FEL SIMULATIONS: HISTORY, STATUS AND OUTLOOK

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

Abstract

The coupled system of radiation interacting with a co-propagating electron beam within an undulator of an FEL exhibits many degrees of freedom. Only in an idealized and simplified model can the FEL equations be solved analytically and a more complete description requires numerical methods. Therefore numerical codes have been developed along with the advances in FEL theory, starting from a simple 1 D model to today's fully time-dependent 3D simulations, utilizing large scale parallel computers. This presentation gives a brief history of FEL simulation and addresses the remaining challenges in FEL modeling which we hope to solve in the near future.

INTRODUCTION

Numerical codes have become an integral component to study and design Free-electron Lasers (e.g. [1] – [4]) worldwide. The underlying theory [5, 6] is rather complex and allows analytical solution only under approximations. The level of complexity increases with a more refined and realistic model of the FEL which include effects such as electron beam misalignment, undulator field errors, and start-up from spontaneous radiation. In addition new concepts (e.g. cascading Free-Electron Lasers [7]) are studied, which are following a rather inhomogeneous set-up as compared to the 'simple' single undulator of a SASE FEL.

The paper gives a brief overview on the FEL model and the most common approaches to solve it. It also discusses some problems in FEL dynamics, which cannot be solved with the current codes to a satisfying level, and how future development can remove these limitations.

THE FEL MODEL AND ITS LIMITATIONS

This section presents the equations of motion for the electron beam and the field equation for the radiation, which are solved numerically by FEL codes. It has to be noted that there is a slight variations in the FEL model from code to code, depending on the underlying assumption or level of details. However they have in common that they solve a coupled set of ordinary differential equations for the electrons and a partial differential equation for the radiation field.

* sven.reiche@psi.ch

Electron Motion

The motion of the electrons are dominated by the periodic field of the undulator, enforcing a sinusoidal oscillation. For a planar undulator the transverse, normalized velocity β_x of this oscillation is given by

$$\beta_x = \frac{K}{\gamma}\sin(k_u z) \quad , \tag{1}$$

where γ is the Lorenz factor of the electron energy, $K = eB_y\lambda_u/2\pi mc$ the normalized deflection strength of the undulator peak field B_y per period λ_u, and k_u the undulator wavenumber. This motion causes also a longitudinal oscillation with $\beta_z = \sqrt{1-\beta_x^2} \approx 1 - \beta_x^2/2$.

When overlapping with a co-propagating radiation field with the amplitude E, phase ϕ, and the wavenumber k the electrons is exchanging energy with the field according to

$$\frac{d}{dz}\gamma = \frac{KeEe^{i\phi}}{mc^2}\left(e^{i\theta} + e^{i\theta-2ik_u z}\right) \tag{2}$$

with the definition of the ponderomotive phase $\theta = (k + k_u)z - ckt$. It is convenient to express the longitudinal position in terms of this ponderomotive phase and the longitudinal 'velocity' becomes

$$\frac{d}{dz}\theta = k_u + k\left(1 - \frac{1}{\beta_z}\right). \tag{3}$$

Eqs. 2 and 3 describe completely the longitudinal dynamics of the electron motion, interacting with a radiation field in an undulator. Note that the longitudinal velocity β_z depends implicitly on the electron energy, the fast oscillation of Eq. 1 as well as an additional 'slow' betatron oscillation, which has a much longer period and thus can be assumed constant over a single undulator period.

The transverse motion is given by

$$\frac{d}{dz}\beta_x = \frac{p_x}{\gamma mc} \tag{4}$$

$$\frac{d}{dz}\beta_y = \frac{p_y}{\gamma mc} \tag{5}$$

$$\frac{d}{dz}p_x = -mc\frac{K}{k_u}\cos(k_u z) + F_x(z) \tag{6}$$

$$\frac{d}{dz}p_y = F_y(z) \tag{7}$$

$$\tag{8}$$

where $p_{x,y}$ are the transverse momenta and $F_{x,y}$ some external focusing (e.g. by quadrupoles).

