω.

If the photon beam is transform limited we also have $\sigma_t\sigma_\omega = 1/2$. However in most practical cases where incoherent light is generated by an electron beam in an undulator, the photon beam is not transform limited and the time-frequency volume is larger than this value. For this reason the minimum value is not used in (I.91).

Consider now the case of many electrons. In the ideal case when the electron beam generating the radiation has a small transverse emittance, smaller than photon beam transverse phase space area

$$\varepsilon_{x,y} \leq \lambda / 4\pi \ , \tag{I.92}$$

the transverse and angular distribution of the photons are still subject to the diffraction limit. Similarly if the electron energy spread is smaller then that of the radiation we can use this quantity to evaluate the brightness. However if the electron beam does not satisfy these conditions we must fold the electron transverse position and angle and the electron energy spread in the definition of brightness. Assuming a Gaussian electron distribution and treating the undulator radiation as a Gaussian mode, the formula for the brightness is

$$B_{ph} = \frac{N_{ph}}{4\pi^2 \Sigma 2\pi\sigma_t\sigma_\omega\omega} \ , \tag{I.93}$$

where the transverse phase space area is

$$\Sigma = \sqrt{\sigma_x^2 + \sigma_r^2}\sqrt{\sigma_{x'}^2 + \sigma_{r'}^2}\sqrt{\sigma_y^2 + \sigma_r^2}\sqrt{\sigma_{y'}^2 + \sigma_{r'}^2} \ , \tag{I.94}$$

$\sigma_{x,y}$, $\sigma_{x',y'}$ are the rms position and angular spread of the electron beam and the radiation beam radius and angular spread are given by (I.62), (I.64). Following the most used convention the brightness is measured in units of number of photons/s/mm²mrad². The relative photon frequency spread is normalized to a "standard" monochromator bandwidth of 0.001 and a factor of 1000 is applied to (I.93). The case of spontaneous planar undulator radiation, which has an azimuthal variation, as opposed to the Gaussian radiation mode considered before, is discussed in (Kim, 1986c). The main effect is to reduce the brightness by a factor of two.
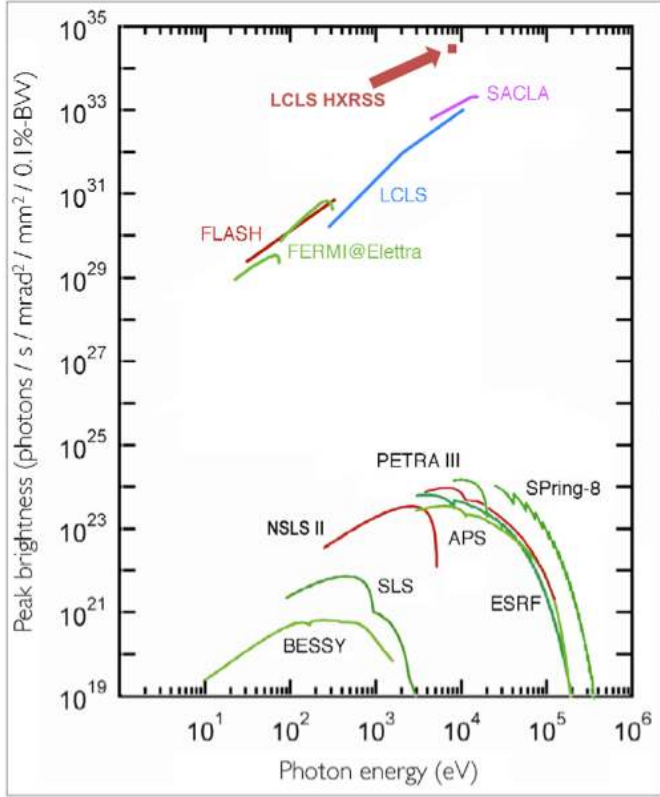
FIG. 12. Peak brightness for undulators in storage rings and operating FELs. From Fletcher et al. (2015).

FIG. 12 shows the peak brightness of storage ring based synchrotron radiation sources and operational X-ray FELs, evaluated using the number of photons and time duration of a single X-ray pulse. The large jump in the peak values, a factor of nine to ten order of magnitudes at about 10 keV photon energy, is due, as we will see later in this paper, to the FEL collective instability increasing the number of coherent photons per electron by a factor of 10^5, a reduction in the pulse time duration by a factor of 10^3 or more, and to a smaller energy spread. The very high brightness point, indicated as HXRSS, is obtained operating LCLS in self-seeding mode, with a smaller line width. This mode of operation will be discussed in Section VII.

Equation Section 2II. The FEL equations

In the previous section we evaluated the spontaneous radiation emitted by a relativistic electron traversing an undulator magnet. Radiation is emitted because the electron is accelerated. In the calculation we assumed that the radiation emitted by the electron does not affect its trajectory, determined only by the undulator magnetic field. In this section we calculate the radiation emitted when the electron moves through an undulator magnet in the presence of a co-propagating electromagnetic wave and consider the energy exchange between the electron beam and the wave. In this process the kinetic energy of the electron is changed and, by conservation of energy, the wave intensity also changes.

To calculate the energy exchange we use a classical approach, using the Lorentz equations to

evaluate the electron trajectories in the combined presence of the undulator magnetic field and the radiation field, and Maxwell equations to evaluate the change in the radiation field driven by the electron beam current.

1. Electron FEL equation in a helical undulator

For simplicity we consider initially the case of a helical undulator and generalize the results to the planar case later. We first consider a one dimensional theory and make some other approximations: the beam transverse size is assumed much larger than the radiation wavelength; the electromagnetic field is described as a plane wave and the beam transverse density distribution is assumed constant; the amplitude and phase of the field, oscillating at the frequency $\omega_r = 2\pi c / \lambda_r$ are slowly changing, and Maxwell equations are simplified accordingly; betatron oscillations and beam emittance effects are neglected. Many of these restrictions will be removed later: a fully 3-dimensional theory is discussed in section V; numerical codes to solve the Lorenz-Maxwell system of equations under realistic conditions, including all known effects, in section VIII.

The E-M Field is described with the vector potential of the undulator and the radiation field

$$\vec{A} = \vec{A}_U + \vec{A}_r . \tag{II.1}$$

We consider circularly polarized plane waves with wave number $k_r = 2\pi / \lambda_r$ and a helical undulator with wave number $k_U = 2\pi / \lambda_U$. The undulator vector potential on axis is the same as in (I.19), using only the lowest order terms

$$\vec{A}_U = \frac{B_0}{k_U}[\sin(k_U z)\vec{x} - \cos(k_U z)\vec{y}].$$

The wave vector potential is assumed to be

$$\vec{A}_r = -\frac{E_r(z,t)}{k_r}[\sin\phi_r \vec{x} + \cos\phi_r \vec{y}], \tag{II.2}$$

where

$$\phi_r = k_r(z - ct) + \psi(z.t) \tag{II.3}$$

and $E_r(z.t),\ \psi(z,t)$ are the slowly varying amplitude and phase of the wave, satisfying the conditions $(1/k_r E_r)\partial E_r / \partial z << 1$, $(1/k_r E_r)\partial E_r / c\partial t << 1$, and $(1/k_r \psi)\partial \psi / \partial z << 1$, $(1/k_r \psi)\partial \psi / c\partial t << 1$.

The electron Hamiltonian is

$$H = c\sum_n [(\vec{P}_n - \frac{e}{c}\vec{A})^2 + m^2 c^2]^{1/2} \tag{II.4}$$

and does not depend on *x* and *y*. Hence the corresponding canonical momenta are constants of the motion.

From these constants we obtain the electron transverse velocity. For a generic electron this is (we omit when not necessary the label n)

$$\beta_x = -\frac{e}{mc^2\gamma}(A_{U,x} + A_{r,x}) + \frac{P_{x,0}}{mc\gamma}, \tag{II.5}$$

$$\beta_y = -\frac{e}{mc^2\gamma}(A_{U,y} + A_{r,y}) + \frac{P_{y,0}}{mc\gamma}. \tag{II.6}$$

Introducing, in addition to the undulator parameter K, the normalized vector potential of the radiation field

$$K_r = \frac{eE_r(z,t)}{mc^2 k_r}, \tag{II.7}$$

we have

$$\beta_x = -\frac{K}{\gamma}\sin(k_U z) + \frac{K_r}{\gamma}\sin\phi_r + \frac{P_{x,0}}{mc\gamma}, \tag{II.8}$$

$$\beta_y = \frac{K}{\gamma}\cos(k_U z) + \frac{K_r}{\gamma}\cos\phi_r + \frac{P_{y,0}}{mc\gamma}. \tag{II.9}$$

The quantity K_r is the work done by the radiation electric field in one radiation wavelength divided by the electron rest mass and in all cases of interest is much smaller than one. It would be equal to one for a field value $eE_r = 2\pi mc^2/\lambda_r$, about $3\times10^{15}V/m$ at a wavelength of 1 nm and larger at shorter wavelength. For this reason from now on, whenever possible, we neglect all terms proportional to this quantity in the equations for the electron motion.

Using the approximation of neglecting emittance effects, we assume for the time being that the two initial values of the canonical momenta, $P_{x,0}$, $P_{y,0}$ are zero. The equations of motion for the longitudinal coordinate and momentum can also be obtained from the Hamiltonian, but it is simpler to use instead the equivalent equations for the change in electron energy and FEL phase,

$$\Phi = k_U z + k_r(z - ct), \tag{II.10}$$

the sum of the phases of the electromagnetic field and of the electron oscillation in the undulator. From the radiation vector potential we obtain the electric field and use it to evaluate the electron energy change. Neglecting terms proportional to the derivatives of the slowly varying amplitude and phase of the radiation field   we obtain the equations for the energy and phase change

$$\frac{d\gamma}{dt} = \frac{ck_r KK_r}{\gamma}\sin(\Phi + \psi), \tag{II.11}$$

$$\frac{d\Phi}{dt} = (k_r + k_U)\dot{z} - k_r c. \tag{II.12}$$

Using (I.26), (I.29) for the longitudinal position and velocity we rewrite the equation for the FEL phase change as

$$\frac{d\Phi}{dt} = k_U \beta_{z0} c \left(1 - \frac{k_r}{k_U} \frac{1+K^2}{2\gamma^2} \right).$$
(II.13)

We define the resonant energy, corresponding to a stationary phase, as

$$\gamma_R^2 = \frac{k_r(1+K^2)}{2k_U},$$
(II.14)

for a given radiation wave number and undulator period and magnetic field. Notice that, here and in the rest of the paper, we use a subscript R to indicate a resonant value of a quantity and r for a quantity referring to the electromagnetic wave. Using this quantity we can rewrite the equation for the phase change as

$$\frac{d\Phi}{cdt} = k_U \beta_{z0} \left(1 - \frac{\gamma_R^2}{\gamma^2} \right).$$
(II.15)

This equation for the FEL phase change and equation (II.11) for the energy change, describe the longitudinal dynamics of the electrons in the FEL process.

The transverse electron betatron motion introduces an extra term in the expression (I.29) of the longitudinal electron velocity and an additional phase change. Following the discussion at the end of Section I-3, we assume that the betatron oscillation wavelength is the same in the two planes and write the betatron oscillation amplitude, the solution of (I.33), (I.34), in the form (I.16) as

$$\xi = \sqrt{2J_x \beta_{B,F}} \sin(\Phi_x(z)),$$
$$\varsigma = \sqrt{2J_y \beta_{B,F}} \sin(\Phi_y(z))$$
(II.16)

To evaluate the longitudinal velocity and phase change introduced by the betatron oscillations we calculate the square of the transverse velocity, assuming that the betatron oscillation wavelength is much larger than the undulator period and the gain length. Under this conditions we can average the square of the velocity over a number of periods of the order of the gain length and obtain (Reiche, 2000; Huang and Kim, 2007) for the phase change the equation

$$\left. \frac{d\Phi}{cdt} \right|_\varepsilon = -\frac{k_r}{2} \frac{J_x + J_y}{\beta_{B,F}} - \frac{k_r K^2 k_U^2}{2\gamma^2} (\xi^2 + \varsigma^2).$$
(II.17)

In the last equation the first term comes from the transverse betatron velocity and the second from the change of the undulator field when moving off-axis. The last term is small at high enough

32

electron energy. We assume that the condition $Kk_U/\gamma \ll 1/\beta_{B,F}$ is satisfied and neglect this term. Including this term in (II.15), the equation for the phase change becomes [Huang and Kim, 2007]

$$\frac{d\Phi}{cdt} = k_U\left[1 - \frac{\gamma_R^2}{\gamma^2} - \frac{k_r}{k_U\beta_{B,F}}(J_x + J_y)\right]. \qquad (II.18)$$

The effect of the betatron oscillation related term is equivalent to an energy spread, as we saw before in section I-9. From this discussion we can see that a theory of high-gain FELs applicable to the X-ray regime, for very large values of the beam energy γ and an external quadrupole focusing system, must in many cases take into account the effect of betatron oscillations on the longitudinal motion and the phase change.

2. Electron FEL equations in a planar undulator

The equation for energy and phase change for a planar undulator can be obtained following a similar procedure. For simplicity we consider only the terms proportional to the undulator magnetic field in the expression for the electron trajectory, neglecting terms proportional to the wave vector potential K_r.

We assume an electric field vector in the same plane of the electron velocity,

$$\vec{E}_r(z,t) = E_r\cos[k_r(z - ct) + \psi]\vec{x}. \qquad (II.19)$$

With the velocity and longitudinal position given by (I.46), (I.47) we obtain for the electron energy change

$$\frac{d\gamma}{dt} = \frac{ecKE_r}{2mc^2\gamma}\{\sin[(k_r + k_U)z - k_rct + \psi] - \sin[(k_r - k_U)z - k_rct + \psi]\}. \qquad (II.20)$$

This equation can also be written, using the Jacobi-Anger expansion, as

$$\frac{d\gamma}{dt} = \frac{ecKE_r}{2mc^2\gamma} \times$$

$$\sum_{m=-\infty}^{\infty} J_m(\xi)[\sin\{[k_r + k_U(1-2m)]\bar{\beta}_0ct - k_rct + \psi\} - \sin\{[k_r - k_U(1-2m)]\bar{\beta}_0ct - k_rct + \psi]\}] \qquad ,(II.21)$$

where $\xi = (k_r + k_U)K^2/8k_U\bar{\beta}_0\gamma^2$. Contrary to the case of a helical undulator, the modulation of the longitudinal position at twice the undulator period allows an energy exchange on axis at all odd harmonics of the fundamental.

The phase equation is

$$\frac{d\Phi}{dt} = k_U\bar{\beta}_0c\left(\pm(1-2m) - \frac{k_r}{k_U}\frac{1-\bar{\beta}_0}{\bar{\beta}_0}\right). \qquad (II.22)$$

For $m > 0$ only the negative sign gives a slowly varying solution at the frequency $k_U(2m-1) = k_r(1-\bar{\beta}_0)/\bar{\beta}_0$, $m > 0$. For $m < 0$ the positive sign gives a slowly varying solution at the frequency $k_U(2|m|+1) = k_r(1-\bar{\beta}_0)/\bar{\beta}_0$, $m < 0$.

Keeping only the terms with a slowly varying phase we obtain for the energy and phase change at harmonic n

$$\frac{d\gamma}{dt} = \frac{ecKE_r}{2mc^2\gamma}F_n(K)\sin(\Phi_n + \psi) \tag{II.23}$$

$$\frac{d\Phi_n}{cdt} = k_U\left[n - \frac{\gamma_R^2}{\gamma^2} - \frac{k_r}{2k_U\beta_{B,F}}(J_x + J_y)\right], \tag{II.24}$$

where the coefficient $F_n(K)$ is given by (I.68). In the last equation we have also included the emittance term, as we did previously for the helical undulator case, equation (II.18). To obtain the last equation we approximated $\bar{\beta}_o$ with 1 when possible.

We can compare the energy and phase change equations for a helical undulator, equations (II.11), (II.18), with the corresponding equations for the first harmonic planar case, (II.23), (II.24). These equations are similar and we can unify their discussion if we make the substitutions $K \rightarrow KF_1(K)/2$ in the energy equations and $K \rightarrow K/\sqrt{2}$ in the equation defining the resonant energy. Notice that the coupling between the electron velocity and the radiation electric field in the energy change equations in the planar case is smaller than that of the helical case. A plot of the coefficient $F_1(K)/2$ versus the undulator parameter is shown in FIG. 13, showing a reduction in energy change in the planar case by a factor between 2 and 3. In the rest of the paper we consider for simplicity the helical case and will use the substitutions just mentioned to use the results for the planar case.
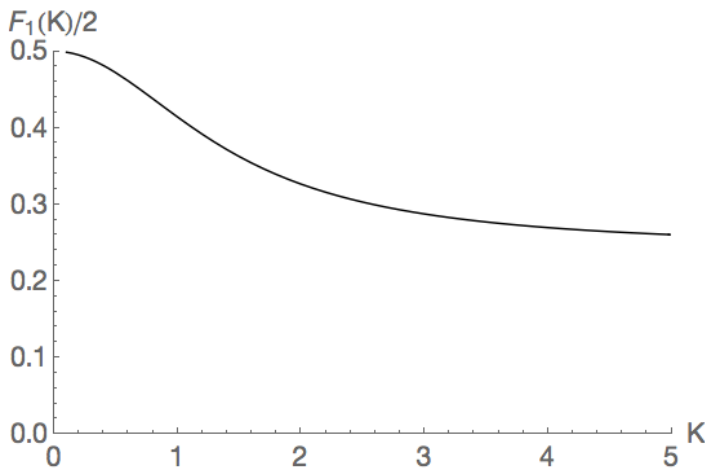

FIG. 13. The first harmonic coefficient $F_1(K)/2$ in the energy exchange equation (II.23).

3. Maxwell equations

Next we need the equations for the field. We write Maxwell equations in the form

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\vec{A} = -\frac{4\pi}{c}\vec{J}, \tag{II.25}$$

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)V = -4\pi\rho. \tag{II.26}$$

The current and charge density are

$$\vec{J}_e(\vec{r},t) = ec\sum \vec{\beta}_n \delta(\vec{r} - \vec{r}_n(t)), \tag{II.27}$$

$$\rho_e(\vec{r},t) = e\sum_{n=1}^{N_e} \delta(\vec{r} - \vec{r}_n(t)). \tag{II.28}$$

We consider first the transverse fields, and introduce the quantities

$$\hat{J}_e = J_{e,y} - iJ_{e,x} = ec\sum_{n=1}^{N_e} \delta(\vec{r} - \vec{r}_n)\frac{1}{\gamma_n}[Ke^{-ik_U z} + K_r e^{i\phi_r}], \tag{II.29}$$

$$\hat{A} = A_{r,y} - iA_{r,x} = i\frac{\alpha}{k_r}e^{ik_r(z-ct)}, \tag{II.30}$$

where

$$\alpha(z,t) = -iE(z,t)e^{i\psi(z,t)}. \tag{II.31}$$

With these definitions, and using the slowly varying amplitude and phase approximation,

$$\frac{1}{k_r\alpha}\frac{\partial\alpha}{\partial z} \ll 1, \quad \frac{1}{k_r\alpha}\frac{\partial\alpha}{c\,\partial t} \ll 1. \tag{II.32}$$

Maxwell equations for the complex field amplitude become

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\alpha - \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\frac{i\alpha}{2k_r} =$$
$$2\pi e\sum_{n=1}^{N_e} \delta(\vec{r} - \vec{r}_n(t))\frac{1}{\gamma_n}\left[Ke^{-i\Phi_n} + \frac{ie\alpha}{mc^2 k_r}\right]. \tag{II.33}$$

This equation, together with the $2N_e$ equations (II.11), (II.18) for the electron energy and phase, are the complete set of FEL equations. This is a rather complicated system of equations and to better understand its physical meaning we discuss initially its solutions in some simplified

cases. The approximations will be removed later to obtain a full description of the system, including the use of numerical simulations methods.

The sum over particles in (II.33) is simplified considering that the problem has two length scales, the bunch length and the radiation wavelength, σ_B, λ_r. In the limit in which all electrons are in a distance smaller than the wavelength $\sigma_B \ll \lambda_r$ all electrons act as a single super particle and the radiation generated is simply that of a single particle tines N_e^2.

This condition can satisfied at long wavelengths, in the millimeter or longer wavelengths, but cannot be satisfied in the X-ray region of the electromagnetic spectrum. In the last case, that of interest to us, we are in a regime where $\sigma_B \gg \lambda_r$. In this case we separate the electron density distribution into two components, a macroscopic distribution that does not change during the time the electrons traverse the undulator, and a microscopic distribution, on the scale of the radiation wavelength, which does change. The term in the current distribution that changes at the microscopic level is the phase Φ_n.

In other words, to evaluate the electron current we take a local average over a length corresponding to one wavelength and we give it a weight factor corresponding to the local macroscopic density, $n_e(x, y, z - \beta_0 ct)$. With these considerations in mind we can rewrite the electron current density in equation (II.33) for the field as

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\alpha - \nabla_T^2 \frac{i\alpha}{2k_r} = 2\pi e n_e(x, y, z - \beta_0 ct)\left[K\left\langle\frac{e^{-i\Phi}}{\gamma}\right\rangle - \frac{ie\alpha}{mc^2 k_r}\left\langle\frac{1}{\gamma}\right\rangle\right], \qquad (II.34)$$

where the symbol $\langle \ \rangle$ indicates an average over the microscopic level distribution and

$$\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \qquad (II.35)$$

The two averages are functions of z, t, the position along the undulator and time. The macroscopic electron density distribution is in general a function of the transverse and longitudinal coordinates and propagates along the undulator with the average electron longitudinal velocity β_{z0}.

In the case of a mono-energetic beam, $\gamma_n = \gamma_0$, the term $\left\langle e^{-i\Phi}/\gamma\right\rangle$ in (II.34), is equal to the bunching factor introduced before in the discussion of the radiation from many electrons. As we will see the evolution of the bunching factor from the initial value, usually very small, to a value near one is a key effect in the FEL dynamics.

We use now the vectors $\vec{x}_B$, $\vec{p}_B$ with components (ξ,ς), $(\beta_B d\xi / cdt, \beta_B d\varsigma / cdt)$ and (I.33), (I.34) to write the betatron oscillation equations. Using also (II.11), (II.18), (II.34) we have the complete set of FEL equations

$$\frac{d}{cdt}\vec{x}_{B,n} = \frac{\vec{p}_{B,n}}{\beta_{B,F}}, \qquad (II.36)$$

36

$$\frac{d}{cdt}\vec{p}_{B,n} = -\frac{\vec{x}_{B,n}}{\beta_{B,F}}, \tag{II.37}$$

$$\frac{d\Phi_n}{cdt} = k_U \beta_{z0}\left[1 - \frac{\gamma_R^2}{\gamma_n^2} - \frac{k_r}{2k_U\beta_{B,F}}(\frac{\vec{x}_{B,n}^2}{\beta_{B,F}} + \beta_{B,F}\vec{p}_{B,n}^2)\right], \tag{II.38}$$

$$\frac{d\gamma_n}{cdt} = \frac{eK}{2mc^2\gamma_n}(\alpha e^{i\Phi_n} + c.c.), \tag{II.39}$$

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\alpha - \nabla_T^2\frac{i\alpha}{2k_r} = 2\pi e n_e(x,y,z-\beta_0 ct)\left[K\left\langle\frac{e^{-i\Phi}}{\gamma}\right\rangle - \frac{ie\alpha}{mc^2 k_r}\left\langle\frac{1}{\gamma}\right\rangle\right], \tag{II.40}$$

for $n = 1,...,N_e$.

This is a system of $6N_e + 1$ nonlinear equations with no simple analytical solution. It includes: radiation diffraction; longitudinal and transverse electron dynamics; betatron oscillations for equal and constant focusing in the horizontal and vertical planes; electron beam distribution in its 6-dimensional phase space. It does not include quantum effects and the effect of the radiation field on the electron velocities. In the following sections we will solve these equations in a number of simplified cases and will later discuss their numerical solutions.

Equation Section 3III. A physical picture of the FEL interaction and scaled FEL equations

In this section we give a physical picture of the FEL interaction and use this insight to write the FEL equations using scaled variables. As we will see the scaled equations greatly simplify the description of the system and, in the one-dimensional case, can be written in a universal form, valid for all wavelengths, using a single parameter.

1 A physical picture of the FEL interaction

Even if the FEL equations are rather complicated we can us a physical pictures to understand the FEL gain mechanism. We will later arrive at the same conclusions by studying their properties analytically.

To increase the peak brightness of undulator radiation we can either increase the electron current or increase the number of photons produced per electrons. An FEL achieves the second goal, increasing the number of coherent photons per electron by about five to seven orders of magnitude.

How do we increase the number of photons emitted per electron? If we have many electrons, say N_e, and they are all grouped within a small fraction of a wavelength, the total intensity at this wavelength would be the single particle intensity times N_e^2, and the number of photons per electron would be increased by the same factor. In practical cases, certainly at X-ray wavelength, the electrons are in a bunch much longer than the radiation wavelength, and their position

distribution on the wavelength scale is completely random. As a result, the radiation fields emitted by different electrons have a random relative phase, and the total intensity is proportional to N_e.

We can, however, increase the number of photons emitted per electron if we take advantage of a collective instability of the electron beam-electromagnetic radiation field-undulator system (Bonifacio, Pellegrini and Narducci, 1984). This instability works as follows:

1. The electron beam interacts with the electric field of the radiation; the electric field is perpendicular to the direction of propagation of the beam (the undulator axis), and is parallel to the wiggling (transverse) velocity of the electrons produced by the undulator magnet, of amplitude K/γ; the interaction produces an electron energy modulation, on the scale λ_r:

$$mc^2 \frac{d\gamma}{dt} = eEV_T \cos[k_r(z-ct)+k_U z]; \qquad \text{(III.1)}$$

2. The electron energy modulation changes the electron trajectory in the undulator magnetic field: high energy electrons have a shorter path than low energy electrons, as shown in FIG. 14, generating bunching of the electrons at the scale λ_r;
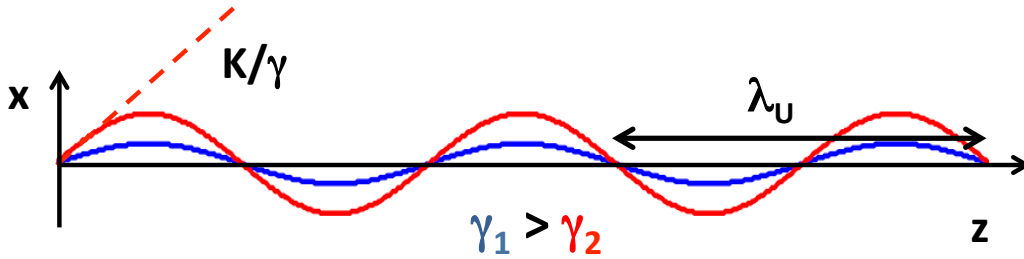
FIG. 14. Electrons trajectories in the undulator, showing a longer path for lower energy electrons respect to higher energy electrons.

3. Electrons bunched within a wavelength emit radiation in phase, thus producing a larger intensity; the larger intensity leads to more energy modulation and more bunching, leading to exponential growth of the radiation; the intensity can reach the limit $I \sim N_e^2$ for the case of extreme bunching.

As we will see later in detail by studying the FEL equations, in the one dimensional limit, neglecting diffraction effects, the process is characterized by two lengths, the gain length, L_G, the radiation intensity exponential growth rate of the instability, and the cooperation length, L_c, the distance over which the radiation emitted by one electron can interact with another electron considering the difference in velocity between electrons and photons. As we have seen in Section I-5 a photon emitted by an electron moves ahead of it, slips ahead, by one wavelength per undulator period. The cooperation length is the slippage in one gain length. The two lengths are proportional to each other $L_c = (\lambda_r / \lambda_U)L_G$. This implies that both gain and coherence length can be

expressed by one parameter only (Bonifacio, Pellegrini and Narducci, 1984). We call this quantity the FEL parameter given, for a helical undulator, by

$$\rho = \left(\frac{K\Omega_P}{4k_U c} \right)^{2/3}, \tag{III.2}$$

where

$$\Omega_P = \left(\frac{4\pi r_e c^2 n_e}{\gamma^3} \right) \tag{III.3}$$

is the beam plasma frequency, n_e the electron density, r_e the classical electron radius. The electromagnetic field instability growth rate, or gain length, is given approximately, as will be shown later in the paper, by

$$L_G \simeq \frac{\lambda_U}{4\pi\rho}. \tag{III.4}$$

The typical value of the FEL parameter for X-ray FELs is about 0.001. The expression of the FEL parameter for a planar undulator will be given later.

The amplitude of the radiation field grows exponentially along the undulator axis, z, as $E = E_0 \exp(z/L_G)$ if the following conditions are satisfied (Pellegrini1988):

a. Beam emittance smaller than the radiation wavelength:

$$\varepsilon < \lambda_r / 4\pi ; \tag{III.5}$$

b. Beam relative energy spread, including the effective term (I.89) depending on the emittance and the focusing beta function, smaller than the FEL parameter ρ :

$$\sigma_E / E < \rho ; \tag{III.6}$$

c. Undulator length larger than the gain length:

$$N_U \lambda_U > L_G ; \tag{III.7}$$

d. Gain length shorter than the radiation Rayleigh range:

$$L_G / Z_R < 1 ; \tag{III.8}$$

where the Rayleigh range is defined using the minimum waist of a Gaussian radiation beam $\pi w_o^2 = \lambda_r Z_R$.

e. The quantum-recoil parameter:

$$q = \hbar\omega_r / mc^2\gamma\rho \ll 1. \tag{III.9}$$

Condition *a* says that for the instability to occur the electron beam must match the angular and transverse space characteristics of the radiation emitted by one electron in traversing the undulator. This is also the condition to obtain diffraction limited spontaneous radiation. Notice that for nanometer wavelength or shorter this condition cannot be met at present by storage ring based synchrotron radiation sources, but it can be satisfied by electron beams produced by a radio frequency, laser driven electron gun (photo-injectors) (Fraser and Sheffield, 1987), as we will discuss in the following sections. Condition *b* limits the beam energy spread to a value such that the width of the FEL radiation spectrum remains smaller than that of the spontaneous radiation line. Condition *c* introduces a requirement on the minimum undulator length for this process to become significant. Condition *d* requires that the radiation generated by the FEL be greater than that lost by diffraction. Conditions *a*, *b*, *d* are a function of beam emittance, electron beam focusing and radiation wavelength, and are not independent. If they are satisfied diffraction and 3-dimensional effects are not important, and we can use with good approximation the 1-dimensional FEL model. The case when diffraction is important will be discussed in Section V.