Field Equation

Because the radiation field needs to have a good overlap with the electron beam, its preferred direction of propagation is along the undulator axis. This allows to separate the dominant oscillation of the field with $\exp(ik(z-ct))$ from the field equation and to express the field evolution as a partial differential equation of the complex field amplitude $u = E_x e^{i\phi}$ with

$$\left[\nabla_\perp^2 + 2ik\left(\frac{\partial}{\partial z} + \frac{\partial}{c\partial t}\right)\right] u = \tag{9}$$

$$iec^2\mu_0 k \sum_j \delta(\vec{r} - \vec{r}_j)\frac{K}{\gamma}\left(e^{-i\theta} + e^{-i\theta+2ik_u z}\right)$$

This approximation is called the "Slow Varying Envelope Approximation" (SVEA), which is also used for the propagation of laser fields in free space [8].

Eikonal Approximation and Limitation

The coupled Eqs. 2 - 9 are called the FEL equations. For a more refined model the electro-static field is added to describe the effect of space charge. It requires to solve another partial differential equation of type of an inhomogenious Laplace equation. The effective electric field is then added to Eq. 2.

All equations of the FEL model oscillate with the undulator period (except for the transverse motion in y) and a solver must resolve this oscillation. This imposes a limitation on the maximum possible integration step size Δz, which has to be much smaller than the undulator period λ_u. However the FEL amplification is a resonant process and requires several undulator periods before a change in the electron motion and radiation field amplitude becomes noticeable. This is the reasoning to average the FEL equations of motion over the distance of an undulator period to eliminate all fast oscillating terms. Only the resonant terms remain and the integration step size can be increased to a fraction of the characteristic scale of the FEL process, the FEL gain length [9].

Because the unaveraged ponderotive phase as a longitudinal oscillation, twice as fast as the transverse oscillation, the electron can couple to many frequencies in the two exponential terms of Eq. 2 and 9. Collecting all terms which are proportional to the fundamental frequency the effective coupling of the electron beam is given by $JJ_0 = J_0(\xi) - J_1(\xi)]$ with $\xi = kK^2/8\gamma^2 k_u$. The average equations are obtained when in Eqs. 2 - 9 the undulator parameter K is replaced with JJ_0K and all terms with an explicit dependence on z (except for the external focusing) are dropped.

This approximation defines an eikonal system of equations [10] and is the basis for most FEL codes. The limitations are of course that the model becomes blind to anything which occurs on the scale of the undulator period. It also restricts the spectral resolution to a bandwidth around the fundamental frequency $\omega = k/c$.

An even further simplification is that no temporal variation is assumed in the electron beam and radiation field amplitude. This is the steady-state model because the time derivative drops out in Eq. 9. The field equation has the form of an inhomogenious Schrodinger equation. In this model, effects such as diffraction and impact of energy spread, emittance and focusing can be studied. However it is insufficient to model SASE FELs.

NUMERICAL METHODS

In this section, a brief description is given over the most common algorithms, used in steady state FEL simulations with macro particles. Time-dependent simulation offers an additional complexity and thus has its own chapter devoted to the problem of finite length electron bunches. The 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 length of the undulator requires some adaptation of the algorithm for the specific case of the Free-electron Laser.

Particle Distribution

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 tracking programs such as ELEGANT [11] is not possible without some extensive manipulation of the external distribution. Nevertheless it is required for a fully consistent start-end simulation of the FEL, and most of the codes have adopted a strategy to import external distributions.

Quiet Loading of a Uniform Distribution

Because the internal labeling of the particles has no relevance for the simulation, sequence elements, can be freely reordered. Guaranteeing no correlation between succeeding sequence elements, such as the random number generator, is not needed. Instead codes are using quasi-random sequences [12], which are highly correlated within its own elements but avoids any correlation with other sequences. The most popular sequence is the Halton sequence [13], which is a generalization of the bit-reversal technique for an arbitrary n-base system to represent numbers, such as $n = 2$ for the dual system, $n = 8$ for the octal system or $n = 10$ for the decimal system. In the Halton sequence,

a running index is encoded in a representation of the base n. Then the order of the digits are reversed and the fraction point is put in front. Tab. 1 shows the first elements for the base $n = 5$ as an example.

Table 1: Halton Sequence h_5 for the base $n = 5$.

Index	Index (5-base)	Fractional (5-base)	h_5
1	1	0.1	0.2
2	2	0.2	0.4
3	3	0.3	0.6
4	4	0.4	0.8
5	10	0.01	0.04
6	11	0.11	0.24
7	12	0.21	0.44
8	13	0.31	0.64
9	14	0.41	0.84
10	20	0.02	0.08
...	...	...	...

The power of the Halton sequence is that the least significant digit is converted to the most significant digit due to the reversal in the order of the digits. Thus the sequence provides rapidly varying elements within the open interval of zero and one and fills out the uniform distribution rather evenly and fast.

To avoid correlation with other sequences, the bases should be chosen differently. In particular it is recommended to use prime numbers for different sequences to avoid accidental correlation between sequences, which shares common factors.

Transformation of a Uniform Distribution

Once an n-dimensional uniform distribution is generated it is transformed to the desired distribution. Most common are Gaussian distributions, but parabolic and Lorenz distribution are used as well. There are several methods, which can be applied. The most direct method is to feed the uniform distribution into the integral of the inverted probability function or by rejection method for some specific functions [14]. Another method is the joint probability distributions [15]. Gaussian distributions are typically generated with the last method. However, it does not preserve the quiet loading of the Halton sequence [16]. It is recommended to use the inverted probability distribution whenever possible, even if requires more computational efforts.

Shot Noise

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 $b = (1/N_e)\sum \exp(i\theta)$ fluctuates for from slice to slice with

$$< b >= 0 \tag{10}$$

and

$$< |b|^2 >= \frac{1}{N_e} \tag{11}$$

with N_e the number of electrons per slice. If a random generator is used to generate the shot noise 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 general idea is to allow an additional degrees of freedom, which represent the bunching of a group of electrons when represented by a macro particle in the simulations. It can be either the value of the bunching factor directly or by adding a dipole component [17], defined by the amplitude and distance of two opposite charge particles.

The most common approach though is to group several macro particles into a beamlet. The macro particle are evenly distributed over the ponderomotive phase while all other dimensions have identical values. Then a small variation in the longitudinal position θ is applied, either following a uniform distribution [18] or a sinusoidal modulation with

$$\theta \rightarrow \theta + 2b_j \sin(\theta + \phi_j) \quad , \tag{12}$$

where b_j and 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 [19]. Alternatively the individual charge of the macro particle can be varied for the same results [20].

Beamlets have the advantage that once they have been generated, the rest of the simulation does not require the distinction between beamlets and the macro particles in a given beamlets. Thus the grouping into beamlets can be dropped. On the other hand, the number of macro particles per beamlet limits the maximum harmonic due to the mirroring process in the quiet loading for the ponderomotive phase. As an examples with 1 macro particles and 3 mirrored particles per beamlet, the fundamental and second harmonics are suppressed by the even spacing between these macro particles. However the evaluation of the bunching factor at the 4th harmonic would add up all macro particles coherently in phase even before applying the shotnoise modulation. Therefore these simulations require at least twice the number of macro particles per beamlet than the harmonics to be resolved. This is still a problem for large HGHG cascades where the final harmonic number can be of the order of hundred or higher.

With the methods briefly described above, the shot noise value per beamlet is static and preserved over drifts. However, recent theories have shown that this "static" modeling might be not sufficient [21] for some schemes to experimentally control the shotnoise, where in addition a energy modulation per beamlet is require to allow for the bunching

factor to grow or decay over distance while still preserving the shot noise statistic at any position.