Condition *e* is similar to *b*, but in this case what matter is the energy change of one electron in a single photon emission. When this condition is violated quantum effects become important (Bosco, Colson and Freeman, 1983; Becker and Zubairy, 1982; Becker and McIver, 1983; Dattoli *et al.*, 1985; Bonifacio and Casagrande, 1985; Bonifacio 1997; Preiss *et al.*, 2012). These papers discuss the quantum theory in the limit of small recoil. The case of large recoil and large FEL gain is discussed in Schroeder, Pellegrini and Chen, 2001; Bonifacio, Piovella, Robb, 2005; Geloni, Kocharyan, Saldin, 2012. The quantum-recoil parameter of condition *e* can be interpreted as the ratio of the axial displacement due to the emission or absorption of a discrete photon to the radiation wavelength. In classical FEL theory $1/q = \rho mc^2\gamma / \hbar\omega_r$ is approximately the number of resonant photons emitted per electron at saturation. This is an important consideration because it shows that in the classical regime one electron can emit many photons, giving the FEL a large gain over spontaneous radiation. In the quantum case number of photons is about one, the efficiency remains low. This condition puts a lower limit on the electron energy to operate the FEL. Reducing the beam energy for a given X-ray photon energy by reducing the undulator period, leads to the quantum regime and low efficiency. The condition of quantum-recoil much smaller than one was satisfied in Madey's theory (Madey, 1971), giving for the gain a classical result. In the rest of this paper we will consider only the classical case of small quantum recoil.

When conditions *a* to *e* are satisfied the radiation field emitted by the electron beam grows exponentially along the undulator length, with a growth rate given by (III.4). At saturation the radiation power is given by

$$P_{sat} \approx \rho(I_{beam} / e)E_{beam}, \tag{III.10}$$

(Bonifacio, Pellegrini and Narducci, 1984), where I_{beam} and E_{beam} are the beam current and energy, and the number of photons per electron is

$$N_{ph} \approx \rho E_{beam} / E_{ph} \qquad\qquad (III.11)$$

In a case of interest, a Soft X-ray FEL with $E_{ph} = 250eV$, $E = 3GeV$, $\rho = 10^{-3}$, we obtain $N_{ph} \simeq 10^4$, i.e. an increase of almost 6 orders of magnitude in the number of coherent photons produced per electron. At higher beam energy and 8 keV photon energy the increase is five orders of magnitudes. This increase is reflected in a much larger brightness for FELs compared to synchrotron radiation sources, as shown in FIG. 12.

When the FEL starts from noise, the SASE case, the power saturates after an undulator length, called the saturation length, about ten electric field gain lengths, or, using (III.4), approximately

$$L_{sat} \approx \lambda_U / \rho . \qquad\qquad (III.12)$$

2. *FEL equations with universal scaling*

We now analyze the FEL equations to justify our physical picture. They are rather complicated and are a large number, $6N_e + 1$. To better understand their meaning and obtain results valid at any wavelength we rewrite them using different variables, scaling according to two characteristics length of the system, proportional to the gain and cooperation lengths in the 1-dimensional limit.

Due to the difference in velocity between the electrons and the radiation field, the field advances in front of the electrons by one wavelength per undulator period, as discussed in Section I-5. The quantity

$$S = N_U \lambda_r , \qquad\qquad (III.13)$$

with N_U the number of undulator periods, is the slippage, telling us how much one photon advances over the full undulator length with respect to the electron that generated it at the undulator entrance. To characterize different FEL regimes we use the ratio of the slippage to the electron bunch length L_B, and consider cases when this ratio is larger or smaller than one.

We write the electromagnetic field amplitude using the new variables

$$z_u = z,$$
$$z_b = \lambda_U (z - \beta_{z0}ct) / \lambda_r \beta_{z0}, \qquad\qquad (III.14)$$

where β_{z0} is the electron average velocity, corresponding to an average energy γ_0. The second variable gives the position of an electron along the bunch, normalized to the ratio of the undulator period to the radiation wavelength.

With this coordinate transformation, and using the approximation $d / cdt = \beta_{z0} \partial / \partial z_u \simeq \partial / \partial z_u$, the field and particle equations, (II.36) to (II.40), become

$$\frac{\partial}{\partial z_u} \vec{x}_{B,n} = \frac{\vec{p}_{B,n}}{\beta_{B,F}}, \qquad\qquad (III.15)$$

$$\frac{\partial}{\partial z_u} \vec{p}_{B,n} = -\frac{\vec{x}_{B,n}}{\beta_{B,F}}, \tag{III.16}$$

$$\frac{\partial \Phi_n}{\partial z_u} = k_U \left[1 - \frac{\gamma_R^2}{\gamma_n^2} - \frac{k_r}{2k_U \beta_{B,F}} (\frac{\vec{x}_{B,n}^2}{\beta_{B,F}} + \beta_{B,F} \vec{p}_{B,n}^2) \right], \tag{III.17}$$

$$\frac{\partial \gamma_n}{\partial z_u} = \frac{eK}{2\beta_{z0} mc^2 \gamma_n} (\alpha e^{i\Phi_n} + c.c.), \tag{III.18}$$

$$\left(\frac{\partial}{\partial z_u} + \frac{\partial}{\partial z_b} \right) \alpha - \nabla_T^2 \frac{i\alpha}{2k_r} = 2\pi e n_e(x,y,z_b) \left[K \left\langle \frac{e^{-i\Phi}}{\gamma} \right\rangle - \frac{i e \alpha}{mc^2 k_r} \left\langle \frac{1}{\gamma} \right\rangle \right]. \tag{III.19}$$

The dependence on z_b describes the shape of the wave packet, including the slippage. In deriving equation (III.19) we have used the relationship $(k_r / k_U)(1 - \beta_{zo}) / \beta_{z0} = 1$. In the field equation, the term in the electron beam current proportional to the electromagnetic field is quite small and will be neglected in the next sections.

To facilitate the analysis of the problem we rewrite the FEL equations scaling the longitudinal variables z_b, z_u as in Bonifacio *et al*. (1994), and the transverse variables to an effective beam radius, r_0, as in Krinsky and Yu (1987):

$$\bar{z} = z_u / L_{sc}, \ \bar{z}_1 = z_b / L_{sc}, \tag{III.20}$$

$$\bar{x} = x / r_0, \ \bar{y} = y / r_0, \tag{III.21}$$

where the scaling length,

$$L_{sc} = \lambda_U / 4\pi\rho, \tag{III.22}$$

is the same quantity given in (III.4) and is an approximate value of the gain length in the 1-dimensional case.

Introducing the cooperation length L_c, defined as the slippage in the length L_{sc} (Bonifacio *et al*., 1994)

$$L_c = \lambda_r / 4\pi\rho, \tag{III.23}$$

and using (III.14) (III.20), (III.22), we can rewrite the variable $\bar{z}_1$ as

$$\bar{z}_1 = \frac{z - \beta_{z0} ct}{\beta_{z0} L_C}, \tag{III.24}$$

the position along the bunch measured in units of the cooperation length.

We also make the assumption that the electron energy changes only by a small value respect to the average beam energy and resonant energy, γ_0, γ_R and introduce the relative energy deviation

$$\eta_n = (\gamma_n - \gamma_R) / \rho \gamma_R. \tag{III.25}$$

42

Condition *b* discussed before, equation (III.6), is satisfied if this quantity is smaller than one.

The field variable is scaled introducing the quantity

$$\hat{\alpha} = \frac{eKL_{sc}}{2\beta_{z0}mc^2\gamma_0^2\rho}\alpha .$$
(III.26)

We also approximate, whenever possible without having an effect on the physics of the system, the beam longitudinal velocity with the light velocity and expand to first order in η. We simplify the field equation neglecting the current term proportional to the field in (III.19). Using these notations and approximations the FEL equation can be written as (Bonifacio *et al.*,1988; Bonifacio *et al.*, 1994):

$$\frac{\partial \vec{x}}{\partial \bar{z}} = \frac{\vec{p}}{\bar{\beta}_{B,F}} ,$$
(III.27)

$$\frac{\partial \vec{p}}{\partial \bar{z}} = -\frac{\vec{x}}{\bar{\beta}_{B,F}} ,$$
(III.28)

$$\frac{\partial \Phi_n}{\partial \bar{z}} = \eta_n - \frac{k_r L_{sc}}{2\bar{\beta}_{B,F}}(\frac{\vec{x}_{B,n}^2}{\bar{\beta}_{B,F}} + \bar{\beta}_{B,F}\vec{p}_{B,n}^2) ,$$
(III.29)

$$\frac{\partial \eta_n}{\partial \bar{z}} = \hat{\alpha}e^{i\Phi_n} + c.c. ,$$
(III.30)

$$\left(\frac{\partial}{\partial \bar{z}} + \frac{\partial}{\partial \bar{z}_1}\right)\hat{\alpha} - \frac{i}{F_D}\nabla_T^2\hat{\alpha} = \left\langle e^{-i\Phi}(1-\rho\eta)\right\rangle ,$$
(III.31)

where the Fresnel diffraction coefficient is defined as

$$F_D = 2r_0^2 k_r / L_{sc}$$
(III.32)

and the betatron wave number is measured in scaled units $\bar{\beta}_{B,F} = \beta_{B,F} / L_{sc}$. In the current term of the field equation we have also assumed $\eta\rho \ll 1$ and approximated it to first order in this quantity.

The set of these $6N_e + 1$ nonlinear equations are our FEL model. As we have been discussing they contain many approximation and assumption, like for instance constant electron beam density, which allows us to consider the FEL parameter as a well-defined number. However also under these conditions they are a quite complicated set of 10^8-10^{10} equations, and cannot be solved analytically. In the two following sections we will study them in some particular cases to establish our understanding of the FEL physics. A more complete solution, removing the assumption and approximation made, require a numerical solution, as will be discussed in section VIII.

In the 1-dimensional limit and neglecting emittance effects the equations assume a very simple form, with no explicit dependence on the radiation wavelength. In this limit the results are valid for any FEL, and they can be compared to experiments at any wavelength to determine their validity, in the 1-dimensional case, for all wavelengths.

Equation Section 4 IV. 1-D theory, FEL small gain and collective instability

To further simplify our model and study the basic FEL physics we consider initially the 1-dimensional case: the electromagnetic field is represented by plane waves, the beam by charged planes with infinite transverse extension and betatron oscillations are neglected. Only the longitudinal variables remain in the problem and the transverse part of the wave equation does not appear, $\nabla_T^2 \hat{\alpha} = 0$. Considering the electron current, we assume that the transverse and longitudinal distributions can be separated. In the X-ray FEL case the FEL parameter is small, 10^{-3} or less, and we neglect also the term proportional to it on the right hand side of the field equation (III.31)

In this case the field and particle equations, (III.29), (III.30), (III.31), become

$$\frac{\partial \Phi_n}{\partial \bar{z}} = \eta_n, \tag{IV.1}$$

$$\frac{\partial \eta_n}{\partial \bar{z}} = \hat{\alpha} e^{i\Phi_n} + c.c., \tag{IV.2}$$

$$\left(\frac{\partial}{\partial \bar{z}} + \frac{\partial}{\partial \bar{z}_1} \right) \hat{\alpha} = \left\langle e^{-i\Phi} \right\rangle. \tag{IV.3}$$

In this section we first consider the FEL 1-dimensional equations in the simplest possible case when we assume that the radiation electric field remains nearly constant during the interaction with the electron beam, the small gain theory. Madey's FEL theory (Madey, 1971) was done for this case and was based on a quantum approach, evaluating the stimulated bremsstrahlung in the undulator magnetic field. The final gain formula is classical and does not contain Planck's constant. Colson (1977a, 1977b) and Hopf *et al.* (1977) developed a classical small gain theory and we will follow this model.

Later we consider the case when the electric field can change by a large amount and show that, under certain conditions, the electron-radiation interaction can lead to an exponential growth of the radiation power, the FEL collective instability or high gain theory.

1. Electron dynamics in a constant radiation electric field

The constant electric field approximation can be used if the change in the field amplitude in crossing the undulator is small. In this case Maxwell equations are not needed and the FEL description uses only the phase and energy equations, (IV.1), (IV.2), with a constant field. The equations can be obtained from the Hamiltonian

$$H = \frac{\eta^2}{2} + 2|\hat{\alpha}| \cos \Phi. \tag{IV.4}$$

The electron motion is like that of particles in a pendulum-like potential, often called the ponderomotive potential, formed by the combined undulator and electromagnetic fields. A phase space plot is shown in FIG. 15.

The quantity $\sqrt{2|\hat{\alpha}|}$ is the small oscillation frequency around the stable point $\Phi = \pi, \eta = 0$ and the small oscillation period is $T_{sync} = \pi\sqrt{2/|\hat{\alpha}|}$. For analogy with the theory of particle accelerators these oscillations are called synchrotron oscillations and the area within the separatrix is called the bucket.
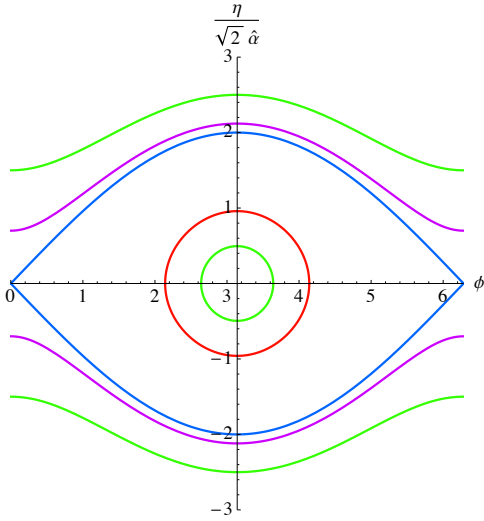
FIG. 15. Phase-space for the FEL pendulum equation, for particle with different initial conditions. The stable oscillation point is $\Phi = \pi, \eta = 0$. The two unstable equilibrium points are $\Phi = 0, 2\pi, \eta = 0$. The curve dividing the close and open trajectories is the separatrix. The maximum possible value of the normalized relative energy change is $\eta_{Max} = \pm 2\sqrt{2|\hat{\alpha}|}$.

The motion of the electrons inside the bucket can be evaluated from (IV.1), (IV.2) for given initial conditions. The results, starting from a monochromatic distribution in energy and a uniform distribution in phase are shown in FIG. 16. The four curves give the electron distribution at times corresponding to $0, T_{sync}/12, T_{sync}/8, T_{sync}/4$ for an initial condition $\eta_0 = 0.1$.

FIG. 16. Evolution along the undulator of the electron distribution in phase space, starting with a monoenergetic beam, $\eta_0 = 0.1$, uniform in phase,. The maximum bunching is obtained after a quarter of the synchrotron period.
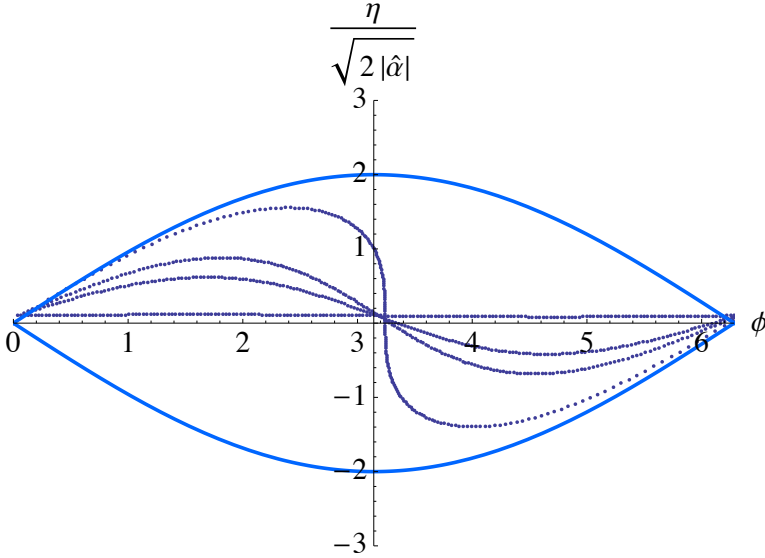
We can also evaluate the change in the average electron energy, the energy spread and the bunching factor as the electrons move along the undulator. Both the energy spread and the bunching factor grow to reach a maximum value for an undulator length corresponding to $T_{sync}/4$. The results are shown in FIG. 17. The electron change in energy would change sign for an initial condition $\eta_0 < 0$. For positive values of this quantity the decrease in the average electron energy corresponds to an increase of the radiation intensity, what we want to achieve in an FEL. For negative values of η_0 the radiation intensity would decrease and electron energy increase. This case is the Inverse FEL that can be used as a particle accelerator (Palmer, 1981; Courant, Pellegrini and Zakowicz, 1985; Van Steenbergen *et al.*, 1996; Musumeci *et al.* 2005).
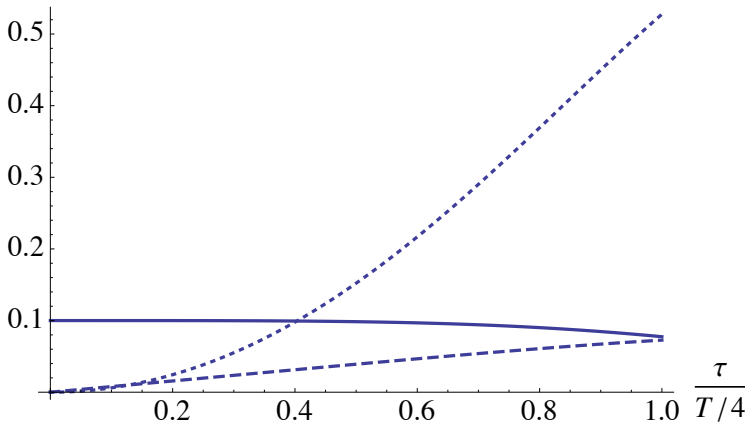

FIG. 17. Average energy change, solid line, beam energy spread, dashed line, and bunching factor, dotted line, along the undulator. We assume an undulator length corresponding to a quarter of the synchrotron oscillation period T .

46

The average energy change, the bunching and the energy spread can be evaluated solving Vlasov equation,

$$\frac{\partial f}{\partial \overline{z}} + \frac{\partial \eta}{\partial \overline{z}}\frac{\partial f}{\partial \eta} + \frac{\partial \Phi}{\partial \overline{z}}\frac{\partial f}{\partial \Phi} = 0 \qquad \text{(IV.5)}$$

for the Hamiltonian system (IV.4). We assume $|\hat{\alpha}| \ll 1$ and write the solution as a power series in this quantity, the square of the small oscillation frequency,

$$f(\Phi,\eta,\overline{z}) = f_0(\Phi,\eta) + |\hat{\alpha}| f_1(\Phi,\eta,\overline{z}) + |\hat{\alpha}|^2 f_2(\Phi,\eta,\overline{z}). \qquad \text{(IV.6)}$$

To lowest order we assume

$$f_0(\Phi,\eta) = \frac{1}{2\pi} g(\eta), \qquad \text{(IV.7)}$$

a distribution uniform in phase and monoenergetic. The first order solution is

$$f_1(\Phi,\eta,\overline{z}) = \frac{2}{\eta}\frac{\partial f_0}{\partial \eta}[\cos\Phi - \cos(\Phi - \eta\overline{z})]. \qquad \text{(IV.8)}$$

To this order the average electron energy change is zero, but the energy spread and bunching are not. To evaluate the second order solution we expand it in a Fourier series

$$f_2(\Phi,\eta,\overline{z}) = \sum_n f_{2n}(\eta,\overline{z})e^{in\Phi} \qquad \text{(IV.9)}$$

Only the term independent from the phase,

$$f_{20}(\eta,\overline{z}) = 2\frac{\partial}{\partial \eta}\left[\frac{1}{\eta^2}\frac{\partial f_0}{\partial \eta}(1 - \cos\eta\overline{z})\right] \qquad \text{(IV.10)}$$

is needed to evaluate the average energy change. The result for the change in the beam energy density is

$$\langle\eta\rangle - \langle\eta\rangle_0 = 4|\hat{\alpha}|^2 \overline{z}^3 \int d\eta\, g(\eta) F(\overline{z}\eta), \qquad \text{(IV.11)}$$

where the function $F(\overline{z}\eta)$, the small signal gain function, is

$$F(\overline{z}\eta) = \left[1 - \cos(\overline{z}\eta) - (\overline{z}\eta/2)\sin(\overline{z}\eta)\right]/\overline{z}^3\eta^3 \qquad \text{(IV.12)}$$

and is plotted in FIG. 18. At the undulator exit we have $\overline{z}\eta = 4\pi N_U(\gamma - \gamma_R)/\gamma_R$. Using (II.14) and (I.59), (I.60), the small signal gain function can be written as

47

$$F(\Delta) = -\frac{1}{8}\frac{d}{d\Delta}\left(\frac{\sin\Delta}{\Delta}\right)^2, \tag{IV.13}$$

where $\Delta = \eta\bar{z}/2 = \pi N_U(\omega - \omega_R)/\omega_R$. Comparing this result with the spontaneous radiation double differential spectrum, equation (I.59), we see that the small signal wave function is proportional to the derivative of the spontaneous radiation spectrum. It is also possible to establish a connection between the average energy change and the derivative of the energy spread $\langle\gamma - \gamma_0\rangle = (1/2)\partial\langle(\gamma - \gamma_0)^2\rangle/\partial\gamma_0$ where γ, γ_0 are the energy values at the undulator exit and entrance (Madey, 1979; Krinsky *et al.*, 1982). These two relations are known as Madey's theorems.

Notice that the average energy variation is proportional to the square of the radiation electric field. If we assume that the change in the beam electron energy is equal to the increase of the radiation energy we can use (IV.11) to evaluate the gain, the ratio of the change in radiation intensity to the initial intensity.

FIG. 18 shows that the gain is an anti-symmetric function of $\eta = (\gamma - \gamma_R)/\gamma_R$ and is zero for $\eta = 0$, $\gamma = \gamma_R$. The gain of an FEL operating with an initially mono-energetic beam and $\eta = \eta_0$, is zero if $\eta_0 = 0$, positive if $\eta > 0$ as shown in FIG. 17, and negative in the opposite case. The same figure shows also the increase in energy spread and bunching factor along the undulator.
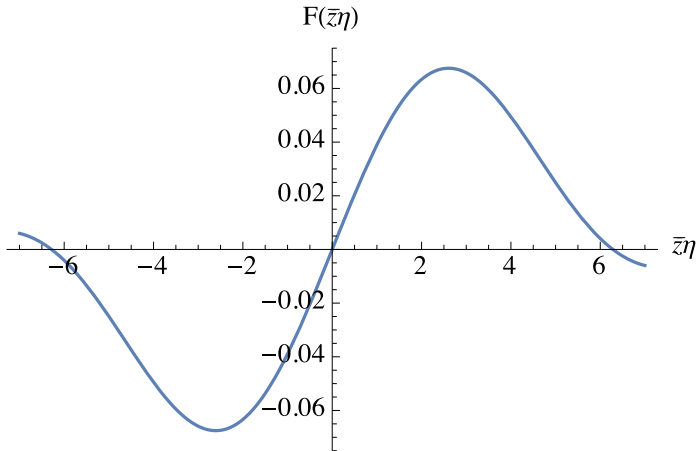
FIG. 18. The small signal gain function. The function is anti-symmetric. The width of the positive peak between zero and 2π is about $\Delta\gamma/\gamma \approx 1/2N_U$.

2. The FEL instability: long bunch case

We remove now the approximation that the electric field remains constant while the electron beam traverses the undulator. We notice that the FEL, described by the set of 6N+1 equations (IV.1), (IV.2), (IV.3), is in equilibrium if:

a. $\eta_n = 0$ for all electrons, the beam is monochromatic and its energy is the resonant energy;

b. The electrons are uniformly distributed in phase, $\langle \exp(-i\Phi) \rangle = 0$;

c. No field is present, $\hat{\alpha} = 0$.

Small deviations from equilibrium can be described in terms of collective variables: bunching factor $B = (1/N_e) \sum_{l=1}^{N_e} \exp(-i\Phi_l)$; energy modulation $P = (1/N_e) \sum_{l=1}^{N_e} \eta_l \exp(-i\Phi_l)$; field variable $\hat{\alpha}$. Using these variables we study the stability of the FEL near the equilibrium state, to determine if it is stable or unstable (Bonifacio *et al.*, 1984; Kim, 1986b). For our discussion we consider two cases: bunch length much larger than or of the order of the slippage distance.

We start by considering a long electron bunch, $L_B \gg N_U \lambda_r$, with a constant electric field along the bunch. Under this conditions the derivative of the field respect to the bunch coordinate can be neglected in (IV.3) and the equation of motion can be obtained from the Hamiltonian

$$H = \sum_{n=1}^{N_e} \left[\frac{\eta_n^2}{2} - i(\hat{\alpha} e^{i\Phi_n} - c.c.) \right] \tag{IV.14}$$

where the conjugate variables are Φ_n, η_n and $\sqrt{N_e}\hat{\alpha}, i\sqrt{N_e}\hat{\alpha}^*$.

We characterize the beam by a distribution function of the electron variables, $f(\Phi, \eta, \bar{z})$ and use Vlasov equation,

$$\frac{\partial f}{\partial \bar{z}} + \frac{\partial \eta}{\partial \bar{z}} \frac{\partial f}{\partial \eta} + \frac{\partial \Phi}{\partial \bar{z}} \frac{\partial f}{\partial \Phi} = 0, \tag{IV.15}$$

to study its evolution from the initial equilibrium state

$$f_0(\Phi, \eta) = g(\eta)/2\pi, \tag{IV.16}$$

an electron beam with uniform phase and energy distribution $g(\eta)$, to a final state .

We expand the distribution function near the equilibrium state as

$$f(\Phi, \eta, \bar{z}) = f_0(\Phi, \eta) + f_1(\eta) e^{i\Phi - i(\mu + \delta_\gamma)\bar{z}}, \tag{IV.17}$$

where

$$\delta_\gamma = \frac{\gamma_0^2 - \gamma_R^2}{2\rho\gamma_0^2} \tag{IV.18}$$

is the detuning, the relative difference between the average initial energy and the resonant energy defined by the radiation field frequency. In fact, using the relationship (II.14) between energy and wave number we can also write this equation as

$$\delta_\gamma = \frac{k_R - k_0}{2\rho k_R} . \tag{IV.19}$$

Scanning the system in frequency is equivalent to scanning in initial beam energy.

We use Vlasov equation to evaluate the first order term and, from it, the bunching factor and the field. We obtain

$$f_1(\eta)e^{i\Phi - i(\mu+\delta_\gamma)\bar{z}} = \frac{\hat{\alpha}e^{i\Phi} + c.c.}{2\pi i(\mu+\delta_\gamma - \eta)} \frac{\partial g(\eta)}{\partial \eta} \tag{IV.20}$$

and

$$\left\langle e^{-i\Phi} \right\rangle = i\int_{-\infty}^{\infty} d\eta \frac{\hat{\alpha}}{\eta - \mu - \delta_\gamma} \frac{\partial g(\eta)}{\partial \eta} . \tag{IV.21}$$

Substituting in the equation for the field, assumed to change as $\exp[-i(\mu+\delta_\gamma)\bar{z}]$, we obtain the dispersion relation

$$\mu + \delta_\gamma + \int_{-\infty}^{\infty} d\eta \frac{\partial g(\eta)}{\partial \eta} \frac{1}{\mu+\delta_\gamma - \eta} = 0 . \tag{IV.22}$$

In the simple case of a monochromatic beam with average value γ_0, $g(\eta) = \delta_f(\eta - \delta)$, the dispersion relation becomes

$$\mu^2(\mu+\delta_\gamma) - 1 = 0 . \tag{IV.23}$$

In the simple case $\delta_\gamma = 0$ the cubic equation has three solutions, one real two complex conjugates

$$\mu_{1,2} = (-1 \pm i\sqrt{3})/2, \ \mu_3 = 1 \tag{IV.24}$$

The first root has a positive imaginary part, generating an exponential growth of the radiation field, $\hat{\alpha} = \exp(\sqrt{3}\bar{z}/2) = \exp(\sqrt{3}z/2L_{sc})$. In the more general case the imaginary part is a function of the detuning as shown in FIG. 19. One can see that there is an exponential growth only if the detuning is less than 1.889.
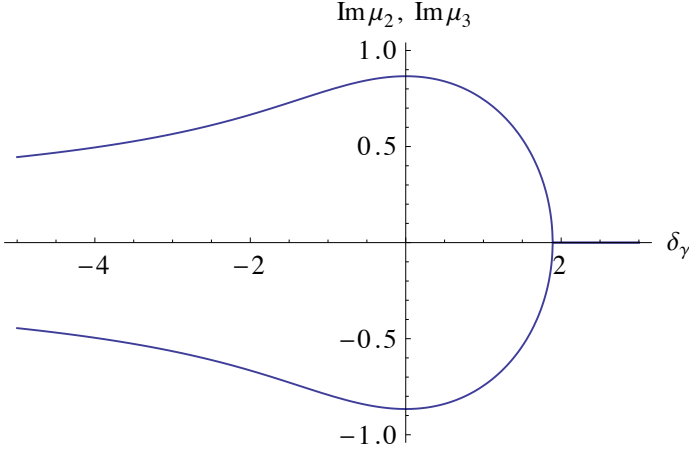
FIG. 19. The imaginary part of the two complex roots of the dispersion relation.

The evaluation of the growth rate from (IV.22) when the beam is not monochromatic in most cases must be done numerically, including the case of a Gaussian energy spread. We can obtain a simple analytic result if we assume a Lorentzian energy distribution

$$g(\eta) = \frac{\Delta}{\pi[(\eta - \delta_\gamma)^2 + \Delta^2]} \qquad (IV.25)$$

Substituting in (IV.22) and integrating we obtain the dispersion relation (Murphy and Pellegrini, 1990), for the case $\operatorname{Im}\mu > 0$,

$$(\mu + \delta_\gamma)(\Delta - i\mu)^2 + 1 + \Delta^2 = 0 \qquad (IV.26)$$
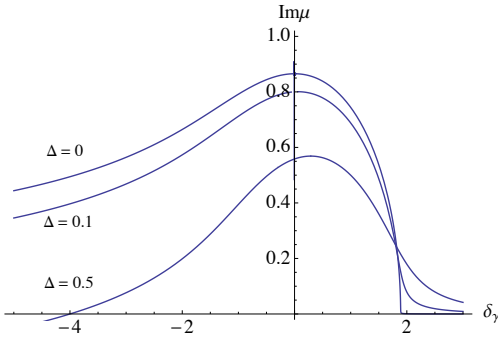
FIG. 20. Dependence of the imaginary part on the beam initial energy spread.