Particle Tracker

The most common solver to be found is the Runge-Kutta solver [22] in forth order, 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 phasespace 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 does not affect the FEL performance unless they are dialed up to an excessive level, which are highly unlikely for a normal FEL configuration. Using this symplectic solver for the transverse variables allows for the a faster execution because the more time-consuming algorithm 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. This would allow for more advance solver which optimize the integration step width with respect to the numerical precision, e.g. in drift section with no coupling to the radiation field, the integration step size could be increase to quickly advance to the next undulator section. Algorithms like the Gear-predictor [23] or the Burlisch-Stoer [24] methods are self-optimizing solver, which have the maximum efficiency for smooth systems with slow changes in the differential equations. However in X-ray FELs, where the computational efficiency is desired the most, the external focusing by quadrupoles prevents a fast calculation because these solvers are trying to resolve the physical length of the quadrupoles, which are typically of the order of a few undulator periods and thus much shorter than the gain length. As a result the average integration step size would be smaller than theoretically possible for the exponential growth of the FEL and the advantage over the more robust solvers, such as the Runge-Kutta, dimishes with little motivation for the developer to implement these algorithm in the codes.

Field Solver

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 wavefront. 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 wavefront 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.

However there are 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 wavefront 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 [25]. 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 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 operator are substituted with difference operators. The general rules to discretize the difference operators is given by Gauss law with

$$\int_A \nabla^2 u dA = \oint_{\partial A} \vec{\nabla} u \cdot \vec{n} ds \tag{13}$$

where A is the area associated to a grid point, ∂A the edge of is grid, and $\vec{n}$ a unit vector, which is normal to the edge and points outward. For a 2D Cartesian grid the Laplace operator at the grid point with the indices i and j is

$$\nabla^2 u_{i,j} \equiv \frac{u_{i,j+1} + u_{i,j-1} + u_{i+1,j} + u_{i-1,j} - 4u_{i,j}}{h_i h_j}, \tag{14}$$

where h_i and h_j are the grid spacing in the both direction, respectively. The basic idea of finite difference methods is to convert the partial differential equation into a matrix equation, where the grid point are access in a given order, forming a vector with $\vec{u} = (u_k) = (u_{i(k),j(k)})$.

One degree of freedom is, where within the integration step the transverse Laplace operator is evaluated. 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:

$$\left[\alpha\mathbf{L} + i\frac{2k}{\Delta z}\mathbf{I}\right]\vec{u}^{l+1} = \left[(\alpha-1)\mathbf{L} + i\frac{2k}{\Delta z}\mathbf{I}\right]\vec{u}^{l} + \vec{s}^{l} \quad (15)$$

where the upper index l 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 source term by the electrons.

The fully explicit solution ($\alpha = 0$) allows 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 [26]. 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 $[\alpha\mathbf{L} + 2ik/\Delta z\mathbf{I}]$ 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 to the RHS of Eq. 15. Instead the algorithms solve for $\vec{u}^{l+1}$ directly.

There are many methods to solve the matrix equation and a detailed discussion is beyond the scope of this book. In general there can be classified as direct solver or iterative solver. Direct solver are finding the exact solution. If the matrix is of tridiagonal shape [27], with non zero elements only in the center diagonal and the diagonals above and below, the system can be solved by successively scaling a line and subtract it from the line below to eliminate elements till the last line is reached with a simple identity for one field grid point. Then the system is rolled back, inserting the now known values for the grid points. A tridiagonal system occurs for a single dimension in the transverse plane (e.g. radial grid). Higher dimension can be solver with the Alternating Direction Implicit (ADI) methods [28], where for each dimension a fully implicit sub-step is done, while the rest of the dimensions is treated explicit. Then a tridiagonal shape of the Laplace operator is enforced and can be solved.

Iterative solver are trying to find a solution which isn't exact but within a reasonable error margin. The advantage is that a single iteration is reasonable simple and fast and thus allows for many iterations for the same computational effort as a direct solver. For convergence to the correct solution all recursion methods require diagonal dominance of the matrix equation to be solved [29], which is trivially fulfilled for the FEL field equation. The iteration process can be regarded as an averaging/relaxation process over neighbor field point, where after a few iteration the error has been averaged out to a negligible value. This averaging process can be optimized by a careful ordering of the grid points, enhancing the correction per iteration (successive over-relaxation [30]) and the extrapolation to multiple coarser grids (Multigrid Methods [31]). A detailed description would be beyond the scope of this paper.

HARMONICS

While the period-average FEL equations allows for large integration step sizes it has its drawbacks, which are mainly the suppression of harmonics in the par-axial equation for the field evolution. In the rest frame, defined by the "effective" particle energy $\gamma_z = \gamma/\sqrt{1+K^2/2}$ the electron is performing a figure-eight motion, which allows the emission of odd harmonics in addition to the fundamental wavelength. The coupling to higher harmonics becomes more pronounced with larger values of $K \gg 1$.

There are also additional mechanisms how the electron beam can couple to harmonics, including even ones [32]: Variation of the radiation field over the finite extension of the wiggling amplitude, "slow" transverse motion due to the betatron oscillation and due to the transverse gradient in the charge distribution of the electron beam.

In the period-average equations, the effective coupling has to be calculated for each harmonics with a Fourier analysis of the electron motion. As an example for a co-propagating wave, the longitudinal oscillation of the figure-8 motion prevents the electron to stay synchronized with the radiation field. This is expressed by a reduced coupling to the field, depending on the harmonics with

$$JJ_n = (-1)^{\frac{n-1}{2}}[J_{\frac{n-1}{2}}(\xi) - J_{\frac{n+1}{2}}(\xi)] \quad (16)$$

with $\xi = nka_w^2/4\gamma^2 k_u$. Similar terms arise for the coupling through the betatron motion, though it scales in addition to the transverse divergence of the electron. Noticeable coupling arises from a misaligned electron beam as well in the typical FODO lattice configuration of the undulator for soft and hard X-ray FELs.

In the simulation the harmonics are treated as independent radiation fields. Depending on the sample rate of the electron beam (in the ideal case it is once every wavelength) the absolute bandwidth of each harmonics is the same, but on the relative scale it shrinks with higher harmonics. This causes the problem that the energy acceptance in the electron beam detuning is more stringent for harmonics than the fundamental in order to keep the resonant wavelength within the bandwidth. This effect is even enhance if a low sampling rate is chosen to keep the amount of samples within a reasonable limit in the case of X-ray FELs at around 1 Ångstrom.

For SASE FELS simulations it is a good approximation that the self-consistent interaction of the harmonics with the electron beam can be neglected once the growth in the bunching factor is driven by the non-linear dynamics of the fundamental [33]. This allows for a faster execution because only the fundamental field is used for advancing the electrons and only the paraxial equation for the fundamental and the harmonic under consideration is solved. It also keeps the memory demand, to hold the radiation field, on a manageable level.

TIME-DEPENDENT SIMULATIONS

The request for time-dependent simulations implies a new level of complexity. In the 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 over come 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 initiatives have been done following this approach. They are described briefly in the last section of the paper.

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 with first solving the impact of diffraction and the contribution by the source term s (emission from the electron beam) with

$$\left[\vec{\nabla}_\perp^2 + 2ik\frac{\partial}{\partial z}\right] u = s \tag{17}$$

and the effect of slippage with

$$\left[\frac{\partial}{\partial z} + \frac{\partial}{c\partial t}\right] u = 0 \quad . \tag{18}$$

The latter is solved by any function with the argument $f(z - ct)$, which is in the co-moving frame of the electron beam 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.