When the energy spread is zero (IV.26) reduces to the monoenergetic case (IV.23). A plot of the growth rate for various values of the energy spread is given in FIG. 20. When the value of the spread, measured in units of the FEL parameter ρ, becomes near one there is considerable growth rate and gain bandwidth reduction.

To a good approximation and for values of the energy spread of the order or smaller than the FEL parameter the dependence of the gain length on the energy spread can be approximated as (Xie, 1996)

51

$$L_G \simeq L_{G0}(1+(\sigma_\gamma / \rho)^2)$$ (IV.27)

3. The FEL instability, analysis in bunch frequency space

We can study again the FEL near the equilibrium point without the long bunch approximation. In this case the collective variables, bunching factor, energy modulation and field, can vary along the bunch, and are defined locally as a function of $\overline{z}$, $\overline{z}_1$. The bunching factor and energy deviation collective variables are

$$B = \frac{1}{\overline{N}_e} \sum_{l=1}^{\overline{N}_e} e^{-i\Phi_l}, \quad P = \frac{1}{\overline{N}_e} \sum_{l=1}^{\overline{N}_e} \eta_l e^{-i\Phi_l},$$ (IV.28)

where $\overline{N}_e$ is the number of electrons in one radiation wavelength.

We obtain a simple set of equations describing the FEL near the equilibrium point if we multiply the particle equations (IV.1), (IV.2), by $\exp(-i\Phi_l)$ and sum over all particles within one wavelength, $\overline{N}_e$. Neglecting second order terms the linearized FEL equations are then given by (Bonifacio *et al.*, 1994)

$$\left(\frac{\partial}{\partial \overline{z}} + \frac{\partial}{\partial \overline{z}_1} \right) \hat{\alpha} = -B,$$ (IV.29)

$$\frac{\partial}{\partial \overline{z}} B = -iP,$$ (IV.30)

$$\frac{\partial}{\partial \overline{z}} P = \hat{\alpha}.$$ (IV.31)

We assume the dependence of the three variables respect to the bunch coordinate $\overline{z}_1$ to be of the form

$$\hat{\alpha}(\overline{z}, \overline{z}_1) = \tilde{\alpha}(\overline{z}, \delta) e^{-i\delta \overline{z}_1},$$ (IV.32)

with similar equations for the bunching and energy deviation. The quantity δ is a measure the frequency deviation from the resonant frequency

$$\delta = (k_r - k)/2\rho k_r = L_c (k_r - k).$$ (IV.33)

Notice the different meaning of δ respect to the similar quantity δ_γ, defined in (IV.18), of the previous discussion.

The linearized FEL equations are now,

$$\left(\frac{\partial}{\partial \overline{z}} - i\delta\right)\tilde{\alpha}(\overline{z},\delta) = -\tilde{B}(\overline{z},\delta) \tag{IV.34}$$

$$\frac{\partial}{\partial \overline{z}}\tilde{B} = -i\tilde{P} \tag{IV.35}$$

$$\frac{\partial}{\partial \overline{z}}\tilde{P} = \tilde{\alpha} \tag{IV.36}$$

The functions $\tilde{\alpha}$, $\tilde{B}$, $\tilde{P}$ are functions of $\overline{z}, \delta$. We look for a solution of the form $\tilde{\alpha}, \tilde{B}, \tilde{P} \propto \exp(-i\mu\overline{z})$. The FEL equations are then satisfied if and only if

$$\mu^2(\mu + \delta) - 1 = 0. \tag{IV.37}$$

This is similar to (IV.23). As discussed before, in the simple case $\delta = 0$ the cubic equation has three solutions

$$\mu_{1,2} = (-1 \pm i\sqrt{3})/2, \; \mu_3 = 1. \tag{IV.38}$$

The imaginary part of μ as function of δ is the same as shown in FIG. 19.

It is important to consider that in the long bunch case, discussed in the previous section, we assumed a constant bunching and radiation field profile, with δ_γ proportional to the difference between the average beam energy and the resonant energy. In the case discussed now, short bunch case, the quantity δ gives the frequency spectrum corresponding to the bunching, energy deviation and field profile along the electron bunch.

The growth rate for the instability at zero detuning, when using the distance z along the undulator, is $L_G = 2L_{sc}/\sqrt{3}$. There is a small difference between the scaling length given in (III.22) and the one-dimensional gain length. Other differences will be due to effects like diffraction, emittance, energy spread and beam focusing, as we will discuss later.

The instability is typically triggered by a non-vanishing initial value of the electric field or of the collective variables B and P. As the exponential growth progresses, the linear approximation is eventually broken and the amplification reaches saturation. While the behavior of the FEL around saturation is strongly non-linear, we can estimate the order of magnitude of the radiation power as well as the energy modulation with a simple argument based on the linear equations.

Note that by definition $|B| < 1$. Saturation occurs when the bunching factor is close to its maximum value, i.e. when $|B| \approx 1$. From the energy deviation equation we have that, during the exponential growth, $|P| = \rho|B|$, which means that around saturation we have

$$|P| \approx \rho. \tag{IV.39}$$

Similarly, from the field equation, the radiation power density during the exponential amplification is given by $\rho n_e \gamma mc^2|B|^2$, the electron beam energy density times the square of the bunching factor. It follows that, close to saturation, the electromagnetic radiation power is

$$P_L \simeq \rho P_{beam},$$
(IV.40)

where $P_{beam} = mc^2 \gamma I_{beam} / e$ is the electron beam peak power, I_{beam} is the beam peak current. The FEL parameter gives the power extraction efficiency of the FEL amplifier, the efficiency of energy transfer from the electron beam to the radiation. These arguments will be confirmed later on with non-linear numerical simulations.

4. Initial Value Problem

We have seen that the bunching factor and the energy deviation are such that $0 < |B| < 1$, $0 < |P| < 1$. The radiation field, instead, scales with the beam power as $P_{rad} \leq \rho P_{beam}$. We solve the linear FEL system, (IV.34), (IV.35), (IV.36), with the Laplace transform method. We define the transform as

$$\tilde{a}_l(\mu, \delta) = \int_0^\infty \tilde{a}(\bar{z}, \delta) e^{i\mu\bar{z}} d\bar{z}.$$
(IV.41)

The linear FEL equations (IV.34), (IV.35), (IV.36) become

$$\begin{aligned}
-i(\mu + \delta)\tilde{\alpha}_l - \tilde{\alpha}_0 &= -\tilde{B}_l, \\
-i\mu\tilde{B}_l - \tilde{B}_0 &= -i\tilde{P}_l, \\
-i\mu\tilde{P}_l - \tilde{P}_0 &= \tilde{\alpha}_l,
\end{aligned}$$
(IV.42)

where the subscript 0 stands for the initial value of the variable. The field variable in the Laplace space is given by

$$\tilde{\alpha}_l = \frac{1}{D(\mu, \delta)} (i\mu^2 \tilde{\alpha}_0 + \mu \tilde{B}_0 + \tilde{P}_0),$$
(IV.43)

where

$$D(\mu, \delta) = \mu^2(\mu + \delta) - 1$$
(IV.44)

is the FEL dispersion function. The Laplace transform can be inverted using the residue theorem

$$\hat{\alpha}(\bar{z}) = \sum_{j=1}^3 \frac{-i}{dD/d\mu \big|_{\mu=\mu_j}} (i\mu^2 \tilde{\alpha}_0 + \mu \tilde{B}_0 + \tilde{P}_0) e^{-i\mu_j \bar{z}},$$
(IV.45)

54

where μ_j are the zeroes of the dispersion equation $D(\mu,\delta)=0$.

The FEL instability can be triggered by a non-zero value of any of the three variables describing the system. However the FEL typically starts by either an initial radiation field, as is the case of seeded FELs, or an initial micro bunching. In the first case we have an FEL amplifier. How to generate a signal to amplify at nanometer or sub-nanometer wavelengths will be discussed in Section VII on seeding. In the latter case, if the initial micro bunching is given by shot-noise, the system is termed self-amplified spontaneous emission (SASE).

The dispersion equation has an unstable root for $\delta<\delta_{th}\simeq1.9$. To understand the spectral properties of the FEL amplifier, it is useful to study the unstable root μ_1 around resonance, $\delta=0$. Taking the derivative of the dispersion equation we find

$$\frac{d\mu}{d\delta}=\frac{\mu}{3\mu+2\delta}, \qquad\qquad\text{(IV.46)}$$

$$\frac{d^2\mu}{d\delta^2}=\frac{d\mu/d\delta}{3\mu+2\delta}-\frac{\mu(3d\mu/d\delta-2)}{(3\mu+2\delta)^2},$$

which yields for the growth rate expansion around resonance

$$\mu_1=-\frac{1}{2}-\frac{\delta}{3}-\frac{\delta^2}{18}+i\frac{\sqrt{3}}{2}(1-\delta^2/9). \qquad\qquad\text{(IV.47)}$$

The power gain is proportional to $\exp\left(\sqrt{3}\bar{z}(1-\delta^2/9)\right)$, giving a power amplification bandwidth $\sigma_\delta=3/\sqrt{2\sqrt{3}\bar{z}}$. In physical units, the relative amplification bandwidth is $\sigma_\omega/\omega=6\rho/\sqrt{2\sqrt{3}\bar{z}}$. Typically an FEL starting from a very low initial field or bunching saturates in about 20 power gain lengths, giving a relative amplification bandwidth $\sigma_\omega/\omega\approx\rho$.

In the absence of an external seed the FEL instability can be triggered by the beam shot-noise in the longitudinal particle distribution. As discussed in Section I-9 for an ensemble of particles with a random longitudinal distribution, the intrinsic discreteness of the particle distribution gives a non-vanishing bunching factor. The statistical average of the shot-noise bunching factor is zero but the average power is given by $\left\langle\left|B\right|^2\right\rangle=1/\bar{N}_e$. In the high-gain approximation and starting from an initial bunching factor B_0, the normalized radiation field is given by

$$\tilde{\alpha}(\bar{z},\delta)=G(\bar{z},\delta)\tilde{B}_0, \qquad\qquad\text{(IV.48)}$$

where

$$G(\bar{z},\delta)=-ie^{i\mu\bar{z}}/(3\mu-\delta) \qquad\qquad\text{(IV.49)}$$

is the gain-function. Since the shot-noise microbunching, $\tilde{B}_0$, is a stochastic variable, the radiation field must be treated statistically. In the time domain, the field is given by the sum over the whole frequency spectrum of the spectral components from shot-noise

$$\hat{\alpha}(\overline{z},\overline{z}_1) = \sum \tilde{B}_0(\delta_n)G(\overline{z},\delta_n)e^{i\delta_n\overline{z}_1} \qquad (IV.50)$$

where $\delta_n = \pi n / k_r\rho L_B$ and L_B is the electron bunch length. For a long bunch we can approximate the summation with an integral and obtain

$$\hat{\alpha}(\overline{z},\overline{z}_1) = \frac{k_r L_B \rho}{\pi}\int \tilde{B}_0(\delta)G(\overline{z},\delta)e^{i\delta\overline{z}_1} . \qquad (IV.51)$$

The normalized radiation power is computed by the following integral in the frequency domain:

$$\left\langle \left|\hat{\alpha}(\overline{z},\overline{z}_1)\right|^2\right\rangle = \frac{k_r L_B \rho}{\pi N_e}\int_{-\infty}^{\infty}\left|G(\overline{z},\delta)\right|^2 d\delta . \qquad (IV.52)$$

A good approximation for the SASE power can be obtained by neglecting the dependence on δ of the residue $1/(3\mu - 2\delta)$. With this approximation the gain function has a Gaussian dependence on the detuning

$$\left|G(\overline{z},\delta)\right|^2 = \frac{1}{9}e^{\sqrt{3\overline{z}}(1-\delta^2/9)} \qquad (IV.53)$$

and the resulting shot-noise power is

$$\left\langle \left|\hat{\alpha}(\overline{z},\overline{z}_1)\right|^2\right\rangle = \frac{2\rho}{3\overline{N}_e}\sqrt{\frac{\pi}{\sqrt{3\overline{z}}}}e^{\sqrt{3\overline{z}}} . \qquad (IV.54)$$

In physical units, the SASE power is given by

$$P_{SASE} = P_{sn}e^{2k_U\rho\sqrt{3\overline{z}}} , \qquad (IV.55)$$

where

$$P_{sn} = P_{beam}\frac{2\rho^2}{3\overline{N}_e}\sqrt{\frac{\pi}{\sqrt{3\overline{z}}}} \qquad (IV.56)$$

is the equivalent shot-noise power.

In the SASE mode, due to the intrinsic randomness of the shot-noise microbunching, the radiation field is composed of several uncorrelated spikes (Bonifacio *et al.*, 1994). The temporal profile of the radiation pulse starting from shot-noise, obtained by performing an inverse Fourier transform of (IV.45) and using shot-noise as the initial trigger is shown in FIG. 21. The temporal structure of the radiation spikes can only be described in the framework of statistical optics. In particular, we are interested in estimating the typical temporal width of a SASE spike. To do so, we compute the autocorrelation function of the radiation field, defined as

$$Q(\overline{z}, \overline{z}_1, \overline{z}_1') = \left\langle \hat{\alpha}(\overline{z}, \overline{z}_1) \hat{\alpha}^*(\overline{z}, \overline{z}_1 + \overline{z}_1') \right\rangle \qquad \text{(IV.57)}$$
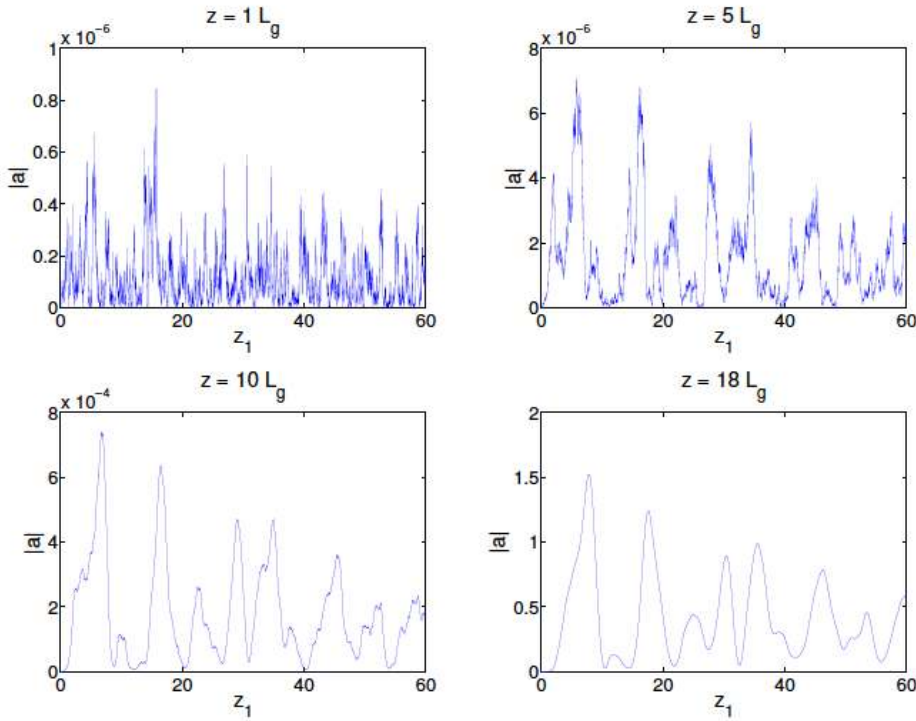

FIG. 21 Temporal profiles of a SASE pulse for different values of $\overline{z}$. As the FEL pulse propagates on the electrons, a longitudinal correlation is developed and the width of the temporal spikes grows.

The physical meaning of the autocorrelation function can be understood as follows: the radiation spikes have no phase relation, i.e., the product of the field of one spike times the field of an independent spike averages to zero. On the other hand, the phase within one spike is well defined, which means that the product $\hat{\alpha}(\overline{z}, \overline{z}_1) \hat{\alpha}^*(\overline{z}, \overline{z}_1 + \overline{z}_1')$ averages to a number larger than zero if $\overline{z}$ is smaller than the width of a spike. In practice, the autocorrelation function yields the characteristic shape of a SASE spike. Using the same line of reasoning as in the average power calculation, we have

$$Q(\bar{z},\bar{z}_1,\bar{z}_1') = \left(\frac{k_r L_B \rho}{\pi N_e}\right)^2 \int \left|G(\bar{z},\delta')\right|^2 e^{-i\delta\bar{z}_1'} d\delta , \qquad \text{(IV.58)}$$

which can be reduced to a Gaussian integral by neglecting the δ dependence of the residue. The frequency integral yields

$$Q(\bar{z},\bar{z}_1,\bar{z}_1') = \left\langle \left|\tilde{\alpha}(\bar{z},\bar{z}_1)\right|^2 \right\rangle e^{-\bar{z}_1'^2 / 2\sigma_{\bar{z}',c}^2} , \qquad \text{(IV.59)}$$

where the normalized rms correlation length is

$$\sigma_{\bar{z}_1,c} = \sqrt{2\sqrt{3}\bar{z}} / 3 . \qquad \text{(IV.60)}$$

There are two facts worth noticing in (IV.60). First of all, the correlation length grows as the square root of $\bar{z}$, the normalized undulator distance. This is because the physical mechanism that introduces temporal correlation along the radiation pulse is propagation: the electrons are totally uncorrelated in phase at the beginning of the FEL interaction, but as the amplification process goes on, the radiation slippage transports the phase information in the forward direction. This introduces a phase correlation between regions of the electron bunch that are within the slippage length. In physical units, the autocorrelation function scale length is the cooperation length, $L_c = \lambda / 4\pi\rho$, the slippage length accumulated in a gain-length, defining the typical temporal scale of SASE. For the typical saturation length of an X-ray SASE FEL, $\bar{z}_{sat} \simeq 20/\sqrt{3}$, we have $\sigma_{\zeta,c} = \sigma_{\tau,c}/k_r = 2L_c$, which means that close to saturation the typical full width of a spike is roughly $6L_c$.

For a SASE FEL, starting from noise in the initial phase distribution, the total intensity, and that of each spike, fluctuate. The probability distribution of the FEL pulse energy, E, is given by a Gamma distribution function (Saldin, Schneidmiller and Yurkov, 1998)

$$p(E) = \frac{M^M}{\Gamma(M)} \left(\frac{E}{\langle E \rangle}\right)^{M-1} \frac{1}{\langle E \rangle} \exp\left(-M\frac{E}{\langle E \rangle}\right), \qquad \text{(IV.61)}$$

where $\Gamma(M)$ is the Gamma-function, $\langle E \rangle$ is the pulse energy average value and M is the number of degrees of freedom, or modes in the radiation pulse. For a transversely coherent pulse, in a single Gaussian mode, the number of degrees of freedom is reduced to that of the longitudinal modes and is given by the number of spikes in the pulse. For $M = 1$, single spike, the probability distribution is a negative exponential and for large values of M it tends to a Gaussian.

The first measurements of the fluctuations (Hogan *et al.*, 1998a; Hogan *et al.*, 1998b) where in good agreement with theory and were also used to find the number of spikes and from this the X-ray pulse length. More recently the same method has been used to measure few femtosecond long pulses at LCLS (Wu *et al.* 2010).

5. Time dependent FEL theory

In this section, we discuss the one-dimensional time-dependent theory of the FEL instability. We assume an arbitrary dependence of the radiation field and study the evolution of the FEL from an initial arbitrary signal directly in the time domain with a Green's function approach (Bonifacio *et al.*, 1994).

The problem we want to solve is the following: given the initial temporal distribution of the radiation field $\bar{a}_0(\bar{z}_1) = \bar{a}(0, \bar{z}_1)$, how does the radiation field evolve as a function of $\bar{z}, \bar{z}_1$? It is convenient to solve the FEL equations by taking a Laplace transform in $\bar{z}$

$$\bar{\alpha}(\mu, \bar{z}_1) = \int \hat{\alpha}(\bar{z}, \bar{z}_1) e^{i\mu\bar{z}} d\bar{z} .\tag{IV.62}$$

We assume that the initial bunching factor and energy modulation vanish, $B(0, \bar{z}_1) = 0$, $P(0, \bar{z}_1) = 0$. The resulting self-consistent equation in the Laplace domain is

$$\frac{\partial \bar{\alpha}}{\partial \bar{z}_1} - i(\mu - 1/\mu^2)\bar{\alpha} = \bar{\alpha}_0(\bar{z}_1) .\tag{IV.63}$$

This equation can be solved writing it in the equivalent form

$$\frac{\partial}{\partial \bar{z}_1}\left[\bar{\alpha} e^{-i(\mu - 1/\mu^2)\bar{z}_1} \right] = \bar{\alpha}_0(\bar{z}_1) e^{-i(\mu - 1/\mu^2)\bar{z}_1} ,\tag{IV.64}$$

with the solution

$$\bar{\alpha}(\mu, \bar{z}_1) = \int_{-\infty}^{\bar{z}_1} \hat{\alpha}_0(\bar{z}_1') e^{-i(\mu - 1/\mu^2)(\bar{z}_1 - \bar{z}_1')} d\bar{z}_1' .\tag{IV.65}$$

As a check, we assume an initial periodic distribution $\bar{\alpha}_0 = \alpha_0 \exp(i\delta\bar{z}_1)$. The integral can be solved and the Laplace transform of the field is

$$\bar{\alpha}(\mu, \bar{z}_1) = \frac{i\alpha_0}{\mu - 1/\mu^2 + \delta} e^{-i\delta\bar{z}_1} = \frac{i\alpha_0\mu^2}{D(\mu, \delta)} e^{-i\delta\bar{z}_1} \tag{IV.66}$$

where $D(\mu, \delta) = \mu^3 + \delta\mu^2 - 1$ is the FEL dispersion function. Note that, up to the oscillatory term $\exp(-i\delta\bar{z}_1)$ the same solution was obtained before (see (IV.43)) and the inverse Laplace transform of (IV.66) gives the usual three FEL modes: an unstable mode, a decaying mode and an oscillatory mode.

The general solution for the time-dependent 1-D FEL theory is given by the inverse Laplace transform of (IV.66). With the chosen definition for the Laplace variable, the inverse transform is given by

$$\hat{\alpha}(\bar{z}, \bar{z}_1) = \frac{1}{2\pi} \int_{-\infty + i\Gamma}^{\infty + i\Gamma} \bar{\alpha}(\mu, \bar{z}_1) e^{-i\mu\bar{z}} d\mu ,\tag{IV.67}$$

59

where Γ is a real number smaller than the imaginary part of any pole of $\bar{\alpha}$. The last equations represent the most general description for the 1-D FEL dynamics. To develop some intuition on the physics at play we derive the Green's function solution to the time-dependent equations. The Laplace transform of the Green's function, for $\bar{a}_0(\bar{z}_1) = \delta_f(\bar{z}_1)$, is given by

$$\tilde{G} = 0 \text{ for } \bar{z}_1 < 0,$$
$$\tilde{G} = e^{i(\mu - 1/\mu^2)\bar{z}_1} \text{ for } \bar{z}_1 > 0. \tag{IV.68}$$

We also note that for an initial pulse localized at $\bar{z}_1 = 0$, the region $\bar{z}_1 > \bar{z}$ falls outside of the radiation slippage length. Since the FEL pulse cannot propagate faster than the speed of light, the Green's function must be localized within the slippage length, i.e.

$$\tilde{G} = 0 \text{ for } \bar{z}_1 > \bar{z}. \tag{IV.69}$$

In the region $0 < \bar{z}_1 < \bar{z}$ the Green's function is given by

$$\tilde{G}(\bar{z}, \bar{z}_1) = \frac{1}{2\pi} \int_{-\infty+i\Gamma}^{\infty+i\Gamma} e^{-i\mu(\bar{z}-\bar{z}_1)-i\bar{z}_1/\mu^2} \, d\mu. \tag{IV.70}$$

The integral in the last equation can be solved exactly in terms of a power series, using the residue theorem. However, it is more instructive to derive a closed form expression for the Green's function using the stationary phase approximation. For large values of $\bar{z} - \bar{z}_1, \bar{z}_1$ the exponential in the integral has a strongly oscillatory behavior except for regions close to the extrema of its argument. We can then approximate the exponential as

$$e^{-ig(\mu)} \simeq e^{-ig(\mu_0)-ig''(\mu_0)(\mu-\mu_0)^2/2} \tag{IV.71}$$

which can be integrated as

$$\int_{-\infty}^{\infty} e^{-ig(\mu)} \, d\mu = \int_{-\infty}^{\infty} e^{-ig(\mu_0)-ig''(\mu_0)(\mu-\mu_0)^2/2} \, d\mu = e^{-ig(\mu_0)} \sqrt{2\pi/ig''(\mu_0)}. \tag{IV.72}$$

The stationary points for the integrand in (IV.72) are given by

$$\bar{z} - \bar{z}_1 - 2\bar{z}_1/\mu^3 = 0, \tag{IV.73}$$

with three solutions,

$$\mu_{st,n} = \mu_n \left(2\bar{z}_1/(\bar{z}-\bar{z}_1)\right)^{1/3}, \tag{IV.74}$$

where the μ_n are the three roots of the dispersion relation on resonance, $\mu^3 - 1 = 0$. In the high-gain regime we only keep the term associated to the unstable root $\mu_1 = -(1-i\sqrt{3})/2$. The asymptotic expansion of the Green's function is then given by

60

$$G(\bar{z},\bar{z}_1) = \sqrt{-12\pi i \mu_1} \frac{2^{2/3}\sqrt{\bar{z}_1}}{(\sqrt{\bar{z}_1}(\bar{z}-\bar{z}_1))^{2/3}} e^{-3i\mu_1(\sqrt{\bar{z}_1}(\bar{z}-\bar{z}_1)/2)^{2/3}} . \quad \text{(IV.75)}$$

The maximum of the Green's function as a function of $\bar{z}_1$ is realized for the maximum of $y = \sqrt{\bar{z}_1}(\bar{z}-\bar{z}_1)/2$, i.e. $\bar{z}_1 = \bar{z}/3$. The peak of the Green's function grows as $G(\bar{z},\bar{z}/3) \propto \exp(-i\mu_1\bar{z})$, meaning that, close to its maximum, the Green's function has the same growth rate as a resonant pulse. Furthermore, the condition $\bar{z}_1 = \bar{z}/3$ can be rewritten as $(z-V_{beam}t)/L_c = z/3L_G$, or

$$z/t = V_{group} - V_{beam}/(1-\lambda_r/3\lambda_U) \simeq V_{beam}(1+\lambda_r/3\lambda_U). \quad \text{(IV.76)}$$

The Green's function travels at a group velocity smaller than the speed of light but larger than the beam velocity. More quantitatively, the FEL pulse travels faster than the electrons by one third of the slippage rate, $V_{group} - V_{beam} = c\lambda_r/3\lambda_U$, and slower than the speed of light by two thirds of the slippage rate, $c - V_{group} = 2c\lambda_r/3\lambda_U$.

Performing a second order expansion of the exponent around its maximum, $d\bar{z}_1 = \bar{z}_1 - \bar{z}/3 \ll 1$, we obtain

$$\sqrt{\bar{z}_1}(\bar{z}-\bar{z}_1)/2 = (\bar{z}/3)^{3/2} - \frac{3^{3/2}}{8\sqrt{\bar{z}}}d\bar{z}_1 . \quad \text{(IV.77)}$$

Close to the peak, the Green's function can be approximated with a Gaussian shape

$$|G| \propto e^{-3^3 d\bar{z}_1^2/2^4\bar{z}} , \quad \text{(IV.78)}$$

with an RMS width $\sigma_{\bar{z}_1} = (2/3)^{3/2}\sqrt{\bar{z}}$. Note that the temporal width of the Green function grows like $\sqrt{\bar{z}}$.

This can be intuitively understood considering the effect of slippage: the initial distribution of the radiation field slips over the electron bunch at the rate of one wavelength per undulator period. As the radiation slips ahead of the electrons, the electrons develop a micro-bunching structure that is longer than the initial field distribution. Since microbunched electrons emit the FEL radiation, the initial short pulse (modeled here as a delta function) is lengthened to a few cooperation lengths. The same behavior was found for the correlation length in the steady-state SASE theory. This is not surprising considering that the SASE signal is given by the convolution of the Green's function with the initial shot-noise. Since the initial noise is totally uncorrelated, the SASE correlation length is essentially the correlation length of the Green's function itself. In the Gaussian approximation, this is given by $\sqrt{2}\sigma_{\bar{z}_1} \propto \sqrt{\bar{z}}$, in agreement with the results obtained in the previous section.
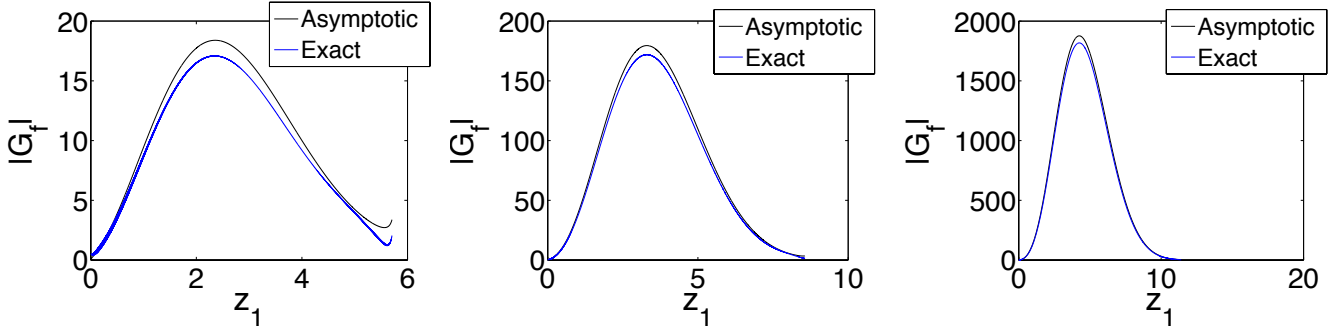
FIG. 22. Green's function for $\bar{z} = 10/\sqrt{3}$, left, $\bar{z} = 15/\sqrt{3}$, center, $\bar{z} = 20/\sqrt{3}$, right. Notice the difference in the vertical scale.