With this assumption the problem can be solved by progressing through the electron bunch from the end to the head. Because the last slice has either no seeding field (SASE FEL) or a well defined, external field (FEL Amplifier), the electron slice can be tracked through the undulator. At each step the field is updated due to the interaction with the electron slice and than advanced forward, temporarily stored till it is feed to the next electron slice after the previous slice has been tracked through the entire undulator. It can easily be seen that the radiation field over the full slippage length and a single electron slice needs to be stored at anytime, which is typically less than the full FEL pulse length for VUV and X-ray FELs.

The resulting algorithm, which loops through the electron bunch slice by slice, is valid as long as some constraints are fulfilled:

1. The difference in the diffraction over the distances λ_u and $\lambda_u + \lambda$ is negligible.
2. The integration step size is much shorter than the gain length to avoid 'FEL amplifier' effects.
3. The radiation field has at least a small degree of longitudinal coherence, so that the periodic boundary condition in the steady-state solver are valid.

The first condition is typically fulfilled for most FELs. The second requires that the radiation field is passed through many slices, so that it cannot act upon itself with a single slice. Otherwise numerical artifacts such as detuning in a chain of 'mini' FEL amplifiers become dominant. The last condition is invalid for the start-up process of a SASE FEL with no longitudinal coherence. However there are 'hot spots', which emits on a higher level than in average. This emission occurs almost unchanged over a fraction of a gain length before the FEL amplification becomes noticeable. At that point longitudinal coherence has been sufficiently build up in these hot spots. The period boundary condition of the radiation field over a short distance has become valid.

This algorithm loops over the electron bunch and undulator in discrete steps, which the undulator being the inner loop. 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 rebinning algorithm of GBytes of macro particles can be computationally expensive as the FEL simulation itself.

NUMERICAL CHALLENGES

While existing FEL codes had tremendous success to reproduce results from FEL experiments and have been benchmarked well, there are some novel ideas for FELs, which will be hard to model due to the limitation in the core algorithm. This chapter lists a few examples, which offer a numerical challenge for the existing FEL codes and most likely become the driving force to extend the codes or to develop new ones (see next section).

High Current Electron Beams

Electron beams, generated by the injection of a femtosecond laser pulse in a plasma, have a length on the micron level with large peak currents of above 10 kA [34]. Despite energies of up to 1 GeV the space charge is strong, which causes a growth in the energy spread and elongation of the electron pulse. This is even enhanced due to the dispersion in the undulator and the effective reduction of the electron energy with $\gamma_z = \gamma/\sqrt{1+K^2/2}$.

The numerical problems is twofold. One is the self-consistent calculation of the space charge field, which has a significant impact on the electron dynamics. Normally FEL codes include only the space charge field on the resonant wavelength because it determines the work to be done by the FEL process to push electrons together to micro bunches against the electro-static field. For the 'long range' space charge field a self-consistent solver has to be adapted by the codes.

The second problem is the strong change in the longitudinal position of the macro particles, in particular in the head and tail of the bunch. This implies a mixing of macro particles of different slices. A sorting procedure transfers macro particles, which are outside of the electron slice, to the corresponding new slice. Besides the penalty of sorting and redistributing electrons, the charge per macro particle has to be the same for all slices. Although this seems natural from a Particle-in-Cell code, enforcing the same charge can result in unbalanced load of nodes a the parallel version of the code.

Short Pulse Operation and Energy Chirp

The FEL equations are derived under the assumption of a slow varying amplitude. That allows to expand the radiation field around a resonant wavelength and thus removing any fast oscillation terms in the field equation. The field and the electron bunching is sampled once every wavelength and assumed to be constant over one wavelength. Therefore any substructure is not allow as well as any changes in the radiation amplitude.

The latter becomes a problem for the superradiant regime above the saturation power level for a spike [35]. The amplitude is growing but the spike has the tendency to get shorter due to the superradiant process. At a point the amplitude rises so quickly that frequency components of the pulse can lie outside of the simulation bandwidth. Instead further shortening, the FEL algorithm numerically filters the signal thus preventing the pulse to become short. This has been verified with simulations [36], which do not rely on the bandwidth limitation of the period-averaged equations.