6. Non linear FEL dynamics and saturation

In the previous sections we have discussed the one-dimensional theory of the FEL in the linear regime. In this regime, the radiation field grows exponentially starting from an initial perturbation in the electron distribution or from an initial value of the radiation field itself. As discussed in sections IV-1 and IV-2 , the linear FEL equations are valid when the collective variables and the normalized field amplitude are much smaller than unity, $|B| << 1$, $|\hat{a}| << 1$, $|P| << 1$. As the FEL instability progresses the system reaches a point where the linear approximation is no longer valid and the amplification stops. This is the saturation effect, and it happens when the bunching factor approaches unity. In this section we discuss the one-dimensional FEL dynamics in the non-linear regime and the physics of saturation. Since the non-linear FEL equations cannot be solved analytically we rely on numerical integrations. However, since the FEL equations are written in a universally scaled form, the results that we find are general and apply to any FEL in the one-dimensional limit.

The non-linear electromagnetic field equation, (IV.3), is the same as in the linear case, since no assumption on the bunching factor has been made to derive it. To this we add the phase and energy equations for all electrons, (IV.1), (IV.2), now coupled through the electromagnetic field. For simplicity we assume a monochromatic FEL pulse, and discuss non-linear effects for a fixed value of δ, as defined in (IV.33).

This is a system of $2N_e + 1$ equations, where the number of electrons is usually as large as 10^8-10^{10}. Several numerical codes have been developed to integrate this system of equation, including also transverse electron density distribution and diffraction effects. A detailed discussion is presented in a later section of this paper. Here we use a Runge-Kutta integration method to obtain information about the FEL dynamics near and after saturation. FIG. 23 shows the growth along the undulator of the normalized radiation field, in units of interaction time, starting from an initial bunching factor of 0.001. The radiation grows exponentially up to the saturation point, which is at about $\bar{z} \simeq 10$. The saturation value is about 1.2, in good agreement with our previous estimates of the saturation power based on the linear regime, $P_{sat} \simeq \rho P_{beam}$.
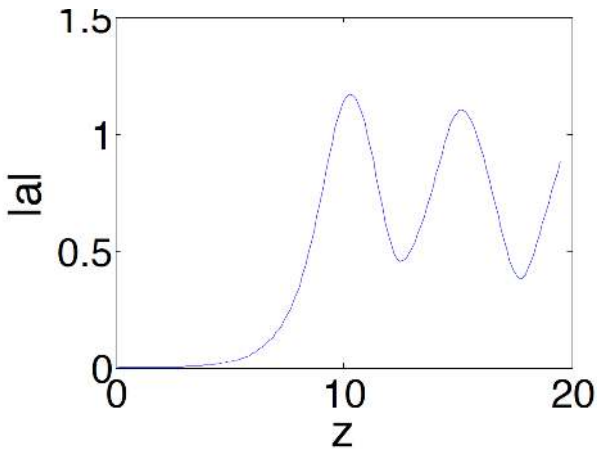
FIG. 23. Normalized radiation field amplitude as a function of interaction time, starting from an initial bunching factor of 0.001.

After the saturation point, the power oscillates in time. FIG. 24 shows the longitudinal phase-space over one radiation wavelength for the same conditions for different values of time. Before saturation, at $\bar{z} = 4$, $\bar{z} = 7$, the behavior of the phase-space distribution can be understood in terms of the linear FEL dynamics: the beam develops an energy modulation, which, in turn, generates micro-bunching. The amplitude of the energy modulation grows exponentially in time as a result of the FEL instability. Close to saturation and after saturation, $\bar{z} = 10$, $\bar{z} = 12$, the longitudinal phase space folds on itself and the micro-bunching stops growing (a process known as wave-breaking in the context of non-linear plasma waves). The oscillating behavior of the FEL after saturation can be understood in terms of the longitudinal oscillations of the electrons in the potential well generated by the radiation and undulator field. The motion of the electrons, for a given value of the radiation field, follows the pendulum equation. If we neglect the time-variation of the FEL field, the electrons oscillate longitudinally with a period $\bar{z} \simeq 2\pi\sqrt{a}$. Conversely, the radiation field follows the electron micro-bunching periodic oscillations of the, which, in turn is caused by the non-linear oscillation of the electrons themselves. Since after saturation the normalized radiation field is always of order of one, we can roughly estimate the periodicity of the non-linear oscillations to be on the order of $\bar{z} \simeq 2\pi$.

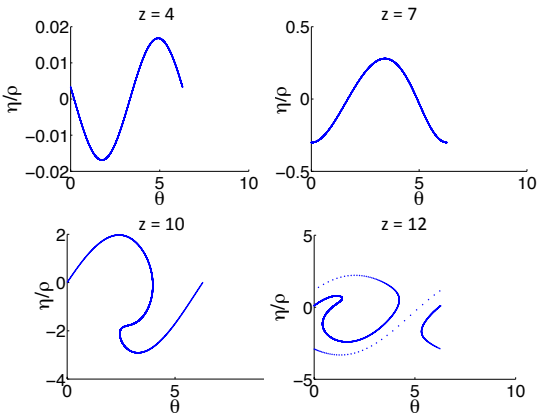FIG. 24. Longitudinal phase-space trajectories for different values of the interaction time $\bar{z}$ starting from $a_0 = 0.001$.

Finally, we study the behavior of the saturation power as a function of detuning. Figure 25 shows the normalized radiation field at saturation as a function of δ. The saturation power is higher for negative values of the frequency detuning and has a decreasing dependence on δ. This

has a simple intuitive explanation: the FEL saturates when the electron bunch energy falls out of resonance. For negative frequency detuning, the electron energy is above the resonant energy $\gamma_R = [(1+K^2)\lambda_U / 2\lambda_r]^{1/2}$, which means that the electrons can transfer more energy to the radiation field before they fall out of the amplification bandwidth.
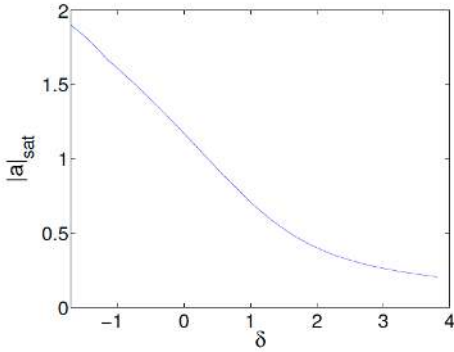

Figure 25 Normalized radiation field amplitude at saturation as a function of detuning

7. Harmonic Gain and non-linear Harmonic Generation

Since, for a planar undulator, spontaneous radiation is emitted on all the odd higher order harmonics, it means that coupling between the energy of the electrons and the radiation field happens on these harmonics too.

The same physical principles used to describe exponential growth on the first harmonic apply to the higher order harmonics, where the electron-radiation interaction generates an energy modulation turned into microbunching by the undulator dispersion and, finally, leads to the emission of more coherent radiation. The theory of the collective instability of harmonics (Murphy, Pellegrini and Bonifacio, 1985) follows that already discussed for the fundamental. The FEL equations for the n-th harmonic, in normalized units, are:

$$\frac{d\eta}{d\overline{z}} = F_n(K)\sin\Phi_n, \tag{IV.79}$$

where $\Phi_n = n\Phi$ is the ponderomotive phase at the n-th harmonic, described by the equation

$$\frac{d}{d\overline{z}}\Phi_n = n\eta \tag{IV.80}$$

and the harmonic coupling coefficient is given by (I.68).
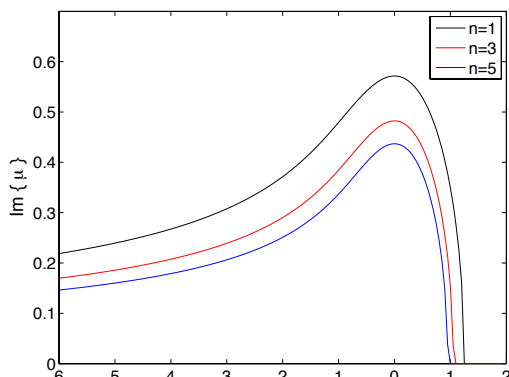
The field equation is similarly normalized as

FIG. 26. Normalized growth rate as a function of detuning for the first, third and fifth harmonics for a planar undulator with $K = 3$.

$$\left(\frac{\partial}{\partial \overline{z}} - i\delta_n\right)\tilde{\alpha}_n(\overline{z}, \delta) = -F_n(K)\tilde{B}_n(\overline{z}, \delta), \qquad \text{(IV.81)}$$

where $\tilde{B}_n = \langle \exp(-i\Phi_n)\rangle$ is the n-th harmonic bunching factor.

The dispersion equation for the n-th harmonic can be derived from these equations following the same method used for the fundamental harmonic. The growth rate is then given by

$$\mu^2(\mu + \delta_n) - \frac{n}{2}F_n^2(K) = 0, \qquad \text{(IV.82)}$$

where $\delta_n = (k - nk_r)/2\rho k_r$ is the frequency detuning from the n-th harmonic.

Note that for $n = 1$, this equation implies that, by re-defining the FEL parameter as $\overline{\rho} = \rho(F_1(K)/2)^{2/3}$, the one-dimensional theory developed so far applies to a planar undulator.

FIG. 26 shows the detuning curve for the first, third and fifth harmonic for a planar undulator with a parameter $K = 3$ while FIG. 27 shows the maximum growth rate as a function of the undulator parameter for the same harmonics. The growth rate can be quite large even for high harmonics. During the exponential growth process, however, the first harmonic dominates the radiation field since it has the strongest gain. However, harmonic lasing can be achieved by suppressing the first harmonic with the use of phase shifters along the undulator line. By inducing a phase difference between electrons and radiation of 2π/n or 4π/n, the lasing process on the n-th harmonic is not affected by the phase shift and can dominate over the first harmonic (McNeil *et al.* 2006; Schneidmiller and Yurkov 2012). The advantage of this concept is to reduce the electron beam energy needed for the given wavelength respect to the operation at the fundamental.
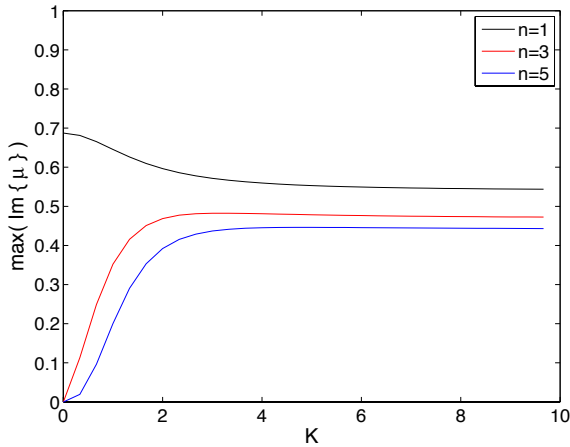
FIG. 27. Maximum normalized growth rate for the first, third and fifth harmonics as a function of the undulator parameter.

Harmonic lasing is interesting for a number of applications. For example, one could drive gain on the third harmonic to increase the energy range of an x-ray free-electron laser (Schneidmiller and Yurkov, 2012). Furthermore, since the slippage length is still determined by the first harmonic, lasing on the n-th harmonic increases the coherence length by a factor n (Xiang, Ding,

Huang and Deng, 2013), thus improving the spectral brightness of the radiation.

In addition to linear harmonic gain, harmonic radiation is generated in a high-gain FEL through a non-linear process close to saturation (De Martini, 1990). The process is similar to that of harmonic generation in a medium, due to the nonlinear dependence of the polarization on the fields driving the emission process and generating a frequency multiplication (Bloembergen, 1965). In this case the nonlinear medium is the electron beam. The bunching process on the fundamental in the nonlinear FEL potential well creates a density distribution composed of microbunches much narrower than the fundamental wavelength, as one can see for instance in FIG. 24, which can be Fourier analyzed as the sum of a large number of harmonics. In a planar undulator, the harmonic microbunching content leads to the emission of coherent radiation on all odd harmonics of the resonant wavelength on axis. The theory of non-linear harmonic generation was developed in (Bonifacio *et al*. 1990; Huang and Kim, 2000) and the first measurements of non-linear harmonic radiation in an FEL were reported in (Biedron *et al*., 2001; Tremaine *et al*., 2002). During the exponential growth regime, it can be demonstrated that the harmonic power growth, driven by the gain on the fundamental frequency, of the n-th harmonic power scales proportionally to P_1^n, where P_1 is the radiation power on the fundamental frequency. Many experiments on a wide range of photon energies (ranging from infrared to x-rays) have reported a saturation level for the third harmonic non-linear radiation on the order of 1% of the fundamental wavelength saturation power.

The second and third harmonics have been measured at LCLS (Ratner *et al.,* 2011). The results depend on the beam quality, peak current, emittance and phase space distribution of the electron beam. In the best conditions, in the soft X-ray region, near 1 keV, the second harmonic is as large as 0.1% of the fundamental and the third harmonic 2.5%. At hard X-rays, 6-8 keV, the third harmonic is 2%.

Equation Section 5V. Three-dimensional FEL theory

The core process of Free-electron lasers occurs in the longitudinal direction of the electron beam motion, interacting with a co-propagating field. This remains valid even in the complete three-dimensional description of the problem (Moore, 1985; Scharlemann, Sessler, and Wurtele, 1985; Sprangle, Ting and Tang, 1987). For the electron model the betatron motion couples to the longitudinal velocity, slowing down the longitudinal motion for larger betatron amplitudes. This and other effects will be discussed later in this section.

A large impact arises from the finite beam size of the electron beam and the localized emission of the radiation field. In a very simple model the electron beam can be regarded as an aperture and the emitted field diffracts depending on the size of the electron bunch. However, unlike an aperture, which has sharp edges, the diffraction pattern depends on the transverse electron distribution. A Gaussian distribution exhibits the least amount of diffraction and it is fruitful to compare the transverse equivalent to the "time-bandwidth" limit. For a Gaussian beam the product of rms beam size and rms divergence angle is given by $\lambda / 4\pi$ (Siegman, 1986). The electron beam equivalent is the geometric emittance (I.39). Note that the geometric emittance is inversely proportional to the beam energy while the resonant wavelength drops as $1/\gamma^2$. For a given normalized emittance, which is for linear beam optics a constant of motion, there is a wavelength, where the emittance becomes larger than that of a diffraction-limited beam. At this point the whole electron beam cannot radiate any longer into the fundamental mode and the overall efficiency is reduced. It is not a hard limit though and it depends on many factors, including the electron beam transverse distribution and the FEL interaction itself. In the following a 3D model is derived under

66

some assumption, which allows for analytical calculation of the FEL process.

1. Three-dimensional FEL equations neglecting emittance effects

We assume for simplicity a cylindrically symmetric electron beam and assume that the beam emittance is very small, neglecting the corresponding term in the phase equation. In contrast to the 1-dimensional theory the 3-dimensional theory has one additional independent parameter, the beam transverse size r_0. We assume it to be the effective source size for the radiation, like the waist of a Gaussian beam, and define a Rayleigh length as

$$Z_R = k_r r_0^2 / 2 . \tag{V.1}$$

The field equation (III.31) is simplified neglecting the dependence on the bunch coordinate, assuming a uniform longitudinal field profile, and operating at the resonant frequency, as we did in section IV-2 for the 1-dimensional case. This approximation is valid for a long bunch, as discussed before. With these approximation we have

$$\left(\frac{\partial}{\partial \bar{z}} - \frac{i}{F_d} \nabla_T^2 \right) \hat{\alpha} = \left\langle e^{-i\Phi} \right\rangle , \tag{V.2}$$

where the Fresnel diffraction parameter is given by (III.32) and is proportional to the ratio of the Rayleigh length to the 1-dimensional gain length. If $F_d \gg 1$ diffraction is negligible over one gain length and the 1-D model becomes a good approximation. Apart form a factor this condition is similar to condition (III.8) introduced before.

To solve the coupled set of equations (IV.1), (IV.2), (V.2) we follow a procedure similar to that used for the long bunch case in section IV-2, and used by several authors for the 3-dimensional case (Krinsky and Yu, 1987; Kim,1986b; Chin, Kim, Xie, 1992). We assume that the system is Hamiltonian and that the electron phase space distribution fulfills Liouville's equation, so that the phase space density is a constant of motion. To simplify the problem the electron beam is rigid with no transverse betatron motion. As a result all equations of motion in the transverse direction are zero and the derivatives respect to the transverse canonical variables are dropped. We use again Vlasov equation (IV.15), assume that the distribution function can be factorized into the transverse and longitudinal coordinates and introduce a function $g(\bar{x}, \bar{y})$ to describe the transverse distribution.

As in the long bunch 1-D model discussed in section IV-2, the phase space distribution is expanded into a Fourier series starting with a term with a uniform phase distribution, giving a bunching factor equal to zero. This time, however we solve the problem as an initial value problem. We assume

$$f(\Phi, \eta, \bar{x}, \bar{y}, \bar{z}) = g(\bar{x}, \bar{y})[f_0(\eta) + f_1(\eta, \bar{z}) e^{i\Phi - i\delta_\gamma \bar{z}}] . \tag{V.3}$$

As we did before the exponential term $\exp(-i\delta_\gamma \bar{z})$ takes into account the energy detuning from the

resonant energy. As in section IV-2 we also assume that the field, like the first order distribution function, is of the form $\hat{\alpha} = \tilde{\alpha}(\bar{z})\exp(-i\delta_\gamma \bar{z})$.

Selecting the first harmonic terms of the Fourier series in Vlasov's equation we obtain

$$\frac{\partial f_1}{\partial \bar{z}} + i(\eta - \delta_\gamma)f_1 + \tilde{\alpha}\frac{\partial f_0}{\partial \eta} = 0, \tag{V.4}$$

with δ_γ given by (IV.19). To solve this equation we apply a Laplace transform

$$\bar{f}_1(\eta,p) = \int_0^\infty f_1(\eta,\bar{z})e^{-p\bar{z}}\, d\bar{z}. \tag{V.5}$$

Since $f_1(\eta,0) = 0$ the solution of (V.4) is

$$\bar{f}_1(\eta,p) = -\frac{\bar{\alpha}(p)}{p+i(\eta-\delta_\gamma)}\frac{\partial f_0}{\partial \eta}. \tag{V.6}$$

Using this solution we evaluate the bunching factor

$$B(\bar{z}) = \left\langle e^{-i\Phi} \right\rangle = \int d\bar{x}\, d\bar{y} g(\bar{x},\bar{y}) \int d\Phi\, d\eta f(\Phi,\eta,\bar{x},\bar{y},\bar{z})e^{-i\Phi}. \tag{V.7}$$

Note that $g(\bar{x},\bar{y})$ is a distribution normalized to one. Using the expression (V.3) of the distribution function we obtain

$$B(\bar{z}) = \int f_1(\eta,\bar{z})d\eta. \tag{V.8}$$

The solution connects the derivative of the initial energy distribution f_0 with the current modulation f_1 of the electron beam.

Inserting the solution (V.7) into (V.2) and using the field Laplace transform we obtain

$$[\nabla_T^2 - ig(x,y)D - iF_d(p - i\delta_\gamma)]\bar{\alpha}(x,y,p) = iF_d\hat{\alpha}(x,y,0), \tag{V.9}$$

with the dispersion term given by

$$D = \int_{-\infty}^\infty \frac{1}{p+i(\eta-\delta_\gamma)}\frac{\partial f_0}{\partial \eta}d\eta. \tag{V.10}$$

The right hand side of the field equation (V.9) is the initial field distribution at the entrance of the FEL and defines the initial value problem for the 3-dimensional FEL model. This term doesn't

change the different solutions – the FEL eigenmodes – of the homogenous partial differential equation but defines only the different coupling of these eigenmodes to the initial field. Without loss of generality we have to find the solution for the homogeneous partial differential equation.

The overall form is similar to a two dimensional Schrödinger equations of the form $[\nabla^2 + V(x,y)]A(x,y) = 0$ and the role of the Laplace parameter p is very similar to that of an eigenvalue in quantum mechanics with the significant difference that the effective potential, and therefore the "eigenvalues", have complex values. We discuss the solution of this problem in section V-3 where we will discuss the radiation modes and the corresponding growth rates.

The 1-dimensional limit, with $\nabla_T^2 \hat{\alpha} = 0$ and $g(x,y) = 1/\pi r_0^2$, requires the explicit solution $D = F_d(p - i\delta_\gamma)$ to satisfy the field equation. This is identical to the 1D dispersion equation (IV.22) if we use the relationship

$$\mu = ipF_d^{1/3}. \tag{V.11}$$

This equation establishes a relationship between the 1-D and 3-D FEL parameters

$$\rho_{3D} = F_d^{1/3}\rho \tag{V.12}.$$

with a corresponding change in gain length. Notice the weak dependence of the 3-D FEL parameter on the Fresnel diffraction parameter.

2. FEL Eigenmodes

The simplest model to study 3D effects is a round beam with constant electron density and radius r_0. The transverse density has the constant value $1/\pi r_0^2$ for any radius smaller than r_0 and zero everywhere else. Even if the electron beam is cylindrically symmetric, the electromagnetic waves might not have the same property. Therefore we expand the general solution for the field in a series of radial and azimuthal modes

$$\tilde{\alpha}(r,\phi) = \sum_{n,m} c_{n,m} R_{n,m}(r) e^{im\phi} . \tag{V.13}$$

The differential equation for the radial modes is

$$\left[\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} + \upsilon^2 - \frac{m^2}{r^2}\right] R_{n,m}(r) = 0,$$
$$\text{with } \upsilon = v = \sqrt{2iD - 2ipF_d}, \text{ for } r < 1, \tag{V.14}$$
$$\upsilon = ig = i\sqrt{2ipF_d} \text{ for } r > 1.$$

Within the electron beam the only valid solution is a Bessel function $J_m(nr)$, yielding a finite value at the origin. Outside the beam the solution is a modified Bessel function $K_m(gr)$, dropping in amplitude faster

than $1/r$ for large values of r, giving a finite value for the total radiation power.

At the electron boundary the amplitude and derivative of the radiation field must be continuous. These requirements can be combined into one continuity condition (Watson, 1987)

$$vJ_{m+1}(v)K_m(g) - gJ_m(v)K_{m+1}(g) = 0,$$ (V.15)

where we used the same notation as in (V.14). For a given azimuthal mode m there are an infinite set of solutions for the "general" dispersion equation, which are marked with the index n. In the following we label the different modes as TEM_{mn}, indicating that the mode is primarily a transverse electrical and magnetic field even if, due to the curvature of the finite size mode, there are also longitudinal field components.
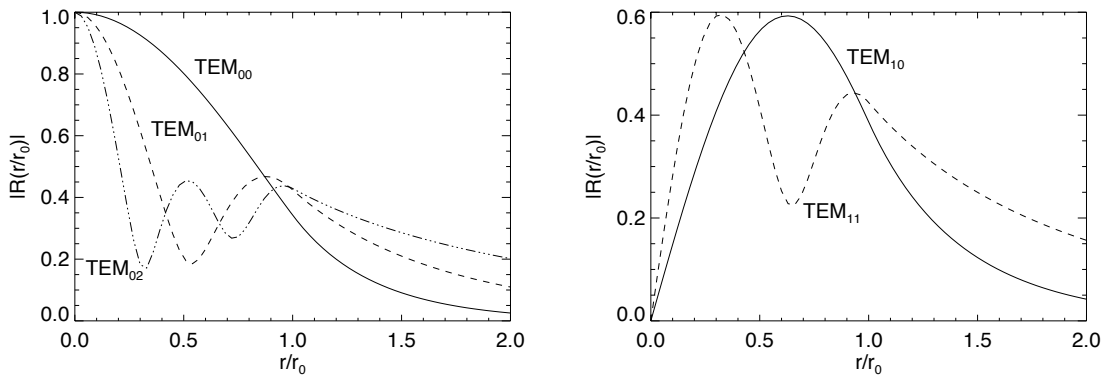

FIG. 28. Lowest order dispersion equation solutions for the FEL eigenmodes.

The lowest order solutions of the dispersion equation for a step beam profile are shown in FIG. 28. The general behavior is very similar to problems in quantum mechanics. In general each mode has $n+m$ roots within the electron beam, though the complex value of the Bessel function argument shifts the value away from zero in field amplitude. The complex argument produces also a phase front curvature, increasing the instantaneous diffraction of the FEL modes. Even the fundamental TEM_{00} mode is not diffraction limited at $\lambda/4\pi$. In addition, higher modes have a larger transverse extension. Unlike the propagation of a Gaussian radiation field in free space, the FEL eigenmodes are constant in shape and size, but grow in amplitude. The effect, called gain guiding, appears like a guiding effect in an optical fiber (Scharlemann, Sessler and Wurtele, 1985). However, unlike a truly guided mode, such as optical modes in fibers, this guiding exists only because the energy flowing in the transverse direction by diffraction is compensated by the field amplification at the electron beam location.
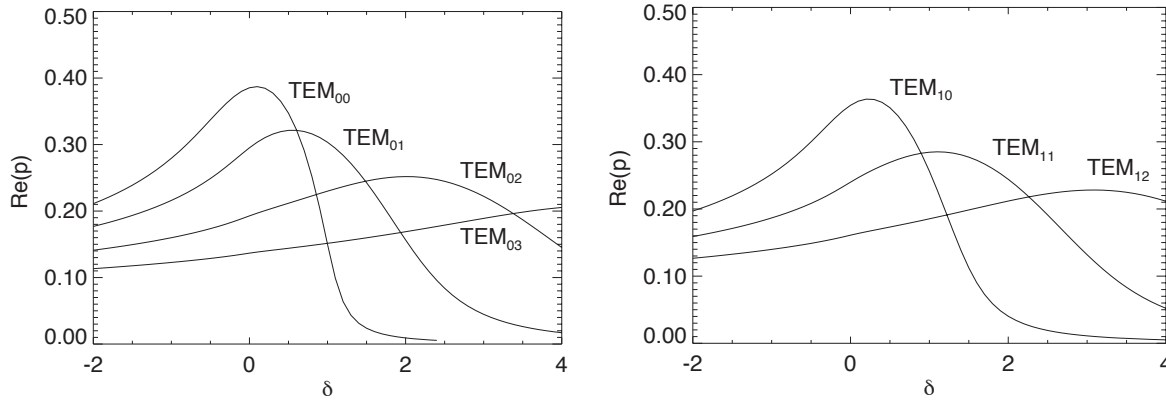
FIG. 29. Growth rate of the lowest $TEM_{n,m}$ modes as function of detuning.

The detuning is an important parameter and has different impact on FEL modes. Similarly to the 1D model, there is one optimum value for the eigenvalue, yielding the largest growth rate. There is still some exponential growth for detuning values smaller than the optimum detuning while there is a hard limit for larger values, where no gain is possible. The difference with the 1-D model is that each mode has a different optimum detuning value as seen in FIG. 29. The optimum resonance is shifted towards larger beam energies for higher FEL modes. The reason is the additional phase slippage, the Guoy phase (Siegman, 1986), introduced when a radiation field goes through a focal waist or - equivalently- is emitted from a finite size source. This additional phase slippage increases the phase velocity of the radiation field, requiring a faster electron velocity to keep radiation field and micro bunches in phase. Hence the shift of the optimum detuning point towards larger values. Note that for the FEL eigenmodes this diffraction effect depends solely on the shape of the eigenmode and not on the exponentially growing amplitude.

Another fact is that the maximum growth rate (the real part of p) is not the same for all modes. Higher modes have smaller growth rates than the fundamental, because of two effects. First, higher modes exhibit stronger diffraction, leaking more radiation transversely out of the electron beam. The field amplitude is effectively reduced and the local FEL amplification is reduced. Second, higher modes have roots, where there is no interaction between radiation field and electrons do not contribute to the mode. As a result the overall number of electrons, which emits coherently into the mode is reduced and the effective current is lower. These effects depend on the Fresnel diffraction parameter and should vanish when the impact of diffraction is smaller. FIG. 29 shows the growth rate for the highest modes.
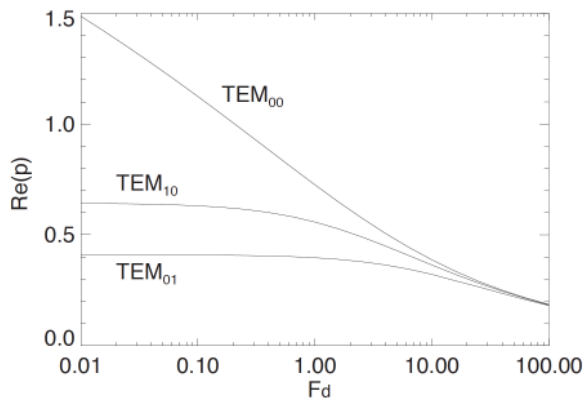
FIG. 30. Growth rates for different modes as a function of the Fresnel parameter.

For large values of F_d the growth rates of all modes merge to one value. The overall dependence in this limit is $F_d^{-1/3}$, confirming the 1-D limit of the 3-D model. The relation in amplitude between all modes is preserved from the initial coupling in the start-up regime till saturation and the amplification is a local effect without any cleaning in the modal structure. The impact on the transverse coherence is discussed in a later section. In the opposite limit of a strongly diffracting radiation field, higher modes are suppressed in comparison to the fundamental mode, but even the fundamental mode growth is slower than predicted by the 1-D model, which diverges with the cubit root of F_d. The penalty of having radiation field leaking out transversely damps the possible growth rate. From an optimization point of view it is beneficial to avoid values of F_d below 0.1. On the other hand large values of F_d, above 10, do not separate the fundamental mode sufficiently enough from the higher modes (Kim, 1986; Saldin *et al.*, 2003) when starting from noise in a SASE FEL, which couples almost equally to all modes. The stronger growth rate of a single mode corresponds to growing transverse coherence.

For strong diffraction the FEL eigenmodes are much wider than the electron beam distribution and the explicit shape of the distribution has little impact. Also the FEL performance scales rather with the current of the electron beam than the electron density. At very short wavelength, typically in the few Ångstrom regime, the FEL mode is actually smaller than the electron beam and sensitive to the distribution because the vanishing diffraction do not washout the feature in the particle distribution. To illustrate this we assume vanishing diffraction (F_d is infinity) and the 1D theory is valid. Also the beam is seeded with an infinite wide radiation field. This assumption is not really needed but simplifies the discussion significantly because SASE FELs start from "hot spots", cluster of electrons that exhibit an above average bunching factor. These hot spots can be small in transverse extensions and thus violate the assumption of vanishing diffraction. However when the transverse coherence grows larger this effect is less pronounced, the bunching and field phase information is spread out with diminishing diffraction. This illustrates again the problem of reaching transverse coherence for large diffraction parameters.

For a flat transverse electron distribution the local growth rate is identical over the entire bunch and the resulting bunching profile is also flat. In contrast, a Gaussian radial distribution gives a transverse dependence of the 1-D FEL parameter. After an undulator length L the field has a radial profile

$$A(r) = \frac{A_0}{3}\exp(\sqrt{3}k_U\rho(r)L)$$ (V.16)

.

After 15 gain lengths the rms mode size of the photon beam is related to the electron beam size with $\sigma_{field} = \sqrt{\ln(15)/3}\,\sigma_{beam}$, and is about 5% smaller than that of the electron beam. This applies also to the bunching profile. Comparing both beam profiles, the flat beam yields a flat bunching profile, while for a Gaussian distribution the bunching profile is marginally smaller, but still Gaussian. The angular distribution of the field emission is proportional to the form factor of the bunching profile, which can be related to the Fourier transform of the profile. So while both types of profiles may have the same beam size the divergence of the flat profile is significantly larger, because of the sinc^2 distribution of the far field pattern. The Gaussian distribution is diffraction limited and thus has the smallest possible divergence.

This example illustrates that for X-ray FELs some substructure in the electron beam distribution can increase the divergence of the beam as observed at LCLS. This higher mode content is also illustrated by FEL theories, which are based on a set of orthonormal modes of propagation (Hemsing, 2008; Webb *et al.*, 2011). Note that these higher modes are phase locked to each other forming a single FEL eigenmode, while the different FEL eigenmodes are not.

The initial value problem has the source term

$$s_m(r) = 2iL_d \int A_0(r) e^{-im\phi} \, d\phi \qquad (V.17)$$

in the field equation. The general solution of this equation is now the sum of a particular solution of the inhomogeneous equation and a solution of the homogenous equation (Saldin *et al.*, 1993). The continuity condition, (V.15), and the Laplace inverse transformation relate the initial field to the amplitude of the FEL eigenmodes as an overlap integral of the initial field and the FEL eigenmodes functions.

When operating an FEL as a seeded amplifier the higher mode content is minimized using a radiation seed radial size much larger than that of the FEL eigenmodes. The reason is that higher modes have zeroes and the field amplitude changes sign over the transverse plane. Assuming the external field to be almost constant radially, the overlap integral amplitude is small because of the variation in sign of the higher modes. However using a much larger seed signal wastes a large amount of radiation power, because non-zero field amplitudes are present outside the electron beam and have no overlap at all. Thus, to increase the coupling efficiency the seed field should match the size of the FEL mode. Although higher modes are excited the fundamental mode is seeded with more power and the total FEL length needed to reach saturation is reduced. Seed field transverse sizes much smaller than the electron beam radius have a larger coupling to the higher modes, because the overlap integral is restricted to a smaller area with less changes in sign of the eigenmodes.

In a similar way the initial bunching distribution couples to the FEL eigenmodes. For SASE FELs, which are starting from incoherent undulator radiation, these areas of initial bunching are small and many FEL eigenmodes are excited almost with the same coupling efficiency, in strong analogy to a Dirac delta function having a wide range of frequencies with almost equal amplitude. Mode clearance only occurs because the different FEL eigenmodes have different growth rates. The far field patterns for a SASE FEL with LCLS like parameters, when the FEL amplification process is stopped at different position along the undulator, are shown in FIG. 31. It can be seen that the distribution merges into a smooth, Gaussian like distribution. The speckle pattern in the first picture corresponds to a single wave front in the FEL pulse. However a SASE FEL has many independent wave front ordered in longitudinal direction and a detector would only measure the sum of all these wave front. In comparison to the single wave front the distribution is broader and smoother, arising from the sum of many independent, unrelated speckle patterns. When a high degree of coherence is achieved, the difference between single wave front and projected distribution is marginal, though the projected distribution is always larger because the poor longitudinal coherence of the FEL pulse means that all frequencies within the FEL bandwidth can be excited, all of them with slightly different divergence angles (Huang, Z. and Kim, K.-J., 2007).
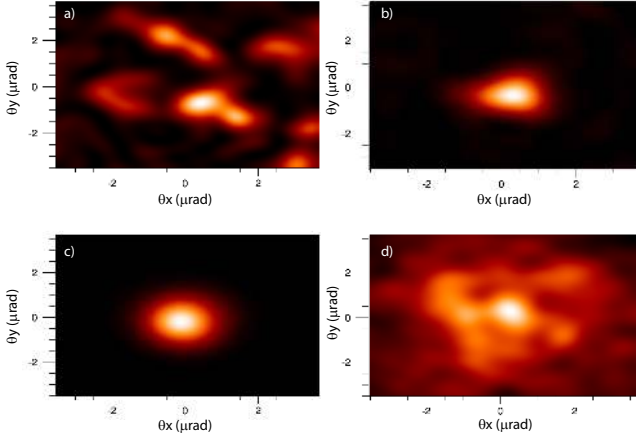
FIG. 31. Farfield pattern of a single wavefront: a. after three gain lengths; b. after seven gain lengths; c. at saturation; d. Summation of the far field pattern of all wave fronts within an FEL pulse after three gain length.

The growth and excitation of FEL eigenmodes are based on the assumption of a rigid electron beam, considered as a passive medium where a bunching phase front can be imprinted. This is an approximation that neglects the effects of the electron transverse betatron oscillations. Transverse coherence can be increased by betatron oscillations, spreading out the local radiation phase over the electron beam, carried indirectly by the bunching phase of the electrons, even with vanishing diffraction.

As discussed in section I, equation (I.16), there are two kinds of electron motion. The first is the betatron motion of the electron itself. This is rather slow and rarely more than one betatron oscillation over the entire undulator length, because stronger focusing would induce an axial velocity spread, acting as a Landau damping term to weaken the bunching process of the FEL, as discussed in detail in the next section.

The second motion is the "breathing" of the beam envelope in the focusing quadrupole array. The electron beam size oscillates with the periodicity of the quadrupole lattice. The beam envelope variation can be significant when the average betatron wave number, β_B, is made smaller. Unlike the main betatron motion the beam envelope variation can be on the scale of the gain length or even smaller. Electrons with large betatron amplitudes are effectively smeared out in transverse position and so is the bunching phase front. This has little impact on the fundamental mode because it has the same sign over the entire electron bunch. Higher modes have nodes and areas with opposite signs.

For a rigid beam the modes for the field and bunching phase front match, but including beam envelope variation the bunching profile is scaled in size and the modes do not overlap any longer. It can occur that there are some overlap of field and bunching with opposite signs and the FEL performance is locally disrupted. This is illustrated in FIG. 32and the impact on the FEL growth mode is shown in FIG. 33 and FIG. 34. As a result SASE FELs with a strong beam envelop variation might have some penalty in gain length but exhibits better coherence because the growth rate of higher FEL eigenmodes is more strongly suppressed (Reiche and Prat, 2012).
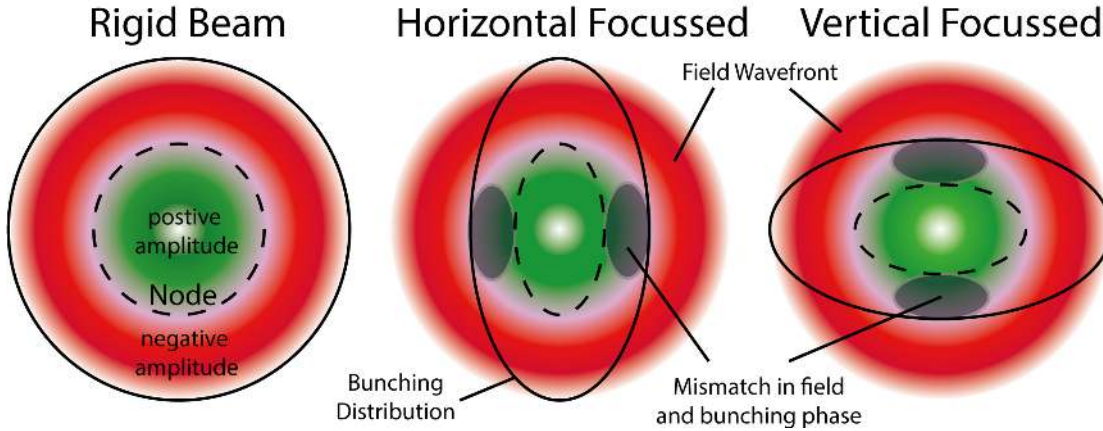
FIG. 32. Overlap between the bunching distribution and radiation wave front for the first higher radial FEL eigenmode. The bunching distribution is distorted by focusing and do not overlap well with the radiation field.

FIG. 33. Radiation power of the fundamental mode, first higher mode with a rigid electron beam and an electron beam in a FODO channel (dotted, solid and dashed line, respectively).

FIG. 34. Growth rate of the first higher FEL mode for a rigid and a focussed electron beam (solid and dashed line, respectively). The oscillation in the dashed line has the same periodicity as the quadrupole spacing of the focusing channel.

75

3. Three-dimensional theory with betatron oscillations

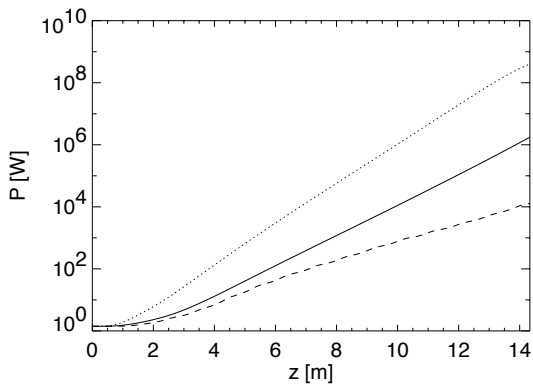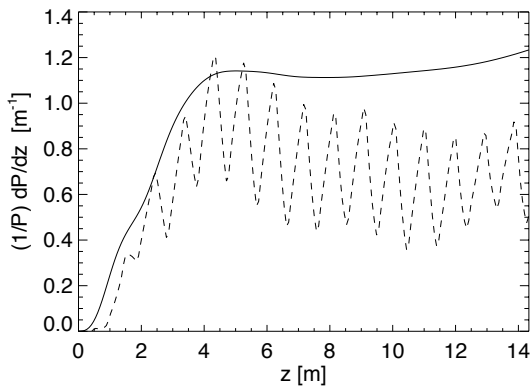The 3-dimensional model discussed in the previous section includes the effects of diffraction and of the electron velocity variation due to their energy spread. It was recognized in the early days of the LCLS studies that an important limiting factor in the high-gain regime at X-ray energies is the transverse position and angular spread due to betatron oscillations around the ideal orbit in the undulator. We considered the betatron oscillations in section I-9 and in section I-10 we evaluated their effect on the line width of the spontaneous radiation, showing that to first approximation they can be described by an effective energy spread, depending, as shown in equation (I.89), on the beam emittance, ε, and the beam focusing, characterized by the betatron oscillation wave number $\beta_{B,F}$.

Kim (1986) first derived an equation satisfied by the guided modes of the radiation field, in the exponential regime. Later on Yu, Krinsky and Gluckstern, (1990) obtained the first approximate solution for the fundamental mode, in terms of four universal scaling parameters. Ming Xie (2000) obtained both exact and variational solutions of the eigenmodes and expressed the gain length in terms of a convenient fitting formula. The three-dimensional initial value problem can be also solved by Van Kampen's eigenmode expansion (Kim, 1986b; Xie, 2001; Huang & Kim, 2000, 2001, 2007).

This comprehensive three-dimensional FEL theory follows the same approach discussed in the previous section, based on the linearized Vlasov equation, extended now to a six-dimensional phase-space. To the equations used before for the electromagnetic field, the electron energy and phase we must now add the equations for the transverse motion, (III.27), (III.28).

The phase space distribution is a function of the longitudinal and transverse variables, $f(\eta,\Phi,\vec{x},\vec{p},\overline{z})$. As before we expand it as a zero order term, uniform in phase, and a first order perturbation term

$$f(\eta,\Phi,\vec{x},\vec{p},\overline{z}) = g(\eta,\vec{x},\vec{p}) + f_1(\eta,\vec{x},\vec{p},\overline{z})e^{i\Phi(1-2\rho\delta)} . \qquad \text{(V.18)}$$

Note that the detuning term in the exponent in this equation is equivalent to $2\rho\delta\Phi = \delta\overline{z}$, appearing in the field envelope equation.

The equation for the first order term is

$$\frac{\partial f_1}{\partial \overline{z}} + \frac{\partial \Phi}{\partial \overline{z}}\frac{\partial f_1}{\partial \Phi} + \frac{1}{\overline{\beta}_{B,F}}\left(\vec{p}\frac{\partial f_1}{\partial \vec{x}} - \vec{x}\frac{\partial f_1}{\partial \vec{p}} \right) + \frac{\partial \eta}{\partial \overline{z}}\frac{\partial g}{\partial \eta} = 0 \qquad \text{(V.19)}$$

Contrary to what we have done in the previous section, this equation cannot be solved explicitly in the Laplace domain because of the presence of the $\overline{z}$ dependent transverse betatron motion.

We assume that the zero-th order distribution function in the transverse coordinates is a Gaussian, and is independent of the bunch coordinate

$$g(\vec{x},\vec{p},\eta) = \frac{\eta_\Phi}{(2\pi)^{3/2}\sigma_\eta}e^{-\frac{\vec{x}^2+\vec{p}^2}{2}-\frac{\eta^2}{2\sigma_\eta^2}} , \qquad \text{(V.20)}$$

76

where $\vec{x}, \vec{p}$ are the transverse position and angle normalized to their rms values $\sigma_x, \sigma_x / \beta_{B,F}$.

We assume that the dependence of the radiation field, bunching and perturbed distribution function on the bunch coordinate is of the form $\hat{\alpha}(\vec{x}, \bar{z}, \bar{z}_1) = \tilde{\alpha}(\vec{x}, \bar{z}) e^{-i\delta \bar{z}_1}$. Using normalized units, the Vlasov equation for the perturbation term, assumed to be much smaller than $g(\vec{x}, \vec{p}, \eta)$, is

$$\frac{Df_1}{D\bar{z}} + i(\eta - \eta_\varepsilon) f_1 - \frac{\eta}{\sigma_\eta^2} \tilde{\alpha} g = 0, \qquad (V.21)$$

where the derivative is taken along the particles betatron trajectory. We also use the fact that the FEL parameter, ρ, is a small quantity. The quantity η_ε is the effective energy spread due to betatron oscillations

$$\eta_\varepsilon = \frac{k_r L_G \sigma_x^2}{2\bar{\beta}_{B,F}^2} (\vec{x}^2 + \vec{p}^2), \qquad (V.22)$$

introduced before in (I.89).

With these assumptions, the field equation (III.31) becomes

$$\left(\frac{\partial}{\partial \bar{z}} - i\delta - \frac{i}{F_d} \nabla_T^2 \right) \tilde{\alpha} = \tilde{B}. \qquad (V.23)$$

The bunching factor is given by

$$\tilde{B}(\vec{x}, \bar{z}, \delta) = \int_{-\infty}^{\infty} d^2\vec{p} \int_{-\infty}^{\infty} d\eta f_1(\eta, \bar{z}, \delta) \qquad (V.24)$$

We can integrate (V.21) using the method of integration along the unperturbed trajectory to obtain

$$f_1 = -\frac{\eta}{\sigma_\eta^2} \int_{-\infty}^{\bar{z}} dz' \tilde{\alpha}(\vec{x}_+(z' - \bar{z}), z') g(\vec{x}, \vec{p}, \eta) e^{i(\eta - \eta_\varepsilon)(\bar{z} - z')}, \qquad (V.25)$$

where $\vec{x}_+ = \vec{x} \cos(\bar{z} / \beta_{B,F}) + \vec{p} \sin(\bar{z} / \beta_{B,F})$ is the transverse position determined by the betatron oscillation and we assume the perturbation to be initially zero.

We use this result to evaluate the bunching with (V.24) and rewrite the field equation (V.23) as

$$\left(\frac{\partial}{\partial \bar{z}} - i\delta - \frac{i}{F_d} \nabla_T^2 \right) \tilde{\alpha}(\vec{x}, \bar{z}) = \int_{-\infty}^{\infty} d^2\vec{p} \int_{-\infty}^{\infty} d\eta g(\vec{x}, \vec{p}, \eta) \int_{-\infty}^{0} d\bar{z}' \tilde{\alpha}(\vec{x}_+(\bar{z}'), \bar{z}' + \bar{z}) e^{-i(\eta - \eta_\varepsilon)\bar{z}'} \qquad (V.26)$$

We consider the field to be a superposition of modes with a transverse profile $\tilde{\alpha}_n(\vec{x}) \exp(-i\mu_n \bar{z})$ and

rewrite the last equation as

$$\left(-i(\mu_n+\delta)-\frac{i}{F_d}\nabla_T^2\right)\tilde{\alpha}_n(\vec{x})=\int_{-\infty}^{\infty}d^2\vec{p}\int_{-\infty}^{\infty}d\eta\,g(\vec{x},\vec{p},\eta)\frac{\eta}{\sigma_\eta^2}\int_{-\infty}^{0}d\vec{z}'\tilde{\alpha}_n(\vec{x}_+(\vec{z}'))e^{-i(\eta-\eta_\varepsilon+\mu_n)\vec{z}'}.\quad\text{(V.27)}$$

This equation admits an infinite number of solutions corresponding to the transverse modes of the high-gain FEL. The fundamental mode is the one with the largest growth rate and is labeled $\hat{\alpha}_0$.

Even in this simplified form, the dispersion relation cannot be solved analytically. Two numerical methods to solve this equation were developed. The first is a matrix method, solving the integro-differential equation by discretizing the transverse space and reducing the equation to a discrete eigenvalue problem (Xie, 2000). This method yields the most accurate solutions, but does not provide much insight into the physics of the problem.

The other is a variational method. In this approach, we assume a trial function for the FEL eigenmode of the form

$$\hat{\alpha}_0(\vec{x})=\hat{\alpha}_0(R)\propto e^{-wR^2}\qquad\text{(V.28)}$$

where $R=\left|\vec{x}\right|$. The dispersion relation is then projected onto this transverse mode to give a scalar integral equation

$$\frac{\mu_0-\delta}{4w}-\frac{1}{2F_d}=\int_{-\infty}^{0}\frac{\tau e^{-\sigma_\eta^2\tau^2/2-i\mu_0\tau}d\tau}{\left(1+\dfrac{iF_d\tau}{2\bar{\beta}_{B,F}^2}\right)^2+4w\left(1+\dfrac{iF_d\tau}{2\bar{\beta}_{B,F}^2}\right)+4w^2\sin^2(\tau/\bar{\beta}_{B,F})}\qquad\text{(V.29)}$$

The variational principle (Morse and Feshbach, 1953; Luchini and Motz, 1990; Yu, Krinsky and Gluckstern, 1990; Xie, 2000) states that the trial function that best approximates the exact solution is the one for which the eigenvalue is stationary, $d\mu_0/dw=0$. This additional condition yields a system of two integral equations with two unknowns that can be solved numerically.

Using the variational method we obtain a gain length fitting formula

$$L_G=L_{G,0}(1+\Lambda),\qquad\text{(V.30)}$$

where L_{G0} is the 1-dimensional gain length and

$$\Lambda=a_1\eta_d^{a_2}+a_3\eta_e^{a_4}+a_5\eta_\gamma^{a_6}+a_7\eta_e^{a_8}\eta_\gamma^{a_9}+a_{10}\eta_d^{a_{11}}\eta_\gamma^{a_{12}}+a_{13}\eta_d^{a_{14}}\eta_e^{a_{15}}+a_{16}\eta_d^{a_{17}}\eta_e^{a_{18}}\eta_\gamma^{a_{19}}.\qquad\text{(V.31)}$$

The diffraction, emittance and energy spread parameters are

$$\eta_d=\frac{L_{G0}}{2k_r\sigma_x^2},\eta_e=\frac{L_{g0}}{\beta_{B,F}}\frac{4\pi\varepsilon}{\lambda_r},\eta_\gamma=\sigma_\eta.\qquad\text{(V.32)}$$

The fitting coefficients are

$$a_1 = 0.45, a_2 = 0.57, a_3 = 0.55, a_4 = 1.6,$$
$$a_5 = 3, a_6 = 2, a_7 = 0.35, a_8 = 2.9,$$
$$a_9 = 2.4, a_{10} = 51, a_{11} = 0.95, a_{12} = 3, \qquad . \qquad \text{(V.33)}$$
$$a_{13} = 5.4, a_{14} = 0.7, a_{15} = 1.9, a_{16} = 1140,$$
$$a_{17} = 2.2, a_{18} = 2.9, a_{19} = 3.2$$

Note that the emittance parameter used by Ming-Xie and the normalized betatron focusing defined earlier are not independent, as the emittance parameter can be expressed as $\eta_e = F_d / 2\bar{\beta}_{B,F}^2$. The two parameterizations are equivalent as they both reduce the parameter space to the minimal set of four independent scaling quantities, when including detuning.

The gain length fitting formula (V.30) is very useful to evaluate the main characteristics of an FEL and has been widely used for parametric studies for the design of LCLS and other X-ray FELs. In FIG. 35 we show how the fitting formula can be used to evaluate the optimum transverse betatron focusing to minimize the gain length.
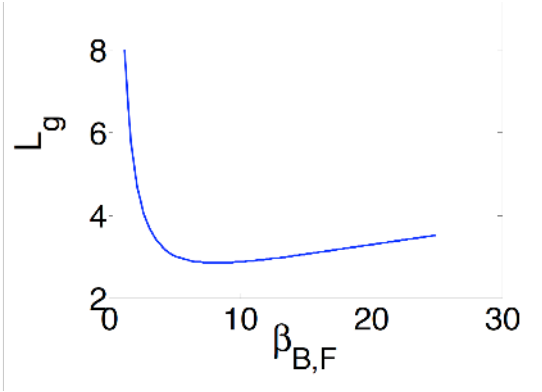
FIG. 35 Gain-length dependence of on the transverse beam focusing beta-function for standard LCLS electron beam parameters, using the gain length fitting formula (V.30). For large values the system is in the 1-D limit and stronger focusing gives higher gain by increasing the electron volume density. As beta decreases the longitudinal velocity spread due to the emittance increases, eventually disrupting the FEL gain. The beta function has an optimum when these two effects balance each other.

VI.   Equation Section 6High efficiency FEL

1. Variable parameters undulators

We saw in the previous sections that the relative energy transfer from the electron beam to the

radiation field is limited to about $1/2N_U$ in the low gain case and to the FEL parameter ρ in the high gain case. In the low gain case this can be seen most easily from the shape of the small signal gain curve in FIG. 18, changing from a positive to a negative maximum for a relative energy change of about $1/2N_U$. In the high gain case the efficiency is given by the FEL parameter, (IV.40), typically of the order of 10^{-3} or less for X-ray FELs.

The reason for the low efficiency is that the FEL synchronism condition is no more satisfied when the electron loses its kinetic energy to the radiation field. To overcome this limitation it has been proposed to use an undulator in which the period or the magnetic field (or both) vary along the axis, so that the undulator characteristics are adjusted to satisfy the synchronism condition over a wide range of electron energies (Sprangle, Tang and Manheimer, 1979; Kroll, Morton and Rosenbluth, 1981), for a fixed radiation wavelength.

We generalize the definition (II.14) of the resonance energy, considering a synchronous electron, with energy γ_s and constant phase Φ_s. The condition for its phase to remain constant as the energy changes is

$$\gamma_s^2(z) = \frac{k_r}{2k_U(z)}(1+K(z)^2).$$
(VI.1)

We assume that the undulator magnetic field, and/or its period, change to satisfy this equation, for a given radiation wavelength, following the electron energy change. If the phase is a constant, Φ_s, the equation for the energy change is obtained from (II.11) as

$$\frac{d\gamma_s^2}{dt} = \frac{ecE_r(z)K(z)}{2mc^2}\sin\Phi_s,$$
(VI.2)

having assumed the electromagnetic field phase equal to zero.

These equations relate the electric field amplitude and the electron energy to the undulator parameter and period, now a function of z, the position along the undulator axis. As the synchronous energy changes along the undulator we have to change the undulator parameter and/or period to satisfy (VI.1). For a given synchronous phase (VI.1), (VI.2) are two equations relating three quantities, λ_U, K, γ_s. The problem is not uniquely defined and we can use different undulator designs for a given energy change.

Note the difference between the "standard" FEL, described before, and the "tapered undulator" FEL described here. In the standard FEL, electrons are injected into the undulator with a range of initial phases, $0, 2\pi$, and an energy spread smaller than the FEL parameter. In the phase-energy space the electron distribution evolves to reach maximum bunching. After that the bunching, and the radiation field intensity, would decrease as shown in FIG. 23, FIG. 24.

In the tapered undulator FEL there is only one phase associated with synchronism and the electron is constantly decelerated. For an initially mono-energetic beam with random phases, the particles can have the correct energy to insure they are synchronous but the bulk of the particles will have phases different from Φ_s. The question is: "Suppose a mono-energetic beam is injected into a tapered undulator with the synchronous energy and a random distribution of the initial phases, what happens to the particles that are not at the synchronous phase?"

Examination of the electron phase space shows that particles near the synchronous phase are indeed decelerated at a nearly constant rate, but particles further away in phase are not, resulting in a loss of

efficiency. It is desirable to choose the synchronous phase, Φ_s, to capture and decelerate the maximum number of particles and provide a desired electron energy reduction, and the corresponding increase in radiation intensity, with as short an undulator as possible.

To study this problem we consider the study the motion of electrons in the phase-energy space, respect to the synchronous particle. We consider the case when the synchronous energy is a slowly varying function of z and neglect its variation over one undulator period,

$$\frac{\lambda_U}{\gamma_s^2}\frac{d\gamma_s^2}{dz} \ll 1. \tag{VI.3}$$

We also assume $\gamma_s = \gamma_R$ and that the energy remains always near the synchronous energy. Measuring the length along the undulator in gain lengths units we write the equations for the energy and phase change respect to the synchronous particle, using the same independent variable as in (IV.1), (IV.2), as

$$\frac{d\eta}{d\bar{z}} = 2|\hat{\alpha}|(\sin\Phi - \sin\Phi_s), \tag{VI.4}$$

$$\frac{d\Phi}{d\bar{z}} = \eta. \tag{VI.5}$$

In writing (VI.4) we neglected a term $2\eta\gamma_s'/\gamma_s$, which is usually negligible if (VI.3) is satisfied. We also neglected terms proportional to E_r in the equation for the phase.

Equations (VI.4), (VI.5) can be obtained from the Hamiltonian

$$H = \frac{\eta^2}{2} + 2|\hat{\alpha}|(\cos\Phi + \Phi\sin\Phi_s). \tag{VI.6}$$

To study in more detail the motion of a particle around the resonant energy we rewrite it as

$$H = \frac{\eta^2}{2} + V(\Phi) \tag{VI.7}$$

where the potential $V(\Phi)$ is

$$V(\Phi) = 2|\hat{\alpha}|(\cos\Phi + \Phi\sin\Phi_s) \tag{VI.8}$$

and is plotted in FIG. 36 ; the areas under the dashed lines represent potential wells where electrons are trapped and oscillate around the synchronous phase.
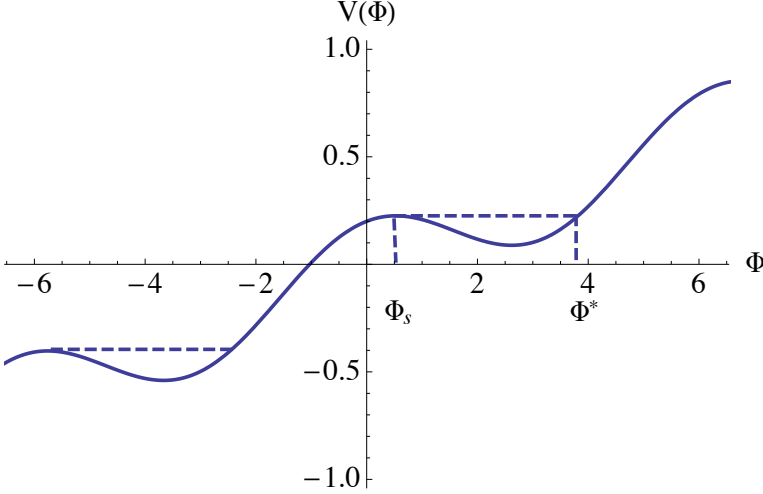
FIG. 36 Tapered undulator potential. In this plot the synchronous phase is $\pi/6$, and $|\hat{\alpha}| = 0.1$. Electrons execute stable oscillations around the synchronous particle in the areas under the dashed line.

The maximum and minimum of the potential in the interval $(0, 2\pi)$ are at the points $\Phi = \Phi_s$, $\pi\text{-}\Phi_s$. The width of the potential well is defined by

$$\Delta\Phi = \Phi^* - \Phi_s, \tag{VI.9}$$

where Φ^* is the solution of the equation

$$V(\Phi^*) = V(\Phi_S). \tag{VI.10}$$

The depth of the potential well is defined by

$$\eta_{Max} = \sqrt{V(\Phi_s) - V(\pi - \Phi_s)} = \atop 8|\hat{\alpha}|\sqrt{\cos\Phi_s - (\pi/2 - \Phi_s)\sin\Phi_s} \tag{VI.11}$$

Particles with $\eta < \eta_{Max}$, $\Phi_s < \Phi < \Phi^*$ will execute stable oscillations around the phase $\pi - \Phi_s$. The maximum energy change for the synchronous particle is obtained, from (VI.2), when $\Phi_s = \pi/2$. However for this value of the synchronous phase the depth of the potential well goes to zero and no electrons can be captured in it. For $\Phi_s = 0$ the depth is a maximum but there is no energy change for the synchronous particle. The optimum value of the synchronous phase is somewhere in between these two values and can be optimized and even changed along the undulator.