A similar problem is to model an energy chirp correctly. While a linear chirp and dispersion yield a smooth compression within the undulator the periodic boundary condition of a given slice in the FEL model, converts the chirp into a saw-tooth distribution in longitudinal phase space. Along the undulator the macro particle distribution would bunch with the resonant wavelength, causing an unphysical level of coherent emission. To avoid this problem, an energy chirp can only be described by a stair-like distribution, where each step belong to an electron slice. Now the slices would not bunch, however the different mean energy per slice would cause an overlapping of the slices, mimicking the compression. A sorting and rearranging of particles would be difficult without generating unphysical phase-space distributions. Note that the problem also occurs in the high current beam case, described above, where the strong space charge field induces an energy chirp naturally.

Large Harmonic Conversion

Echo-enabled Harmonic Generation (EEHG) [37] or a cascade of several stages of High Gain Harmonic Generation (HGHG) can yield very high harmonic conversion in the order of 50 and higher, pushing the modeling of harmonics to a limit. First the simulation must carry all harmonics from the beginning to avoid artificial noise when performing the harmonic conversion. In addition the electron beam has its finest sampling with the long wavelength of the initial seed laser. As a result the thickness of the electron slices remains the same over the entire cascade and the final bandwidth, resolved by the simulation is a few percent or even less. Even the slightest variation in the beam energy can shift the resonant wavelength out of the resonant bandwidth.

EEHG has also the problem that after the first stage, a beamlet is violently distributed over many slices. Therefore the concept of beamlet breaks apart (However ignoring this and enforcing periodic boundary condition have given reasonable results for estimating the EE HG performance numerically). As for the HGHG cascade, rearranging the particles seems straight forward but would introduce too much numerical noise as well as correlation over the wavelength.

OUTLOOK

With growing, accessible computer the utilization of parallel computers or distributed computing allow for new and more precise algorithms. In this last section a few promising concepts are described, which could lead to the next generation of FEL codes. They will be written with parallel computer architecture code in mind, unlike existing codes, which have been ported from a single processor machine and thus remain a single processor core inside.

One approach is to drop the period-average equation and resolve the electron motion on a finer scale [38]. Although this restricts the integration step size to a sub period length the increase in the computational power compensates for it. This Ansatz is attractive because it removes the need to model the harmonics as individual fields. Instead only one field representation is needed with combines the information of the fundamental and the harmonics. Effects as the narrowing of a superradiant spike can be modeled correctly. Because the entire field is advanced in a single step,

the codes allows for particle distributions with a substructure on a scale smaller than the wavelength (e.g. energy chirp).

An interesting idea is to overcome the large disparity between undulator length and electron bunch. By a careful chosen Lorentz transformation the bunch length is increased while the undulator field becomes an incident radiation field with the same length as the bunch length in the new frame [39]. This is the smallest possible domain suitable for a traditional particle-in-cell code approach.

Using the distributed memory of a cluster, the entire field information and electron distribution can be stored in memory while resolving each individual electron. With around 10 billion electrons as an upper limited for an FEL pulse a total memory of 500 GByte would be require to hold the entire distribution. The field information is about of the same order for an Ångstrom FEL or even less for longer wavelength. This demand can be fulfilled by cluster of the size of 500 nodes and more, which is not considered large by today's size of supercomputers. In these one-one simulation a lot algorithms can be simplified to a trivial level, e.g. a random number generator could be used to generate the particle distribution, while providing the correct bunching factor at any wavelength, including the coherent enhancement by the bunch shape at longer wavelength. Also the resorting of particles (in particular for violent redistribution in an EE-HG) is simplified.

In conclusion it can be expected that in the near future numerical tools will be develop to master the remaining challenges left in FEL beam dynamics.