For particles in the potential well the average energy change in traversing the wiggler is equal to the change in the resonant energy, so that the efficiency in the transfer of energy from the electrons to the radiation field is

82

$$\varsigma = \frac{\Phi^* - \Phi_s}{2\pi} \frac{\gamma_s(L_U) - \gamma_s(0)}{\gamma_s(0)},$$ (VI.12)

a value that can be of the order of 10%, much larger than the efficiency at saturation, of the order of the FEL parameter, about 0.1% for X-rays FELs.
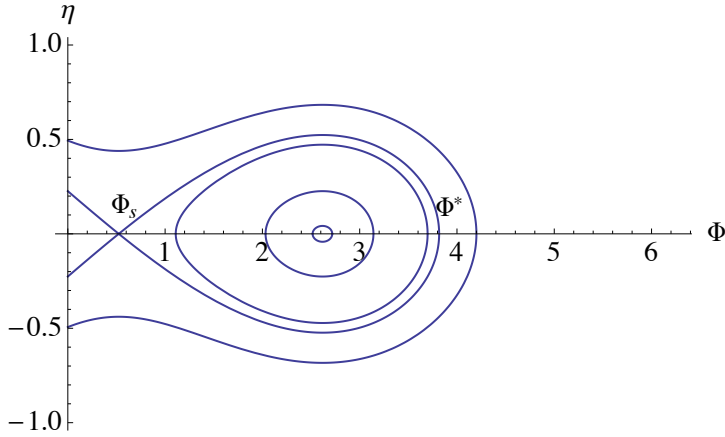

FIG. 37 Tapered undulator phase space. Electrons within the potential well with phases between Φ_s, Φ^* and $\eta < \eta_{Max}$ oscillate around the synchronous particle and follow, on average, its energy change.

There are however limitations on when to start the tapering and how fast one can change the resonant energy and on the number of electrons that remain trapped in the potential well during this process (Scharlemann, 1990; Colson, 1990). If the initial radiation field is smaller that the saturation value it is convenient to operate with a constant period until saturation is reached. Just before this point is reached tapering can start. Initially the change in synchronous energy must be small, with a synchronous phase near to zero, to optimize the trapping and maintain a large phase acceptance. Later the synchronous phase can be increased to maximize the energy transfer. In most cases these considerations lead to an undulator design with a quadratic decrease of the magnetic field starting just before the saturation point.

2. 3-dimensional effects, sidebands and other effects limiting the efficiency

Tapering the undulator is a good approach to increase X-Ray FELs efficiency. However several effects, not included in the previous discussion, limit the efficiency also for this case. The main effects are: variations of the radiation electric field in the longitudinal and transverse directions; diffraction effects; sideband instabilities.

A SASE FEL radiation intensity profile is not uniform and consists of spikes, as discussed in the previous sections. When tapering a SASE FEL it is not possible to optimize the synchronous phase for all electrons along the bunch, leading to an efficiency reduction (Fawley *et al.*, 2002) when compared to the monochromatic FEL amplifier discussed before. While tapering still gives an advantage respect a non-tapered undulator, as shown FIG. 38, the efficiency is nearly a factor of two below the amplifier case. In the following we will discuss the FEL amplifier starting from a

small signal, reaching saturation and continuing with a tapered undulator section. The small nearly monochromatic signal can be obtained with one of the seeding methods discussed in the next section.
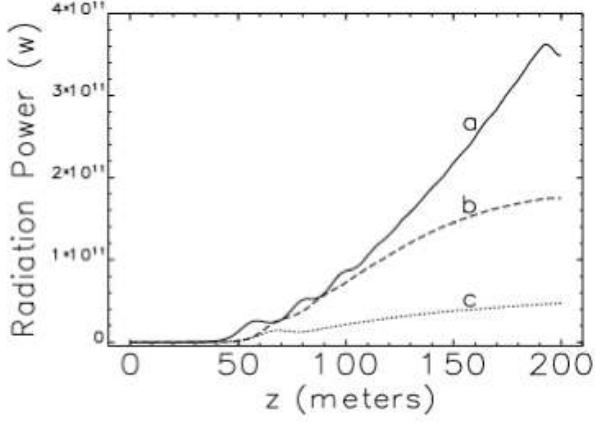

FIG. 38 Comparison of power output for: a) a tapered monochromatic amplifier; b) a tapered SASE FEL; c) a non-tapered SASE FEL. From Fawley et al., 2002. Simulations based on the Ginger code (Fawley 1995).

Other important effects are the radial dependence of the radiation electric field (Luccio and Pellegrini, 1980), the transverse electron distribution (Emma et al., 2014) and diffraction effects. Existing simulation codes, discussed in Section VIII, take into account all these effects and can be used to simulate high efficiency tapered X-ray FELs. However the requirements for computing time is quite large. A physical model to optimize the tapering to obtain the largest X-FEL peak power for given beam characteristics, including all these effects, has also been developed (Jiao, *et al*., 2012). The model describes optical guiding of the radiation and diffraction effects using an optical fiber approximation, relating refractive focusing with the electron beam bunching (Scharlemann, Sessler, and Wurtele, 1985; Sprangle, Ting and Tang, 1987; Hafizi *et al*., 1990). As shown in these references, the coherent interaction between the radiation and electrons can optically guide and focus the light. Because of its microbunching, the electron beam has an effective complex index of refraction

$$n-1 = \frac{1}{k_r}\left(\frac{d\psi_r}{dz} - \frac{i}{E_r}\frac{dE_r}{dz}\right), \qquad (VI.13)$$

obtained identifying the change in phase and amplitude of the field with the real and imaginary part of the index of refraction. The real part of the refractive index gives refractive guiding and describes a change in the wave phase velocity from inside the beam to outside the beam. When the phase velocity in the beam is smaller than in vacuum the wave front is distorted and the radiation is focused. The imaginary part corresponds to an increase in field amplitude and to gain.

Using (II.31) and (IV.3) and neglecting the dependence on the bunch coordinate, the index of refraction can be written as

$$n-1 = -\frac{i}{k_r L_G \hat{\alpha}}\frac{d\hat{\alpha}}{dz} = \frac{i}{k_r L_G \hat{\alpha}}\left\langle e^{-i\Phi}\right\rangle. \qquad (VI.14)$$

Following the exponential gain regime in the initial part of amplifier, where the gain is large, the imaginary part of the index of refraction becomes small, near to zero, and the real part of n, describing the refractive guiding, becomes important. In this region the radiation field amplitude increases more slowly with $\bar{z}$ than in the exponential regime. The index of refraction decreases with increasing field amplitude, as shown in (VI.14), slowing down the energy transfer and possibly leading to a new saturation (Fawley, 1995; Fawley, 1996). This result shows the importance of maintaining a large bunching factor to maintain refractive focusing in the tapered part of the undulator.

Another effect important for tapered FELs, longer than the initial saturation length, is the growth of sidebands at the synchrotron frequency, the oscillation frequency of electrons in the ponderomotive potential well (Kroll, Morton and Rosenbluth, 1980). The sideband growth reduces the energy gain of the primary wave and causes a loss of electrons from the potential well, reducing the bunching factor. The combination of sideband and three-dimensional effects can lead to a second power saturation, after the exponential regime saturation, and thus put an upper limit to the efficiency of an X-ray FEL even when using tapered undulators, as shown in FIG. 39.
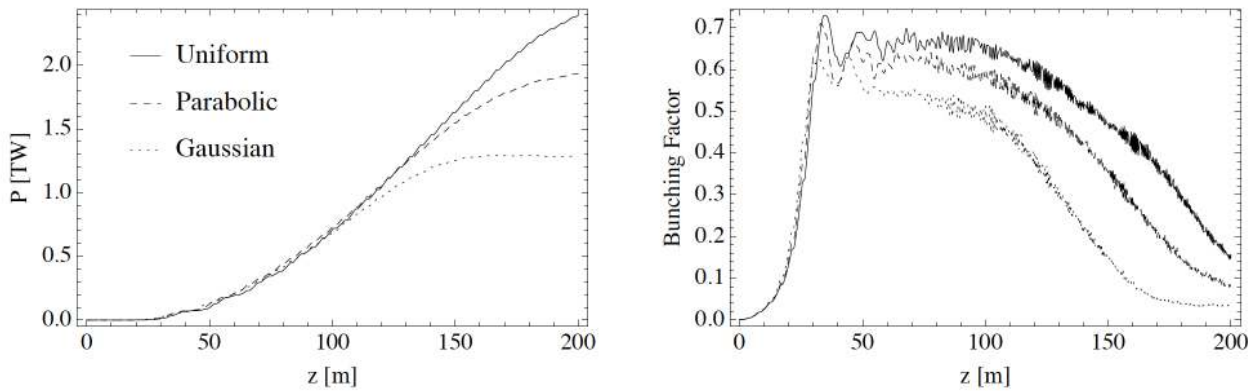

FIG. 39 Power and bunching factor as a function of longitudinal distance along the undulator for Gaussian, green, parabolic, red, and uniform electron transverse distribution. The wavelength is 1.5A and the electron bunch length, FWHM, is 16 fs. The first, exponential regime saturation is at about 40 m. A second saturation is seen around 149 m for the Gaussian case and at a longer distance for the other cases. The stronger bunching factor reduction for the Gaussian beam gives more diffraction and a larger X-ray spot size at the undulator exit. From Emma, Wu and Pellegrini, 2014.

Changing the transverse electron distribution from a Gaussian to a uniform or parabolic shape can make the radial distribution of the electric radiation field more uniform and help to maintain a large bunching factor along the tapered undulator (Emma *et al.*, 2014), as can be seen in FIG. 39, thus giving better radiation focusing. The radiation modes for the parabolic or uniform electron distribution are no more a simple Gaussian. The transverse coherence of the X-ray pulse in these configurations has also been evaluated (Emma *et al.*, 2014) and found to satisfy the requirements for coherent imaging.

Increasing the efficiency of an FEL amplifier with undulator tapering was first demonstrated at a wavelength of about 1 cm (Orzechowski *et al.*,1985). More recently an experiment was done with a seeded FEL at Brookhaven at 793 nm reaching an efficiency of about 1% (Wang *et al.*, 2009] and increasing the energy per pulse by a factor of 3 on the fundamental, preserving a narrow spectral line.

An experiment at LCLS (Ratner *et al.*, 2009)) using the limited tapering capability, about 1%,

of the undulator, has demonstrated a factor three increase in efficiency in good agreement with simulations. Extrapolation to longer undulators (Fawley *et al.,* 2011; Serkez *et al*., 2014) shows that the efficiency can be increased to several percent while preserving the transverse coherence properties needed for coherent diffraction imaging (Emma *et al*., 2014), using the FEL as a self-seeded amplifier.

Equation Section 7VII. Seeding methods for FELs

In recent years enormous progress has been achieved in the theoretical understanding and experimental demonstration of FEL seeding. In this section we review the state of the art for FEL seeding using different techniques and explore their potential to produce radiation pulses approaching the transform limit in a wide range of experimental conditions.

The purpose of seeding for FELs is threefold: overcoming the inherent limitation of longitudinal coherence in a SASE FEL configuration (Saldin, Schneidmiller and Yurkov, 2008) (and thus improving on the brilliance of the FEL output signal), synchronizing the FEL signal with an external signal, and improving the stability of the FEL power from shot-to-shot by introducing a well-defined seed signal unlike the white noise fluctuation of the spontaneous radiation within the FEL bandwidth. The major advantage of the FEL is the resonant wavelength tunability, obtained by changing either the driving electron beam energy or the strength of the undulator field. Therefore any seeding source must exhibit the same tuning capability. The fundamental problem is finding a suitable source, which can be tuned in the same range as the FEL.

Once a tunable seed source has been identified it has to fulfill a second constraint, which is to overcome the shot noise power of the electron beam. Seeding with a power below the power level of the spontaneous radiation would result in SASE performance. This puts a limit on the shortest wavelength, at which seeding can be achieved.

Seeding sources typically have lower output power efficiency at shorter wavelength. On the contrary the shot noise power actually grows with frequency as (Gianessi, 2004)

$$P_{noise} \approx mc^2\gamma\omega_r\rho^2 / 2, \qquad\qquad (VII.1)$$

where ρ, ω_r, $mc^2\gamma$ are, as usual, the FEL parameter (Bonifacio, Pellegrini and Narducci, 1984), the radiation frequency and the electron energy. As an example, for the SwissFEL (Ganter, 2010) parameters at 5 nm and a beam energy of 2.1 GeV the shot noise power is around $P_n = 100$ W. Note that for seeded FELs only 1/9th of the power couples to the exponentially growing mode and that further losses are given by the mode mismatch between the seeding mode and the FEL eigenmode (optimum cases have about 50% coupling efficiency). Thus a seeding power level equivalent to the shot noise power level would be around 2 kW. For an improved signal-to-noise ratio the seeding power has to exceed that value by a wide margin, 10 to 100 times.

1. Direct seeding

Direct seeding refers to any methods where the seed signal wavelength is the same as the FEL resonant wavelength and its power level is above the shot noise power but below the FEL saturation power. Note that the 1D theory puts a limit on how much seed power can be coupled to the exponentially growing

mode. Only one ninth is amplified and about two gain lengths are required before any change in the radiation power becomes measurable, overcoming the lethargy regime of a seeded FEL. The coupling efficiency is further reduced by the need to match the seed mode and the fundamental FEL eigenmode (Saldin, Schneidmiller and Yurkov, 1993). With strong distortion of the phase front or unmatched mode sizes, the effective seed power is significantly reduced. Finally, the bandwidths have to be matched as well. If the seed signal has a bandwidth larger than then FEL bandwidth (e.g. by seeding with a pulse length shorter than the coherence length of the FEL) only a fraction is picked up and amplified. Note that in this case the peak brightness of the FEL output is not improved with respect to a SASE FEL because the entire FEL bandwidth is excited.
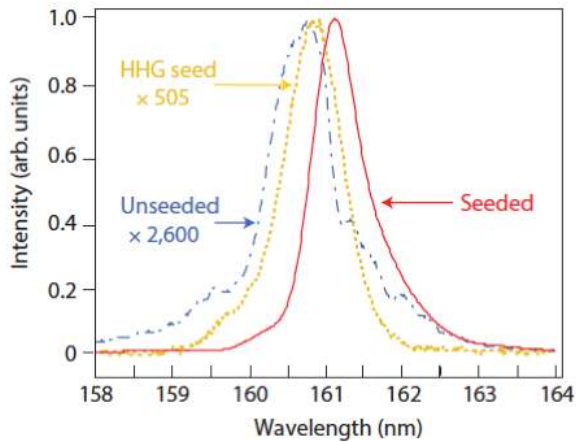

FIG. 40: Spectra of seeded and SASE FEL at the SCSS Test Facility. The SASE spectrum was scaled in amplitude to fit in the plot (Lambert *et al.*, 2008)

Lasers based on High Harmonic Generation (HHG) in noble gases are the most promising seed source for direct seeding (Ferray *et al.*, 1988; Labat *et al.*, 2011) at wavelengths in the ultraviolet and extreme ultraviolet. In the HHG process a strong drive laser field strips off electrons from the atoms by tunnel ionization and accelerates them away from the atom. A half cycle later these electrons are accelerated back towards the ions with a chance of recombination. In this case a photon is emitted with energy much higher than the drive laser but phase locked. The resulting spectrum exhibits a rich content of odd harmonics, reaching into the VUV region and thus suitable for seeding. The HHG process preserves the transverse coherence properties of the drive laser, even if the noble gas ionization by electrons that are not recombined makes the phase matching between the drive laser and the emitted photons difficult. In reality only short pulses of a few tens of fs can be achieved with enough spectral purity in the high harmonics to be suitable for seeding.

There have been several experiments, which have demonstrated successful seeding of an FEL with an HHG signal down to 39 nm (Lambert *et al.*, 2008; Giannessi *et al* 2008; Lechner *et al.*, 2012). However it became apparent that, to overcome the shot noise limit of the electron beam towards shorter wavelength, the HHG sources need to deliver much more spectral power. This is the current focus of the HHG research. Methods with corrugated capillaries or counter propagating laser beam can reduce the phase mismatch limitation, while longer drive wavelength can reach harmonics at shorter wavelengths with more intensity. In addition long transport lines of the HHG signal should be avoided to reduce transport losses. The best solution would be a HHG source inline with the undulator axis, using an electron bypass (chicane or dogleg) to overlap the radiation with the electron beam.

2. Seeding by electron beam manipulation

Instead of being seeded by a radiation field the FEL can also be seeded with a coherent bunching at the resonant wavelength. In the beginning of the FEL the beam will emit coherently and the power will grow linearly till the FEL amplification process starts after a few gain lengths. The induced bunching must be significantly above the shot noise level.
The primary method is based on the energy modulation of the beam with a seed laser, where both wavelength and power are reasonably achievable. The induced energy modulation is converted into a current modulation by a dispersive section (magnetic chicane), following a modulation stage. If the contrast between energy modulation and energy spread is large the current spike is rather pronounced and contains higher harmonics. In a second stage the FEL, also called the radiator, is tuned to one of the harmonics. The emission is only partially coherent but sufficient to start the FEL process. This concept, called High Gain Harmonic Generation (HGHG), was proposed by Li-Hua Yu (Yu, 1991) and can be generalized to a cascade system with more than one frequency multiplication.
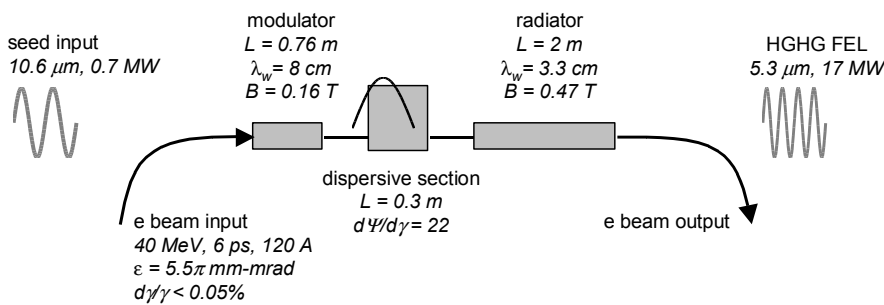

FIG. 41 shows the general configuration HGHG for the first proof-of-principle experiment (Doyuran *et al.*, 2001) done at the SDL laboratory at Brookhaven National Laboratory.
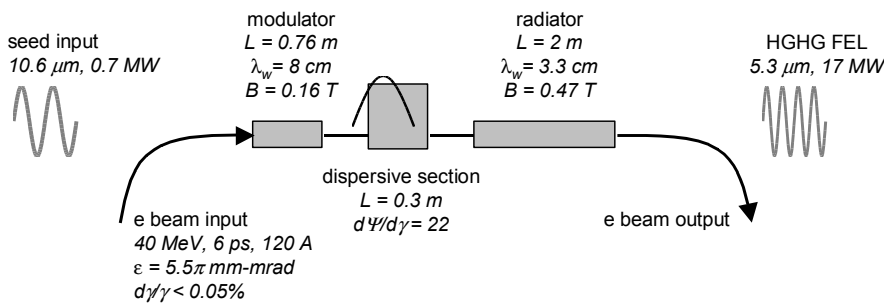

FIG. 41 Schematic layout for the proof of principle HGHG experiment at SDL. From Doyuran *et al.*, 2001.
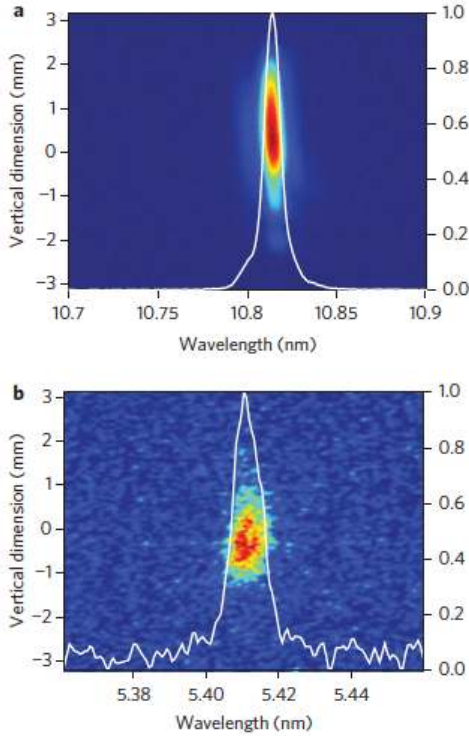
FIG. 42 Measured spectra of the Fermi FEL using two step cascaded HGHG and the fresh bunch technique. From (Allaria *et al.*, 2013a).

The required degree of energy modulation can be supplied either directly by a high power seed laser or an FEL process in the modulator, stopped before saturation is reached at the optimum energy level. The limiting factor is the induced energy modulation, $\Delta\gamma$, during the FEL amplification. The maximum bunching at the nth-harmonic is given by $B_n = \exp[-(n\sigma_\gamma / \Delta\gamma)^2 / 2]$ and drops quickly when the energy modulation is smaller than the intrinsic electron beam energy spread, σ_γ, times the harmonic number. Therefore high harmonic conversions require large energy modulations. On the other hand the final energy spread, as seen in the final radiation stage, still has to fulfill the FEL requirement $\sigma_\gamma / \gamma \ll \rho$, where ρ, the FEL parameter, is of the order of 10^{-3} for X-ray FELs. If the condition is violated the beam will emit partially coherently in the radiator, but the FEL will not reach saturation. As a consequence HGHG FELs must operate with a much smaller energy spread than comparable SASE FELs.

Fermi@Elettra in Trieste is the only FEL designed to operate in HGHG mode (Allaria *et al.*, 2012b) and has successfully operated at wavelengths from about 100 to 4 nm (Allaria *et al.,* 2013). It uses a two-step cascaded HGHG and the fresh bunch technique (Ben-Zvi, Yang and Yu, 1992) to reduce the effect of the energy spread induced by the FEL interaction. In this technique one part of the bunch is used for the initial seeding, and another part is used to generate the final radiation on the harmonic wavelength. The successful results of this method are shown in FIG. 42.

The limitation in the achievable harmonic of the HGHG scheme is overcome in a more complex configuration with two modulators and dispersive sections prior to the final radiator, in the so-

called echo-enable harmonic generation (EEHG) FEL scheme (Stupakov 2009; Xiang and Stupakov, 2009). The purpose of the first stage is to over-compress the energy modulation well beyond maximum bunching with a strong magnetic chicane. The phase space then exhibits narrow bands instead of a continuous smooth distribution at the beginning of the seeding section. The second stage will then operate as a HGHG stage generating a current spike for each band. To achieve maximum bunching the current spikes of all energy bands need to be spaced at the final radiation wavelength to add up coherently. A typical phase space distribution for the different stages of the EEHG process is shown in FIG. 43.

The advantage of the EEHG compared to the HGHG is that the first stage artificially reduces the intrinsic energy spread per band due to the strong over-compression, which allows much higher harmonic conversion in the HGHG stage at the cost of a slight increase in the energy spread. In theory this method can achieve very high harmonics with a bunching efficiency of up to $B_n = 0.39 / n$. The scheme has been demonstrated successfully at lower harmonics numbers (Xiang et al., 2010; Zhao et al., 2012). Its potential for achieving very high harmonics numbers makes it an attractive alternative to HGHG, despite its complexity of coupling two energy modulations and two chicane strengths. The limiting factor is the ability to preserve the energy bands throughout the seeding line and avoiding any blurring effects. There are two sources of degradation: the quantum fluctuation in the emission of photons of incoherent synchrotron and undulator radiation (Sands, 1955) and intra-beam scattering (Stupakov, 2011). While the first can be mitigated with gentle bending angles and long chicanes the latter requires a layout as compact as possible. Both effects combine to limit the practical use of EEHG to wavelengths larger than 1 nm.
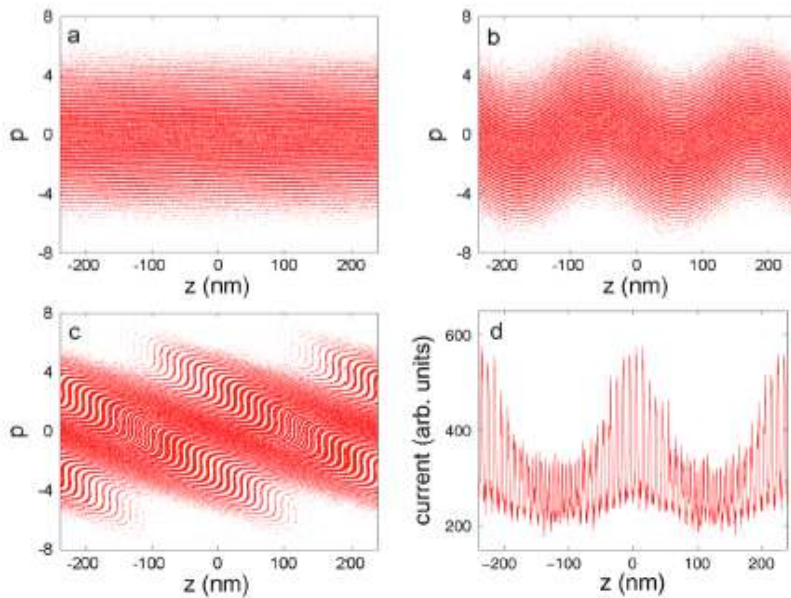

FIG. 43 Longitudinal phase space distribution in EEHG configuration after the first modulator and chicane compressor (a), the second modulator (b), and the second chicane (c). The resulting current profile is shown in the lower right plot (d). From Xiang and Stupakov (2009).

3. Cascades and hybrid configurations

When direct seeding or HGHG/EEHG seeding is not sufficient to reach the desired wavelength the processes can be combined or chained together. The most common approach is the cascaded HGHG where the output of the radiator from the first stage is used as the modulating signal of the second stage. As an example, if both stages have a harmonic conversion of $n = 5$ the final harmonic conversion would be 25. A third stage would yield an overall harmonic of 125.

The problem is the accumulated energy spread for each stage, which can easily degrade the performance of the final radiator. In addition the shot noise is amplified with the harmonic conversion and even a strong input signal can get lost in the noise of the final radiator.

The problem is partially solved by supplying an unused part of the electron bunch for each cascading stage. This is done with a delaying chicane between the radiator of the previous cascade and the following modulator. However this fresh bunch cascade requires the entire process to operate only locally, with a slice of the bunch moving slowly from the tail to the head of the bunch for each cascading step. The amount of electrons contributing to lasing in the final radiator is small and the overall pulse energy is smaller than in SASE operation. This penalty gets larger the more cascading stages are needed. So far only two stage cascades have been operated successfully at Fermi@Elettra (Allaria, 2013).

4. Self-seeding mechanism

All the previous methods require an external signal synchronized to the beam arrival time at the undulator location. The stability of the time jitter between the seed signal and the electron bunch must to be less than the bunch length. To relax the time tolerance most externally seeded FELs consider operating at lower current. As a result the FEL parameter and the power at saturation are reduced. In addition the lower FEL bandwidth restricts the amount of energy modulation of the seeding schemes, reducing the ability to scale to wavelength shorter than a few nanometers.

If the requirement for an externally locked FEL pulse is abandoned, the seed signal can be derived from the same bunch in a two-stage configuration, as shown in FIG. 44. The first stage operates as a SASE FEL amplifier, stopping before reaching saturation. The beam energy spread induced by the FEL process is limited to a value that allows successive amplification of an external signal to full saturation. Following the SASE FEL the electron beam and radiation field are separated and the radiation is filtered to select a narrow bandwidth. The filtered signal is then recombined with the electron beam and injected into the second stage of the FEL, operating as an FEL amplifier. The electron bypass has two purposes: primarily, to match the arrival time of the beam with the filtered signal and, secondarily, to remove the induced FEL micro-bunching using the momentum compaction factor of the chicane. This configuration is referred to as "self-seeding".
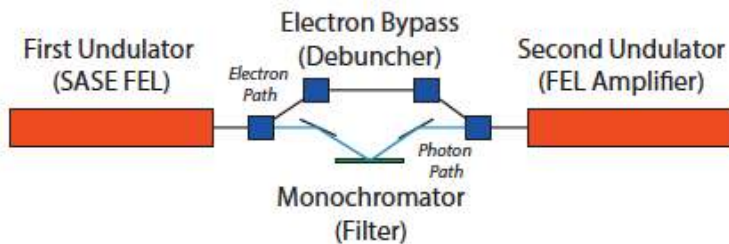
FIG. 44 Schematic layout of a self-seeding configuration

The initial idea was proposed for the soft X-ray facility FLASH (Feldhaus *et al.*, 1997) but had the conceptual difficulties that the delay in the photon path way would require a long electron bypass line, where the transport needs to be controlled by many quadrupole and sextupole magnets to preserve the electron beam properties. It was never realized on FLASH but has been successfully implemented at LCLS (Ratner, 2015), where a more compact design of a soft X-ray self- seeding chicane, less than 4 m long, has been developed (Feng *et al.*, 2010). The LCLS soft X-ray self-seeding system has a measured resolving power of 200-3000 in the photon energy range of 500 to 1000 eV. Although the resolving power is not sufficient to yield full longitudinal coherence, it still gives a significant improvement in the bandwidth.

Implementing this system in the few Ångstrom hard X-ray region would require a longer bypass because of the larger electron energy needed at these photon energies. This problem is solved using a novel concept, introduced by Geloni, Kocharyan and Saldin, 2011, that uses features in the transmission around the stop band of a Bragg reflection instead of grating monochromators. The transition between total Bragg reflection to almost full transmission has frequencies component which are significantly delayed by the crystal and ringing off behind the main SASE signal of the first stage. The electron bunch is delayed and overlaps with the trailing signal, which is the interference of the two edge frequencies of the stop band. For photon energies around 8 keV the seed signal has still significantly large amplitude well above the shot noise level. The attractive feature of this method is that the overall electron-photon delay is of the order of a few tens of femtosecond. The chicane needed to generate this delay is small chicane and can be easily integrated in the undulator line. The method has been successfully demonstrated at LCLS (Amann *et al.*, 2012) using a diamond crystal Bragg reflector. A narrowing of the FEL bandwidth by a factor 50 has been measured, as shown in FIG. 45. However the system is very sensitive to the electron beam energy jitter. For LCLS, the FEL wavelength shot-to-shot variation due to the electron beam jitter is of the order of 0.2%, much larger than the band width around the fixed wavelength defined by the diamond crystal of the Bragg reflector. As a result the FEL intensity at the second undulator exit fluctuates by almost 100%.

One inherent problem with self-seeding based on the Bragg system is that the delayed part of the transmitted field, which is seeding the second stage, exhibits a transverse shift in the position due to the effective index of diffraction around the Bragg stop band (Geloni *et al.*, 2013). This effect is mitigated for near perpendicular incident angles. Therefore the Bragg diffraction has to be optimized for different wavelengths, utilizing various planes of the crystal lattice.

In self-seeding schemes a narrow-bandwidth filter stretches the short SASE spikes with a coherence length typically smaller than the bunch length to a coherence length longer than the bunch length, so that in the second stage a well-defined radiation phase acts over the entire bunch. This effect can also be achieved if the slippage is increased in some other way to cover the entire bunch. In SASE FELs the slippage is one radiation wavelength per undulator period and the characteristic cooperation length L_c is the slippage in one gain length. The cooperation length becomes shorter for shorter wavelengths, assuming an overall constant gain length. There are several methods proposed (Thompson and McNeil, 2008; Wu, Marinelli and Pellegrini, 2012; Schneidmiller and Yurkov, 2012; Xiang *et al.*, 2013b) to artificially increase the slippage per gain length by either breaking up the undulator and interleaving the modules with small chicanes delaying the bunch or by operating on a sub-harmonic of the FEL, where the slippage is increased by the harmonic number. All methods reduce the FEL bandwidth up to a point where the bunch length is limiting the spectral width. At this point these methods are equivalent to

self-seeding methods except that they avoid filters intercepting the radiation. These are attractive alternatives if the heat load on the monochromator, mirrors or crystal is an issue. Similar to these slippage-enhancing methods is the feedback of a fraction of the FEL signal to the succeeding bunch in a high repetition machine. There the slippage is accumulated over many turns, defining the regenerative amplifier FEL (RAFEL) (Nguyen *et al.*, 2000).
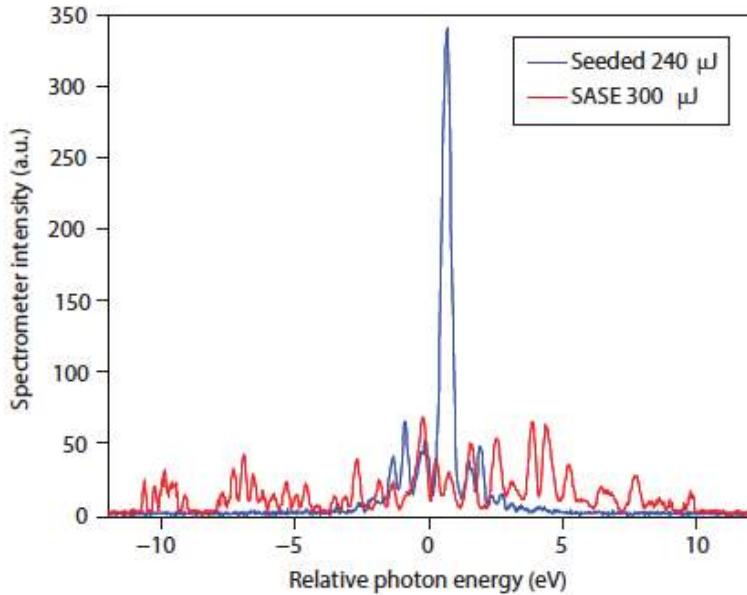

FIG. 45 Sample Spectra at LCLS for SASE, red, and self- seeded, blue, operation. From Amann *et al.* (2012).

5. *Comparison of seeding methods*

While the SASE FEL is an established and robust method to generate coherent X-rays, seeding allows for more stability in the output power, an increase in the spectral brightness and/or the synchronization with an external signal. There are three common approaches: direct seeding with HHG, induced bunching with HGHG or EEHG as well as self-seeding. Each has been successfully demonstrated though the records for the shortest wavelength was done at LCLS at a wavelength of 1.5 Åwith self-seeding.

Direct seeding has its limitations due the growing shot noise for shorter wavelengths, which requires increased power with a narrower bandwidth. HHG as the most promising source can offer attractive solutions for seeded FEL above 30 nm but it requires a significant improvement and R&D towards shorter wavelengths. It is unlikely that in the next couple of years wavelengths below 10 nm become feasible with direct seeding

HGHG made tremendous progress towards shorter wavelength down to 5 nm with a fresh bunch technique in a cascade configuration. Shorter wavelengths seem feasible, but they operate with long bunches and lower current as compared to SASE FELs at the same wavelength. The pulse energy is lower but the signal is well synchronized to an external signal. One fundamental limit is the energy

spread, which has to be smaller for short wavelength.

No apparent limitations occur in self-seeding schemes, which can be extrapolated to very short wavelength in the Ångstrom regime, assuming a sufficient filter exists to clean up the spectrum. The electron beam parameters are the same as for SASE operation and the FEL brightness increases. The drawback is that the FEL pulse is not locked to an external signal.

Equation Section 8VIII. Numerical codes

The coupled system of radiation interacting with a co-propagating electron beam within an FEL undulator exhibits many degrees of freedom. Only in an idealized and simplified model the FEL equations can be solved analytically, as we have seen in sections IV and V. A full analytical description, including 3-dimensional effects, time dependent effects, spiking, and a non-ideal electron distribution in the transverse and longitudinal phase spaces, goes beyond our capabilities.

The solution of the equations when these effects are important requires numerical methods. Simulation codes allow us to include details of the electron longitudinal and transverse particle distributions, external focusing for long undulators, effects of errors in the magnetic fields or other system components, longitudinal variation of the undulator field to enhance radiation power, diffraction effects and start-up from spontaneous radiation. Systems like the Fermi HGHG FEL (Allaria *et al*., 2013) require a rather more complicated set-up as compared to the 'simple' single undulator of a SASE FEL.

A paper by Reiche (2003) reviews FEL codes development over the years. The codes have advanced along with FEL theory, starting from a simple 1-D model to today's fully time-dependent 3-D simulations utilizing large-scale parallel computers (Reiche , 1999).

The development of numerical codes to follow the electrons from their source, through the accelerator to the undulator entrance and the combined electron beam-electromagnetic field through the undulator, "start-to-end simulations" (Borland *et al*., 2002), have played a major role in interpreting experimental results, and designing, commissioning and operating X-ray FELs. In fact, numerical codes have become an integral component in their design, construction and operation (Cornacchia, 1998; Shintake *et al*., 2001; Altarelli *et al*., 2006; Patterson *et al*, 2010).

As an example of the usefulness of numerical codes in correctly understanding experimental results and analyzing the electron beam dynamics, we show in Figure 46 a measurement of the far field angular radiation distribution in the VISA experiment (Murokh *et al*., 2003). Surprisingly the measurement gave a hollow, doughnut like distribution, far from the normally expected Gaussian. The result was reproduced numerically when the asymmetric and energy correlated horizontal phase space distribution in the electron beam, originated in the strongly nonlinear electron beam transport system, was used in the FEL simulations. Having this unusual result reproduced with GENESIS indicated a new level of insight into the dynamics of the electron beam and SASE FEL system, obtained by the synthesis of copious diagnostic measurements and rigorous, detailed simulations.
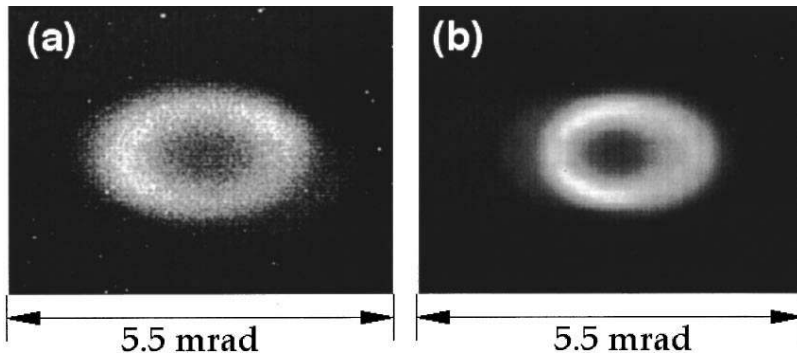
Figure 46: Far field angular distribution of the SASE radiation: (a) measured on the CCD camera and (b) simulated with GENESIS.

This section reviews the FEL simulation codes and addresses the remaining challenges in X-ray FEL modeling that will be hopefully solved in the near future.

The equations of motion for the electron beam and the field equation for the radiation, in the Slowly Varying Amplitude and Phase approximation, to be solved numerically by FEL codes, have already been given: they are (II.34) for the field, (II.23) and (II.24) for the longitudinal electron dynamics, (I.33), (I.34) for the transverse electron motion in a helical undulator or the corresponding equations, **Error! Reference source not found.**, **Error! Reference source not found.** for a planar case. In the last equations the external electron beam focusing can now be evaluated for a set of localized focusing and defocusing quadrupoles.

For a more refined model the electro-static field is added to describe the effect of space charge. It requires the solution of another non-homogenous Laplace partial differential equation.

We divide the simulation codes in two groups: steady state, when the time dependence in the field equation is neglected assuming a constant electromagnetic field profile along the electron bunch, and time dependent. The most common algorithms, used in steady state FEL simulations with macro particles, are considered first. Time-dependent simulation offers an additional complexity and is discussed later.

In the steady state model the field equation has the form of a non-homogenous Schrödinger equation. In this model we can study effects like diffraction, the impact of energy spread and emittance, the effect of external focusing. However a full description of a SASE FELs requires the more complex time dependent simulations.

The simulation codes major numerical challenges are: generating the particle distribution, advancing the macro particle and solving the field equation. Most of these are common with other simulation codes, e.g. particle-in-cell (PIC) codes. However the narrow bandwidth, Ångstrom resolution over up to 300 m of undulator length, requires some adaptation of the algorithm for the specific Free-electron Laser case.

Generating the particle distribution follows similar strategies as particle-in-cell codes. Although a random distribution would be the most straight-forward solution, a quiet loading mechanism is typically used to reduced the number of macro particles while keeping the numerical noise under a acceptable limit (e.g. charge discretization on a grid for space charge calculation).

The particle loading is done in three steps: creation of a uniform n-dimensional distribution, transformation into the desired distribution, and applying the shot noise in a controlled manner. Because the particle distribution has to resolve sub-wavelength current and energy modulation a direct import from other programs, such as ELEGANT (Borland, 2000), tracking the electrons

95

through the accelerator system, is not possible without some extensive manipulation of the external distribution. Nevertheless it is required for fully consistent start-to-end simulations of the FEL, from the cathode where electrons are generated, through the accelerator, to the undulator and the radiation generation. Most codes have adopted a strategy to import electron distributions generated in other codes into the codes evaluating the electron propagation through the undulator and the electron beam-radiation interaction.

The SASE FEL is driven by the intrinsic shot noise in the current due to the finite number of electrons per wavelength. For most cases the wavelength is much shorter than the bunch length and any correlation of the electron beam over a radiation wavelength can be neglected. The longitudinal position can be regarded as purely random. The corresponding bunching factor fluctuates from bunch slice to slice with

$$\langle B \rangle = 0, \ \langle |B|^2 \rangle = 1/N_e \ , \tag{VIII.1}$$

where N_e is the number of electrons per slice. If the shot noise is produced by a random number generator, the fluctuation in the bunching factor is given by the number of macro particles, with $N_p \leq N_e$. Therefore all shot noise algorithms remove any residual fluctuation from the quiet loading and then apply the shot noise in a controlled way.

The most common approach is to group several macro particles into a beamlet. The macro particles are evenly distributed over the FEL phase while all other dimensions having identical values. Then a small variation in the longitudinal position θ is applied, either following a uniform distribution (Penman and McNeil, 1992) or a sinusoidal modulation with

$$\theta \rightarrow \theta + 2b_j \sin(\theta + \phi_j) \ , \tag{VIII.2}$$

where b_j, ϕ_j are derived from a negative exponential and uniform distribution, respectively, for the jth beamlet. The latter method has the advantage that it can also be generalized for higher harmonics (Fawley, 2002). Alternatively the individual charge of the macro particle can be varied for the same results (McNeil et al., 2003).

The most common solver for particle tracking is the Runge-Kutta forth order solver (Kaps and Rentrop, 1979), either with fixed or adaptive step-size. It is has the advantage of being very stable and robust, but not necessarily "intelligent" to adapt to the "environment" with the most efficiency. As an example it will use almost the same computational time for drift spaces between undulator modules as in the undulator modules themselves.

Transverse motion is split into the fast oscillation of the undulator field, which couples with the radiation field, and the "slow" betatron, which is defined by the focusing properties of the undulator beam line with alternating quadrupoles and the natural focusing of the undulator itself. The betatron motion is rather secondary, because the core FEL dynamics occurs in the longitudinal phase space by energy modulation and bunching. Thus, the transverse motion can be split from the longitudinal and advanced with transport matrix to first order. It has been shown that for the FEL dynamics higher multipole components do not affect the FEL performance unless they are dialed up to an excessive level, highly unlikely for a normal FEL configuration. Using this symplectic solver for the transverse variables allows for a faster execution because the more time-consuming algorithms are only applied for the ponderomotive phase and particle energy.

Because the fast oscillation has been incorporated into coupling constants in the FEL equations the FEL problem is reduced to a slow process with the gain length as its characteristic length. The integration step width could scale with the gain length resulting in roughly the same amount of integration step for both a short, few meters, or a 100 m long undulator. However in X-ray FELs the external focusing by quadrupoles prevents large integration step size since their length is of the order of the undulator period and much shorter than the gain length. Any solver must resolve these externally imposed length scale of the undulator lattices.

Except for one-dimensional codes, where the field equation is treated like an ordinary differential equation, the field solver for a partial differential equation offers the highest numerical challenge. The continuous field has to be discretized to handle the information to describe the wave front. This is typically done by expanding the field into a set of orthonormal modes such as Gauss-Hermite or Gauss-Laguerre modes (Finite Mode Solver) or by defining the field wave front on a transverse grid (Finite Difference Solver).

The Finite Mode has the advantage that it isn't necessarily limited to a boundary condition, unlike the grid edge in a Finite Difference Solver. There can be modes which extends to infinity (Gauss - Hermite modes), but also modes, which follow a particular aperture or vacuum chamber, such as a waveguide. The strongest advantage is in connection to further transport of the radiation field through a long drift or optical cavity, making this approach most attractive for simulating FEL Oscillator configuration.

There are, however, also some drawbacks. Certain sets of modes, such as the Gauss-Hermite, are not unique. They have the complex source point as a free parameter and a given wave front can be represented by a few modes in one set but has many higher modes in another. Thus the number of modes with significant amplitude can be high if the source point is not chosen well. To avoid this problem an optimization algorithm could change the source point per integration step so that the number of modes stays low (Best and Faatz, 1990). Another drawback is the calculation of the source term. Normally there is no fast calculation, which could reuse the coupling of the electrons to a previous calculated mode in some form the current mode. The total computational effort scales with $N \times M$ where N and M are the number of macro particles and modes, respectively.

Finite Difference Solvers are more common and several advanced algorithms have been developed since the dawn of numerical physics. The basic idea is that the field is discretized at grid points and that the differential operators are substituted with difference operators. The general rules to discretize the difference operators is given by Gauss law with

$$\int_A \nabla^2 u \, dA = \oint_{\partial A} \vec{\nabla} u \cdot \vec{n} \, ds \,, \qquad\qquad \text{(VIII.3)}$$

where A is the area associated to a grid point, ∂A the edge of is grid, and $\vec{n}$ a unit vector, normal to the edge and pointing outward. For a 2D Cartesian grid the Laplace operator at the grid point with the indices i, j is

$$\nabla^2 u_{i,j} \equiv \frac{u_{i,j+1} + u_{i,j-1} + u_{i+1,j} + u_{i-1,j} - 4u_{i,j}}{h_i h_j} \,, \qquad\qquad \text{(VIII.4)}$$

where h_i, h_j are, respectively, the grid spacing in both directions. The basic idea of finite difference methods is to convert the partial differential equation into a matrix equation, where the grid point are

97

accessed in a given order, forming a vector $\vec{u} = (u_k) = (u_{i(k),j(k)})$.

In advancing the field equation numerically from position z_j to z_{j+1}, the evaluation "time" of the Laplace operation can be chosen freely between the starting and end point. It can be fully explicit before the field is advanced in z, fully implicit after the field has been advanced, or a weighted sum of both. The latter yield the general field equation, expressed in matrix form notation

$$\left[\alpha \mathbf{L} + i\frac{2k}{\Delta z}\mathbf{I} \right]\vec{u}^{\,l+1} = \left[(\alpha - 1)\mathbf{L} + i\frac{2k}{\Delta z}\mathbf{I} \right]\vec{u}^{\,l} + \vec{s}^{\,l}, \qquad \text{(VIII.5)}$$

where the upper index indicates the step along the undulator axis, $\mathbf{L}$ the matrix representation of the Laplace operator, Δz the integration step size and $\vec{s}$ the electron source term.

The fully explicit solution ($\alpha = 0$) allows us to calculate the new field values directly with minimum effort, but unfortunately is an unstable solver. Stability analysis shows that a value of at least $\alpha \geq 1/2$ has to be chosen for a stable solution. The highest stability occurs for $\alpha = 1$ which clashes with the highest precision at $\alpha = 1/2$.

The problem is solved once the inversion of the matrix $\left[\alpha \mathbf{L} + (2ik / \Delta z)\mathbf{I} \right]$ is done. While this matrix is sparse with most matrix elements being zero, the inverted matrix is not. Therefore it is not recommended to calculate the inverted matrix and then multiply with the RHS of (VIII.5). Instead the algorithms solve for $\underline{u}^{l+1}$ directly.

There are many methods to solve the matrix equation and a detailed discussion is beyond the scope of this paper. In general there can be classified as direct solver or iterative solver. Direct solvers are inverting the matrix to find the solution to Equation VIII.5. For 1-D grids (e.g. a radial discretization of the radiation field) a tri-diagonal solver (Press, 1992) avoids the memory extensive calculation of the inverted matrix but iterates line-by-line through the set of linear equations and eliminating the leading terms by scaling and subtracting the preceding equation in this linear set. Higher dimension can be solved with the Alternating Direction Implicit (ADI) methods (Ames, 1977), where for each dimension a fully implicit sub-step is done, while the rest of the dimensions is treated explicitly. This reduces the system into a set of multiple field equations with tri-diagonal shape of the Laplace operator, which is solved as described above.

The request for time-dependent simulations implies a new level of complexity. In addition to the transverse direction the radiation field and the electron beam is sampled at many longitudinal positions. The distance between sample points defines a slice. While steady-state simulations only model a single slice, assuming periodic boundary condition, time-dependent simulations can easily require ten thousands of slices or more, depending on the bunch length and the resonant wavelength. With this huge increase in the data size to model the electron bunch and radiation field the codes require an efficient memory management to overcome the limitation of computer resources in the past.

A second aspect is that the field equation is now a mixed partial differential equation with second and first order derivatives. A self-consistent field solver would be different than the steady-state solver, described previously in this paper. Second it would require the entire radiation field and electron beam to be in memory for each integration step. This can be done only with the support of a computer cluster where the memory demand is distributed over many nodes. Recently, some work has been done in this direction (Reiche, 2014).

A simpler algorithm exists if two approximations are made. The first assumes that

information can only propagate in the forward direction, namely by the slippage of the radiation field by one radiation wavelength per undulator period. The slippage over the entire undulator length is called the slippage length.

Secondly, the field equation can be split into two steps, first solving the impact of diffraction and the contribution by the source term s (emission from the electron beam) with

$$, \left[\nabla_{\perp}^2 + 2ik \frac{\partial}{\partial z} \right] u = s \,, \tag{VIII.6}$$

and the effect of the slippage with

$$\left[\frac{\partial}{\partial z} + \frac{\partial}{c \, \partial t} \right] u = 0 \,. \tag{VIII.7}$$

The latter is solved by any function with the argument $f(z - ct)$, which in the co-moving frame of the electron beam is a shift of the radiation field by one slice over an integration distance of one undulator period. The former equation is identical to the steady-state problem and the same algorithm can be used in this two-step process.

The algorithm loops over the electron bunch and undulator in discrete steps, which the undulator being the inner loop tracking on electron slice through the entire undulator length and then progressing to the next slice ahead. While this allows for the least amount of required memory it also suffers from the drawback that the current profile is fixed over the entire undulator length because the slices cannot exchange particles. In particular those who are falling back and slipping into the given slice would require the knowledge of the slice ahead.

Therefore some codes have changed the loop order with the inner loop pointing along the electron bunch. Though much more data space is required, only one radiation and electron slice has to be in memory, while the rest is temporarily stored on an external hard disk. This allows in theory to exchange particles/beamlets among slices after each integration steps. However the sorting and re-binning algorithm of GBytes of macro particles can be computationally expensive as the FEL simulation itself.

Equation Section 9 **IX. Present status of X-ray FELs**

In this section we discuss the main characteristics of X-ray FELs, operating and under construction, like pulse duration, intensity, line width, wavelength range. Another review, including results on recent improvements in the electron beam phase space density, is found in Pellegrini (2010). We also present some recent experimental work aimed at improving the FEL longitudinal coherence and generate two colors spectra. Experimental results on seeding, in its various forms, and high efficiency tapered FELs have been discussed before.

The validity of the high gain theory and start-up from noise was confirmed in experiments of the late 1990s (Hogan, 1998a; Hogan, 1998b). In the early 2000s three SASE FELs, LEUTL at Argonne national Laboratory, VUV-FEL at DESY and VISA, a UCLA-SLAC-Brookhaven National Laboratory collaboration, reached saturation (Milton *et al.*, 2001; Ayvazyan et al., 2002; Murokh *et al.*, 2003) at wavelengths between 800nm to 109nm. In 2005 the VUV-FEL at DESY –now called FLASH- lased at a wavelength of 32nm (Schreiber, 2005) and later at 4.2nm (Schreiber *et al.*, 2012). Two seeded FELs,

using the high gain harmonic generation mode, operate at Fermi@Elettra in Trieste: FEL-1, lasing in the wavelength range 65 to 29 nm (Allaria *et al.*, 2012) and FEL-2 reaching 4 nm (Svandrlik *et al.*, 2014). SACLA in Japan (Ishikawa *et al.*, 2012) is lasing at wavelengths as short as 0.6A.

The initial operation of LCLS at 1.5Å wavelength (Emma *et al.*, 2010) is an important milestone in the development of X-ray FELs. The growth of the FEL power along the undulator length in LCLS, reaching a saturation level over 20GW, is shown in Figure 47. The results are in good agreement with the simulations based on the high gain SASE theory, including 3-dimensional effects.
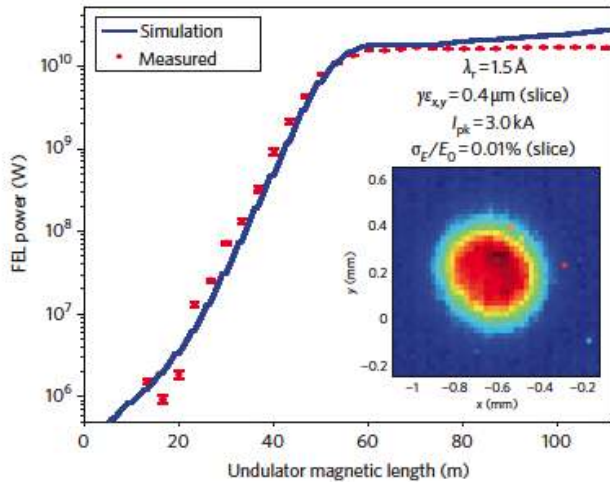

Figure 47 LCLS measurements of power versus undulator length (red points). The blue line is a simulation with the code Genesis. Error bars represent the r.m.s. statistical uncertainty in the measured power when averaging 30 beam pulses. The measured power gain length is 3.5 m. From ref. (Emma *et al.*, 2010).

A comparison of the LCLS, SACLA and FLASH brightness with that of the other existing X-ray sources is shown in FIG. 12. The increase of peak brightness by about nine orders of magnitudes in the Angstrom region obtained at LCLS, a very large jump in performance, is a remarkable achievement.

FLASH, Fermi and LCLS have demonstrated outstanding capabilities and have increased by 7 to 10 orders of magnitude the photon peak brightness, as shown in FIG. 12. The LCLS X-ray pulse duration can be changed, by varying the electron bunch charge from 250 to 20 pC, from about 100 to a few femtosecond, over the full wavelength range of 2.2 to 0.12 nm (Galayda *et al.*, 2011). This flexibility in the choice of pulse duration and intensity is very important, since it allows tailoring the X-ray pulse to the experiment one wishes to do. These X-ray FELs provide, for the first time, high intensity coherent radiation pulses at a wavelength and time scale as short as 1Å and few femtoseconds, typical of atomic phenomena, a breakthrough characteristic, making possible the scientific exploration of the structure and dynamics of atomic and molecular processes. No other electromagnetic radiation source can do this.

The main characteristics of X-ray FELs, are given in Table 1, for systems with wavelength shorter than 1nm, and Table 2, for system with longer wavelength.

The general characteristics of X-ray pulses can be summarized as follows:

1. Pulse energy hundreds of μJ to few mJ;
2. Line width in SASE mode about 10^{-3}, order of magnitude of the FEL parameter ρ; about ten times smaller than SASE when using self-seeding;
3. Pulse duration from a few to about 100 fs;
4. About 10^{3} photons/electron, at 1Å, compared to about 10^{-2} for spontaneous radiation; more at longer wavelengths.

Table 1 Characteristics of hard X-ray FEL in operation or under construction (identified by the * sign).

	LCLS	SACLA	European XFEL*	Korean X-FEL*	Swiss X-FEL*	LCLS-II Cu RF*
Electron energy, GeV	2.15-15.9	5.2-8.45	8.5-17.5	4-10	2.1-5.8	2.5 - 15
Wavelength range, nm	0.11-4.4	0.275-0.063	5.1-0.04	0.6-0.1	7-0.1	1.2 – 0.05
X-ray pulse energy, mJ	1-3 for $0.1<\lambda<1.5$	0.2-0.4 for $0.08<\lambda<0.275$	0.67-8.5 for $0.04<\lambda<5.1$	0.81-1 for $0.1<\lambda<0.6$nm	0.5-1.3 for $0.1<\lambda<7$	1-4.5 for $0.05<\lambda<04$
Pulse duration, rms, fs	5-250 for $0.1<\lambda<1.5$	4.3 for $0.08<\lambda<0.275$	1.68-107 for $0.04<\lambda<5.1$	8.6-26 for $0.1,\lambda<0.6$nm	2-20 for $0.1<\lambda<7$	5-50
Line width, rms, % SASE	0.5-0.1 for $0.1<\lambda<1.5$	0.11-0.37 for $0.08<\lambda<0.275$	0.02-0.25 for $0.04<\lambda<5.1$	0.15-0.18 for $0.1<\lambda<0.6$nm	0.06-0.4 for $0.1<\lambda<7$	0.2-0.1
Line width, rms,% seeded	0.01-0.005 for $0.1<\lambda<1.5$	0.01-0.03* for $0.08<\lambda<0.275$	0.04-0.005 for $0.04<\lambda<5.1$	0.002 -0.002 f $0.1<\lambda<0.6$nm	0.01-0.002 for $0.1<\lambda<7$	0.02

The transverse coherence of X-ray FELs is quite good as predicted by theory. The X-ray beam is nearly diffraction limited, as shown in experimental measurements using a double slit (Vartanyants *et al.*, 2011) and shown in FIG. 48. Spatial coherence is essential for applications like coherent diffraction imaging, x-ray holography and x-ray photon correlation spectroscopy. The recovery of structural information from coherent imaging experiments relies on a high degree of spatial coherence in the incident field to enable the phasing of the diffraction pattern produced by its scattering from the sample.

Presently all hard X-ray FELs operate in SASE mode. In this mode the line width is the FEL parameter ρ, of the order of 10^{-3} for all systems. The X-ray pulse is spiky, as discussed before in Section IV and shown in FIG. 21, and its width is larger than the transform limited line width, ρ/N_s, by a factor equal to the number of spikes. Single spike, transform limited FEL pulses can be obtained if the electron bunch length is shorter than the cooperation length. This condition can be realized reducing the electron bunch charge from the present level of tens to few hundreds pC to a few pC (Rosenzweig *et al.*, 2008; Reiche, Musumeci, Pellegrini and Rosenzweig, 2008; Petrillo *et al.*, 2009). Another possibility to

generate short radiation pulses is to reduce the bunch length and chirp the electron bunch energy in a magnetic compressor, obtaining a correlated energy spread much larger than the slice energy spread and generating radiation with a large frequency spectrum, much larger than the SASE width. Propagating the electron beam in a tapered undulator of a SASE FEL generates a short pulse with a single spike, as observed at the Laboratori Nazionali di Frascati SPARC test facility (Giannessi et al., 2011).