REFERENCES

[1] Linac Coherent Light Source, Conceptional Design Report, SLAC-R-593

[2] T. Shintake *et al*, *SPring-8 Compact SASE Source* SPIE2001, San Diego USE (2001)

[3] The European X-Ray Free Electron Laser TDR, DESY 2006-097 (2006)

[4] B.D. Patterson *et al*, New Journal of Physics **12** (2010) 035012

[5] A.M. Kondratenko and E.L. Saldin, Part. Accl. **19** (1980) 207

[6] R. Bonifacio, C. Pellegrini, L.M. Narducci, Opt. Comm. **50** (1984) 373.

[7] L.H. Yu and J. Wu, Nucl. Inst. & Meth. **A483** (2002) 493

[8] A.E. Siegman, *Lasers*, (University Science Books, 1986, Sausalito)

[9] M. Xie, *Design Optimization for an X-ray Free Electron Laser Driven by SLAC Linac* Proceeding of the PAC Conference, Dallas, USA (1995) 183

[10] N. Kroll, P. Morton, M. Rosenbluth, IEEE J. Quantum Elec., **QE-17** (1981) 1436

[11] M. Borland, *elegant: a Flexible SDDS-Compliant Code for Accelerator Simulation* APS LS-287 (2000)

[12] I.M. Sobol, USSR Comp. Math. & Math. Phys. **7** (1967) 86

[13] J.H. Halton, Numerische Mathematik **2** (1960) 84

[14] L. Devroye, *Non-Uniform Random Variate Generation*, (Springer, New York, 1986)

[15] G.E.P. Box and M.E. Muller, Ann. Math. Statist. **29** (1958) 610

[16] I. Zagorodnov and M. Dohlus, *Numerical FEL Studies with a New Code ALICE*, Proceeding of the FEL Conference, Liverpool, UK (2009) 71

[17] V. Litvinienko, *private communication*

[18] C. Penman and B.W.J. McNeil, Opt. Comm. **80** (1992) 82

[19] W.M. Fawley, Phys. Rev. STAB **5** (2002) 070701

[20] B.W.J. McNeil *et al*, Phys. Rev. STAB **6** (2003) 070701

[21] A. Gover and E. Dyunin, Phys. Rev. Lett. *102* (2009) 154801

[22] P. Kaps and P. Rentrop, Numerische Mathematik, **33** (1979) 55

[23] C.W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations* (Englewood Cliffs, Prentice-Hall, 1971)

[24] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis* (Springer, New York, 1980)

[25] R.W.B. Best and B. Faatz, J. Phys. D: Appl. Phys. **23** (1990) 1337

[26] R.D. Richtmyer and K.W. Morton, *Difference Methods for Initial Value Problems* (John Wiley and Sons, New York, 1967)

[27] H.B. Keller, *Numerical Methods for Two Point Boundary Value Problems* (Waltham, MA:Blaisdell) (1986) 74

[28] W.F. Ames, Numerical Methods for Partial Differential Equations (Academic Press, New York, 1977)

[29] D.M. Young, *Iterative Solution of Large Linear System* (Academic Press, New York, 1971)

[30] R.S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, New Jersey, 1962)

[31] A. Brandt, Math. Comp. **31** (1977) 333

[32] M. Xie, Nucl. Inst.& Meth. **A 483** (2002) 527

[33] J.B. Murphy *et al*, Opt. Comm. **53** (1985) 197

[34] F. Gruner *et al* Appl. Phys. *B 86* (2007) 431

[35] T. Watanabe *et al*, Phys. Rev. Lett. *98* (2007) 034802

[36] B.W.J. McNeil *et al*, *An Improved 1-D Model for Ultra High Power Radiation Pulse Propagation in the Helical Wiggler Free Electron Laser*, Proceeding of the PAC Conference, Portland, USA (2003) 953

[37] D. Xiang and G. Stupakov, Phys. Rev. STAB **12** (2009) 030702

[38] C.K.W. am *et al*, *Unaveraged Three-Dimensional Modelling of the FEL* Proceeding of the FEL Conference, Gyeongju, South Korea (2008) 47

[39] W.M. Fawley and J.-L. Vay, LBNL Paper LBNL-1003E (2008)