Chirping the electron bunch energy can also be used in combination with a monochromator to selects a short slice of the bunch with short time duration (Schroeder, 2002; Krinsky and Huang, 2003). Use of an emittance spoiling foil to select a short part of the electron bunch was proposed in 2004 (Emma, 2004) and successfully implemented in LCLS

Table 2 Characteristics of Soft X-ray FELs operating and under construction. LCLS-II has two undulators, SXR and HXR.					
	FLASH	Fermi FEL-1	Fermi FEL-2	LCLS-II SXR Und *	LCLS-II HXR Und *
Electron energy, GeV	0.35-1.25	1.0 - 1.5		3.6 – 4.0	3.3 – 4.0
Wavelength range, nm	52-4.2	100-20	20-4	6 – 1.0	1.2 – 0.25
X-ray pulse energy, mJ	0.2 @ λ_{Max}, 0.5 @ λ_{min}	0.3 @ λ_{Max}, 0.1 @ λ_{min}	0.1 @ 10.8 nm, 0.01 @ λ_{min}	0.9 @ λ_{max}, 0.4 @ λ_{min}	1.1 @ λ_{max}, 0.02 @ λ_{min}
Pulse duration, rms, fs	15-100 @ λ_{Max} 15-100 @λ_{min}	Depending on seed pulse duration and harmonic order, typically 40-100		6 - 50	6 - 50
Line width, rms, % SASE	0.2 @ λ_{Max} 0.15 @λ_{min}			0.1	0.2 – 0.05
Line width,rms, % seeded	$5 \times 10^{-4} - 10^{-3}$	0.06 @λ_{Max} 0.03 @ λ_{min}	0.06@10.8 nm,0.02@5.4 nm,0.04@λ_{Min}	0.02	--

Reducing the line width with self-seeding and HGHG has already been demonstrated, as we discussed in the seeding section. Much attention has been dedicated recently to the generation on two or more colors, using a variety of techniques. Two-color operation of x-ray FELs is important for: physical chemistry to extend traditional optical techniques of stimulated Raman spectroscopy to the x-ray regime; in the condensed phase, stimulated resonant inelastic x-ray scattering in solids could bring key time resolution; extending x-ray scattering techniques such as multi-wavelength anomalous diffraction (MAD) to time-resolved interactions allows phase retrieval for diffraction studies of femtosecond scale dynamically evolving molecular structures.
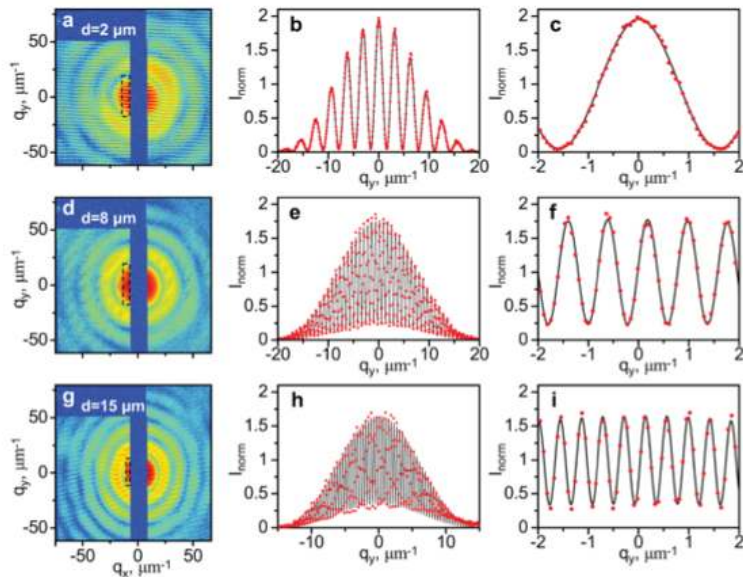
FIG. 48 LCLS measurement of transverse coherence (Vartanyants *et al*., 2011)

Five different schemes have been developed and tested. They can be used separately or in combination. One method splits the undulator in two parts, separated by an electron bunch delay element, with different values of the undulator parameter (Lutman *et al*., 2013; Hara *et al*., 2013). The FEL operates in the SASE regime, the two colors are tunable, the color separation can be large, the time delay between the two pulses is controlled by the electron bunch delay, but the pulses cannot be simultaneous.

The gain modulated FEL (Marinelli *et al*., 2013), using a periodic modulation of the undulator parameter, is another way to generate two or more spectral lines. The periodicity controls the number of lines and the gain per line. In this case different colors pulses can be practically simultaneous, and the line width is smaller than that of a SASE undulator because the modulation gives an extended cooperation length. An example of a two-color spectrum from a gain modulated SASE FEL is shown in FIG. 49.

The Fermi FEL uses a 750 fs long electron bunch seeded with two 180fs long laser pulses with different wavelengths (Allaria *et al*., 2013). The color separation is limited by the gain bandwidth and the two pulses are separated in wavelength and in time between by 300 and 700 fs.

A new scheme for the generation of ultra-short pulse trains in an FEL uses the emission from a multi-peaked electron energy distribution (Petrillo *et al*., 2013). Two electron beamlets with energy difference larger than the FEL parameter are generated by illuminating the cathode with two ps-spaced laser pulses, followed by a rotation of the longitudinal phase space by velocity bunching in the linac. The resulting SASE FEL radiation shows a double-peaked spectrum and a temporally modulated pulse structure.

A similar method was successfully developed at LCLS at X-ray energies, with the generation of two electron bunches separated in energy and time using a double laser pulse at the cathode (Marinelli et al, 2015). The final energy difference and time separation between the two bunches when they enter the undulator is determined by the initial time separation at the cathode and the compression applied using two magnetic chicanes during the acceleration. The two bunches generate different colors, and their separation can be much larger than the FEL bandwidth. The advantage of this scheme over previously

developed methods is that the X-rays can be amplified to saturation, improving the peak power by an order of magnitude with respect to other methods developed at hard X-rays. Filtering the two colors through the hard X-ray self-seeding crystal ((Amann et al., 2012; Lutman et al., 2014) gives a two-color spectrum with narrow bandwidth.
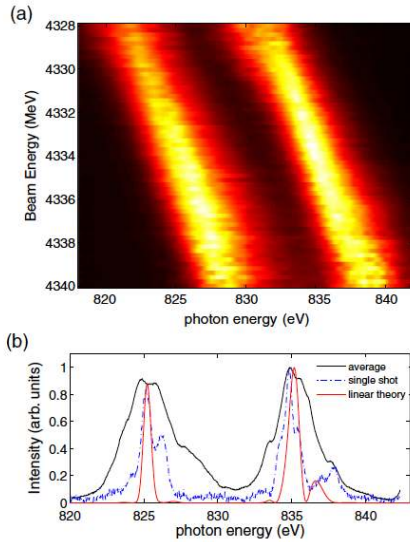

FIG. 49. A two-color spectrum of a gain modulated FEL obtained at LCLS. The figure shows the experimental data and a comparison between theory and experimental data. From ref. (Marinelli *et al*. 2013).
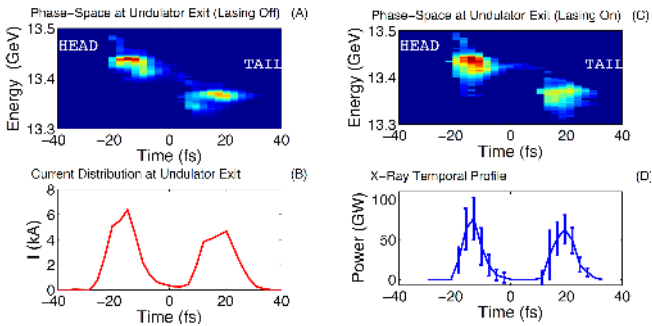

FIG. 50. Measured longitudinal phase-space of the twin electron bunches with lasing off (A) and lasing on (C). The current profile and the reconstructed X-ray temporal profile are shown in subfigures (B) and (C) respectively. From (Marinelli *et al*., 2015).

Conclusions

The theoretical picture of X-ray FELs in the classical regime, for small values of the recoil parameter, has been successfully tested in many experiments and is now well established. The numerical codes based on the theory and developed over the years can describe with good accuracy the experimental results and have been used to predict successfully new regimes, for instance two colors spectra.

One important result from the early experience with X-ray FELs is the reliability and flexibility of the system, in particular the possibility of tailoring the X-ray pulse to the requirements of a wide class of

experiments, controlling the pulse duration, its intensity and spectral properties.

The next steps in the development of X-rays FELs will raise the peak power to the TW level and reduce the pulse length below the femtosecond range, further increasing the peak brightness. Bringing together these enhanced capabilities with developments in X-ray optics, data acquisition and analysis, other active areas of developments, the X-ray FEL will continue to give new opportunities to extend our knowledge of atomic and molecular science, imaging of non-periodic systems and the evolution of systems in non-equilibrium states.

Acknowledgements

References

Ackermann, W. *et al.*, 2007, Nature Photonics **1**, 336.

Alferov D. F., Y.A. Bashmakov and E. G. Bessonov, 1974, Sov. Physics Tech. Physics **18**, 1337.

Allaria, E. *et al.*, 2012a, Nature Photonics 6, 699.

Allaria, E., *et al.*, 2012b, New Journal of Physics **14,** 113009.

Allaria, E. *et al.*, 2013 a, Nature Photonics 7, 913.

Allaria, E. *et al.*, 2013 b, Nature Communicarions 4, 2476.

Altarelli, M. *et al.* edited by, 2006, The European X-ray Free-electron Laser Technical Design Report, DESY 2006-097.

Amann, J., *et al.*, 2012, Nature Photonics **6**, 693.

Ames, W.F., 1977, *Numerical Methods for Partial Differential Equations* (Academic Press, New York)

Ayvazyan, V., et al., 2002, Physical Review Letters **88**, 104 802.

Becker W. and M.S. Zubairy, 1982, Physical Review A **25**, 2200.

Ben-Zvi, I., K. M. Yang and L. H. Yu, 1992, Nuclear Instruments Methods A **318**, 726.

Bertolotti, M., 2005, *History of the laser*, (Institute of Physics Publishing, Bristol).

Best, R.W.B., and B. Faatz, 1990, Journal of Physics D: Appl. Physics **23**, 1337.

Biedron S. G., *et al.,* 2001, in *Proceedings of the 2001 International Free Electron Laser Conference*, Darmstadt, Germany.

Bloembergen, N., 1965, *Nonlinear optics*, Ch. 2, (Benjamin, New York).

Bonifacio, R., F. Casagrande and G. Casati, 1982, Optics Communication **40**, 219.

Bonifacio, R., C. Pellegrini and L. Narducci, 1984, Optics Communication **50**, 373.

Bonifacio R. and R. Casagrande, 1985, Nuclear Instruments Methods in Physics Research A **237**, 168.

Bonifacio, R., F. Casagrande and C. Pellegrini, 1987, Optics Communication **61**, 55.

Bonifacio, R., C. Maroli , N. Piovella, 1988, Optics Communication **68**, 369.

Bonifacio R., *et al.*, 1990, Nuclear Instruments and Methods in Physics Research A **293**, 627.

Bonifacio, R., L. De Salvo, P. Pierini, N. Piovella and C. Pellegrini, 1994, Physical Review Letters **73**,

70.

Bonifacio R., 1997, Nuclear Instruments Methods in Physics Research A **400**, 165.

Bonifacio R., N. Piovella, G.R.M. Robb, 2005, Nuclear Instruments and Methods in Physics Research A **543**, 645.

Borland, M., 2000, Advanced Photon Source report LS-287, Argonne National Laboratory, Lemont, Illinois.

Borland, M. *et al*., 2002, Nuclear Instruments Methods in Physics Research A **483**, 268.

Bostedt, C., S. Boutet, Fritz, D.M., Huang, Z., Lee, H. J., Lemke, H. T., Robert, A., Schlotter, W. F., Turner, J. J., and Williams, G. J., 2015, Review of Modern Physics, in press.

Chapline G. and L. Wood, 1975, Physics Today **40**, 8.

Chin Y.H., K.-J. Kim, M. Xie, 1992, Physical Review A **46,** 6662-6683.

Coisson R., 1981, IEEE J. Quantum Electronics, **QE-17**, 1409.

Colson W.B., 1977a, *Free Electron Laser Theory*, Ph. D. Thesis (Stanford University)

Colson W. B., 1977b, Physics Letters A **64**, 190.

Colson W. B., 1981, IEEE J. Quantum Electronics **QE -17**, 1417.

Colson, W.B., 1990, in *Laser Handbook Vol. 6*, Edited by W. B. Colson, C. Pellegrini and A. Renieri, (North Holland, Amsterdam) p.115.

Cornacchia, M., 1998, Editor, Linac Coherent Light Source Conceptual Design Report, SLAC-R-593, SLAC National Accelerator Laboratory, Menlo Park, California.

Courant, E. D. and H.S. Snyder, 1958, Annals of Physics **3**, 1.

Courant E. D., C. Pellegrini, and W. Zakowicz, 1985, Physical Review A **32**, 2813.

Csonka, P., 1978, Physical Review A **13**, 405.

Dattoli, G., A. Marino, A. Renieri and F. Romanelli, 1981, IEEE Journal of Quantum Electronics **QE-17**, 1371.

Dattoli G., *et al.,* 1985, Nuclear Instruments Methods in Physics Research A **237**, 93.

Deacon, D.A.G., L. R. Elias, J. M. J. Madey, G. J. Ramian, H. A. Schwettman and T. I. Smith, 1977, Physical Review Letters **38**, 892.

De Martini, F., 1990, in *Laser Handbook, Vol. 6*, edited by W.B. Colson, C. Pellegrini, A Renieri, (North Holland, Amsterdam) p. 195.

Di Mitri, S. and Cornacchia M., 2014, Physics Rep. **539**, 1.

Doyuran, A., *et al.*, 2001, Physical Review Letters **86**, 5902.

Elias, L. R., W. M. Fairbank, J. M. J. Madey, H. A. Schwettman and T. I. Smith, 1976, Physical Review Letters **36**, 717

Ellaume, P., 1983, Journal Physics Colloques**, 44**, 333.

Elleaume, P., 1990, in *Laser Handbook Vol. 6*, edited by W. B. Colson, C. Pellegrini and A. Renieri, (North Holland, Amsterdam).

Emma, C., J. Wu and C. Pellegrini, 2014, arXiv:1403 [physics.acc-phys].

Emma, C., J. Wu, K. Fang, S. Serkez and C. Pellegrini, 2014, Physical Review Special Topics Accelerators and Beams **17**, 110701.

Emma, P. *et al*., 2010, Nature Photonics, **4**, 641.

Faatz, B. *et al*., 2009, in *Proceedings of the 2009 Free-electron Laser Conference,* Liverpool, UK.

Fawley W.M., 1995, Center for Beam Physics Tech Note-104, Lawrence Berkeley Laboratory (Berkeley, California).

Fawley, W. M., 1996, Nuclear Instruments And Methods in Physics Research A **375**, 550.

Fawley W. M., *et al*., 2002a, Nuclear Instruments and Methods in Physics Research A **483** 537.

Fawley, W.M., 2002b, Physical Review Special Topics Accelerator and Beams **5**, 070701.

Fawley, W. M. *et al.,* 2011, in *Proceedings of the 2011 FEL Conf.,* Shanghai, China, p.160.

Feldhaus, J., *et al.*, 1997, Optics Communications **140,** 341

Feng, Y., *et al.*, 2010, in *Proceedings of the 2010 FEL Conference*, Malmo, Sweden, p.270

Ferray, M., A L'Huillier, X F Li, L A Lompre, G Mainfray and C Manus, 1988, Journal of Physics B: Atomic Molecular and Optical Physics **21**, L31**.**

Fletcher, L. B., et al., 2015, Nature Photonics **9**, 274.

Fraser, J.S. and R.L. Sheffield, 1987, IEEE Journal of Quantum Electronics **QE-23**, 1489.

Galayda, J., Huang. Z., LCLS/LCLS-II team, P.A.Heimann, 2011, in *Proceedings of the 2011 Particle Accelerator Conference*, York, NY, USA, p. 724.

Galayda, J. N., 2014, in *Proceedings of the 2014 International Particle Accelerator Conference*, Dresden, Germany, p.935.

Ganter, R., 2010, edited by, *SwissFEL Conceptual Design Report*, Paul Scherrer Insitute, PSI report 10-04, Villigen, Switzerland.

Gea-Banacloche, J., G.T. Moore and M. Scully, 1984, in *Proceedings SPIE* **453**, 393.

Geloni, G., V. Kocharyan and E. Saldin, 2011, Journal of Modern Optics **58,** 16.

Geloni, G. V. Kocharyan, E. L. Saldin, 2012, in *Proceedings of the 2012 FEL Conf.,* Nara Japan, p. 29.

Geloni, G. *et al.*, 2013, Report DESY 13-013, Hamburg, Germany.

Geloni, G., V. Kocharyan, E. Saldin, 2014, arXiv:1407.4591v1[physics.acc-phys].

Gianessi, L., 2004, in *Proceedings of the 2004 FEL Conference*, Trieste, Italy.

Gianessi L. *et al.*, 2008, Nuclear Instruments and Methods in Physics Research A **594,** 132.

Giannessi *et al*., 2011, Physical Review Letters **106**, 144801.

Goodman J. W., 1985, *Statistical Optics*, (Wiley-Interscience Publishers, New York), p. 168.

Gover, A. and P. Sprangle, 1981, IEEE J. Quantum Electronics **QE-17**, 1196.

Hafizi, B., *et al*., 1990, Physical Review Letters **64**, 180.

Hecht J., 2008, The History of the X-ray Laser, Optics and Photonics News, 19 (2008).

Hemsing, E., 2008, Physics Review A **77**, 063831.

Hettel, R. and H.-D. Nuhn, 2014, private communication.

Hogan M. J. *et al*., 1998a, Physics Review Letters **80**, 289.

Hogan M. J. *et al*., 1998b, Physics Review Letters **81**, 4867.

Hopf F. A. *et al*., 1977, Physics Review Letters **37**, 1215.

Huang, Z. and K.-J. Kim, 2000, Physical Review E **62**, 7295.

Huang, Z. and K.-J. Kim, 2001, Nuclear Instruments Methods in Physics Research A **475**, 59.

Huang, Z. and K.-J. Kim, 2007, Physical Review Special Topics, Accelerators and Beams **10**, 034801**.**

Huang, Z., *et al*., 2009, Physical Review Letters, **102**, 254801.

Ishikawa T. *et al*., 2012, Nature Photonics **6**, 540.

Jackson J. D., 1998a, *Classical Electrodynamics*, 3rd edition, (John Wiley, New York), p. 663.

Jackson J. D., 1998b, *Classical Electrodynamics*, 3rd edition, (John Wiley, New York), p. 664.

Jackson J. D., 1998c, *Classical Electrodynamics*, 3rd edition, (John Wiley, New York), p. 666.

Jackson J. D., 1998d, *Classical Electrodynamics*, 3rd edition, (John Wiley, New York), p. 669.

Jackson J. D., 1998e, *Classical Electrodynamics*, 3rd edition, (John Wiley, New York), p. 676.

Jerby, E. and A. Gover, 1985, IEEE J. Quantum Electron. **QE-21**, 1041.

Jiao, Y., *et al*., 2012, Physics Review Special Topics Accelerator and Beams, **15**, 050704.

Kaps, P., and P. Rentrop, 1979, Numerische Mathematik, **33**, 55.

Kim E.-S. and Yoon M., 2009, IEEE Transaction in Nuclear Science **56,** 3597.

Kim, K.-J., 1986a, Nuclear Instruments Methods in Physics Research A **250**, 396.

Kim, K.-J., 1986b, Physical Review Letters **57**, 1871.

Kim, K.-J., 1986c, Nuclear Instruments And Methods in Physics Research A **246**, 71.

Kim, K.-J., 1996 Physical Review Letters,76 1244.

Kincaid B. M., 1997, J. Appl. Physics, **48**, 2684.

Kondradenko, A.M. and E.L. Saldin, 1980, Particle Accelerator **10**, 207.

Krinsky S., J.M. Wang and P. Luchini, 1982, Journal of Applied Physics **53**, 5453.

Krinsky, S. and L.H. Yu., 1987, Physical Review A **35**, 3406.

Krinsky S. and Z. Huang, 2003, Physical Review Special Topics Accelerator and Beams **6**, 050702.

Kroll, N.M. and W.A. McMullin, 1978, Physical Review A **17**, 300.

Kroll N. M., P. L. Morton and M.N. Rosenbluth, 1979, Physics of Quantum Electronics, **7**, 104

Kroll, N. M., P. L. Morton, and M. N. Rosenbluth, 1980, in *Free-Electron Generators of Coherent Radiation*, edited by S. F. Jacobs, H. S. Pilloff, M. Sargent, M. O. Scully, and R. Spitzer, Physics of Quantum Electronics Vol. 7 (Addison-Wesley, Reading, MA), p. 147.

Kroll N. M., P. L. Morton and M.N. Rosenbluth, 1981, IEEE Journal of Quantum Electronics **QE-17**, 1436.

Labat, M., 2011, Physical Review Letters **107**, 224801.

Lambert G., et al., 2008, Nature Physics 4, 296.

LCLS Design Study Group, 1998, Linac Coherent Light Source Design Study Report, SLAC-R-521, SLAC National Accelerator Laboratory (Menlo Park, California).

Lechner C., et al., 2012, in *Proceedings of 2012 FEL Conference*, Nara, Japan.

Luccio, A., and C. Pellegrini, 1980, in Free Electron Lasers, edited by S. Martellucci and A. N. Chester, E. Majorana International Science Series (Plenum Press, New York) p. 243.

Lutman, A.A. *et al*., 2013, Physical Review Letters 110, 134801.

Lutman, A.A. *et al.*, 2014, Physical Review Letters 113, 254801.

Madey, J.M.J. 1971, J. Appl. Physics 42, 1906.

Madey M. J. M., 1979, Nuovo Cimento B 50, 64.

Marinelli, A., A. Lutman, J. Wu, Y. Ding, J. Krzywinski, H.-D. Nuhn, Y. Feng, R. N. Coffee and C. Pellegrini, 2013, Physical Review Letters **111**, 134801.

Marinelli, A., *et al*, 2015, Nature Communications DOI: 10.1038/ncomms7369.

Matthews D.L. *et al*., 1985, Physical Review Letters **54**, 110.

McNeil, B.W.J., *et al.*, 2003, Physical Review Special Topics Accelerator and Beams **6**, 070701.

McNeil B.W. J., *et al*., 2006, Physical Review Letters **96**, 084801.

Milton S. V., *et al*., 2001, Science, **292**, 2037.

Moore, G.T., 1984, Optics Communication **52**, 46.

Moore, G.T., 1985, Nuclear Instruments Methods in Physics A **239**, 19.

Morse, P., H. Feshbach, 1953, *Methods of Theoretical Physics*, (McGraw-Hill, New York) p.1108.

Motz H., 1951, Journal of Applied Physics **22**, 527.

Motz, H., 1953, Journal of Applied Physics **24**, 826.

Murokh, A. *et al.*, 2003, Physical Review E **67**, 066501.

Murphy J. B. and C. Pellegrini, 1985, Journal of the Optical Society of America B **2**, 259.

Murphy J. B., C. Pellegrini and R. Bonifacio, 1985, Optics Communication **53**, 197.

Murphy J.B. and C. Pellegrini, 1990, in *Laser Handbook, Vol. VI*, Edited by W.B. Colson, C. Pellegrini, A Renieri, (North Holland, Amsterdam) p. 9.

Musumeci P. *et al*., 2005, Physical Review Letters **94**, 154801.

Neal, R.B. (Editor), 1967, *The Stanford Two Mile Accelerator,* W.A. Benjamin Inc., New York. The book has been digitized and can be found at http://www.slac.stanford.edu/library/2MileAccelerator/2mile.htm

Nguyen, D. C., *et al.*, 2000, in *Proceedings of the 2000 Linac Conference*, Monterey, California.

Orzechowski, T. *et al*., 1985, Physical Review Letters **54**, 889.

Palmer, R.V., 1972, Journal of Applied Physics **43**, 3014.

Palmer R. V., 1981, IEEE Trans. Nuclear Sci. **28**, 3370.

Pantell, R.H., G. Soncini and H.E. Puthoff, 1968, IEEE Journal of Quantum Electronics **QE-4**, 905.

Patterson, B.D., *et al.*, 2010, New Journal of Physics **12,** 035012.

Pellegrini, C., 1988, Nuclear Instruments and Methods in Physics Research **A272,** 364.

Pellegrini, C., 1992, in *Proceedings of the Workshop on 4th Generation Light Sources,* edited by M. Cornacchia and H. Winick, Stanford Synchrotron Radiation Laboratory 92-02, p. 341.

Pellegrini, C., 2010, Review of Accelerator Science and Technology **3**,185.

Pellegrini, C., 2012, European Physics Journal H. **37**, 609.

Petrillo, V., I. Boscolo, A.Bacci, M. Boscolo, M. Ferrario, M. Serluca, L. Giannessi, C. Ronsivalle, 2010, Nuclear Instruments and Methods in Physics Research A **621**, 1.

Petrillo, V., *et al*., 2013, Physical Review Letters **111**, 114802.

Preiss, P. *et al*., 2012, in *Proceedings of the 2012 FEL Conference,* Nara Japan, p. 93.

Press, W.H., 1992, *Numerical Recipes in Fortran 77* (Cambridge University Press, Cambridge).

D. Ratner *et al*., 2009, in *Proceedings of the FEL 2009 Conference*, Liverpool, UK, p. 221.

Ratner, D., *et al,* 2011, Physical Review Special Topics Accelerator and Beams, **14**, 060701.

Ratner, D. *et al.*, 2015, Physical Review Letters 114, 054801.

Reiche S., 1999, Nuclear Instruments and Methods in Physics Research A **429**, 243.

Reiche, S., 2000, Nuclear Instruments and Methods in Physics Research A **445**, 90.

Reiche, S., 2003, Computation of FEL processes, in *Proceedings 2003 Particle Accelerator Conference*, p. 203.

Reiche S., P. Musumeci, C. Pellegrini, J.B. Rosenzweig, 2008, Nuclear Instruments and Methods in Physics Research A **593,** 45.

Reiche, S., E. Prat, 2012, in *Proceedings of 2012 FEL Conference*, Nara, Japan, p. 25..

Reiche, S., 2014, in *Proceedings of 2014 FEL Conference*, Basel, Switzerland, p. 403.

Robinson, K.W., 1985, Nuclear Instruments Methods Physics Research A **239**, 111.

Rosenzweig, J. B, 2003, *Fundamental of Beam Physics*, Oxford University Press, p. 52 and following.

Rosenzweig, J. B. *et al*., 2008, Nuclear Instruments and Methods in Physics Research A **593,** 39.

Saldin E. L., E.A. Schneidmiller, M.V. Yurkov, 1993, Optics Communication **97,** 272.

Saldin E.L., E.A. Schneidmiller, M.V. Yurkov, 1998, Nuclear Instruments and Methods in Physics Research A **407,** 291.

Saldin, E. L., *et al*., 2003, Nuclear Instruments and Methods Physics Research A **507**, 106.

Saldin, E. L., E. A. Schneidmiller, M. V. Yurkov, 2008, Optics Communication **281** 1179.

Sands, M., 1955, Physical Review **97,** 470.

Sasaki, S., K. Miyata and T. Takada, 2002, Journal of Applied Physics **31**, 1794.

Scharlemann, E.T., A.M. Sessler and J.S. Wurtele, 1985, Physical Review Letters **54**, 1925.

Scharlemann, E.T., 1990, in *Laser Handbook, Vol. 6*, edited by W. B. Colson, C. Pellegrini and A. Renieri, (North Holland, Amsterdam) p.291.

Schneidmiller E. A. and M.V. Yurkov, 2012, Physical Review Special Topics Accelerator and Beams **15**, 080702.

Schreiber, S. *et al*., 2005, in *Proceedings of the 2005 FEL Conference*, Stanford, California, p. 12.

Schreiber, S., B. Faatz, J. Feldhaus, K. Honkavaara, R. Treusch, M. Vogt, 2012, in *Proceedings of the 2012 FEL Conference*, Nara, Japan, p. 37.

Schroeder, C. B., C. Pellegrini, and P. Chen, 2001, Physical Review E **64**, 056502.

Schroeder, C. B., C. Pellegrini, S. Reiche, J. Arthur and P. Emma, 2002, Journal of the Optical Society of America **19**, 1782

Sessler, A. M., D. H. Whittum and L. H. Yu, 1992, Physical Review Letters **68**, 309.

Shintake, T., *et al*., 2001, SPring-8 Compact SASE Source, SPIE, San Diego, California.

Siegman, A.E., 1986, *Lasers* (University Science Books, Sausalito, California).

Smith, T. I. *et al.,* 1979, Journal of Applied Physics, **50**, 4580.

Smythe W.R.,1950, *Static and Dynamic Electricity* (McGraw-Hill, New York), p. 277.

Sprangle, P., C.M. Tang and W./M. Manheimer, 1979, Physical Review Letters 43, 1932.

Sprangle, P. and R.A. Smith, 1980, Physical Review A **21**, 293.

Sprangle, P., C.M. Tang and C.W. Roberson, 1985, Nuclear Instruments Methods in Physics Research A **239**, 1.

Sprangle, P., A. Ting, and C. M. Tang, 1987, Physical Review Letters **59**, 202.

Stupakov, G., 2009, Physical Review Letters **102,** 074801.

Stupakov, G., 2011, in *Proceedings of 2011 FEL Conference,* Shanghai, China.

Suckewer S. *et al*., 1985, Physical Review Letters **55**, 1753.

Suckewer S. and P. Jaeglé, 2009, Laser Physics Letters **1**, 411.

Svandrlik, M. *et al.,* 2014, in *Proceedings of IPAC 2014*, Dresden, Germany, p. 2885.

Thompson, N. R. and B. W. J. McNeil, 2008, Physical Review Letters **100**, 203901.

Tremaine, A. et al., 1998, Physical Review Lett*ers* **81***, 5816.*

Tremaine, A., *et al*., 2002, Physical Review Letters, **88**, 204801.

Van Steenbergen A. *et al*., 1996, Physical Review Letters **77**, 2690.

Vartanyants, I.A., *et al*., 2011, Physical Review Letters **107**, 144801.

Vinokurov N. A. and A. N. Skrinsky, 1982, in *Relativistic Microwave Electronics,* Gorkyi, Russia.

Wang, J.-M. and L.-H. Yu, 1986, Nuclear Instruments Methods in Physics Research A **250**, 484.

Wang, X. J. *et al*., 2009, Physical Review Letters **103**, 154801.

Watson G.N., 1987, *A treatise on the Theory of Bessel Functions*, (Cambridge University Press, Frankfurt/Main).

Webb, S. *et al*., 2011, Physical Review Special Topics Accelerator and Beams **14**, 051003.

Wu J. *et al*., 2010, in *Proceedings of the 2010 FEL Conference,* Malmo, Sweden, p. 147.

Wu J., A. Marinelli and C. Pellegrini, 2012, in *Proceedings of the 2012 FEL Conference,* Nara, Japan, p. 237.

Xiang, D. and G. Stupakov, 2009, Physical Review Special Topics Accelerator and Beams **12**, 030702.

Xiang, D., *et al.*, 2010, Physical Review Letters **105**, 114801.

Xiang D., Y. Ding, Z. Huang, H. Deng, 2013a, Physical Review Special Topics Accelerators and Beams **15**, 080702.

Xiang, D. *et al.*, 2013b, Physical Review Special Topics Accelerator and Beams **16**, 010703.

Xie, M., 1996, in *IEEE Proceedings of 1995 Particle Accelerator Conference*, Dallas, Texas, p. 183.

Xie, M., 2001, Nuclear Instruments Methods in Physics Research A **475**, 51.

Yu, L.-H., S. Krinsky and R. Gluckstern, 1990, Physical Review Letters **64**, 3011.

Yu, L.-H. *et al.*, 2000, Science **289,** 932.

Zinth, W., A. Laubereau and W. Kaiser, 2011, European Physics Journal H **36**, 153.