Title

X-ray FEL linear accelerator design via start-to-end global optimization

Author

Qiang, Ji

Publication Date

2022-03-01

DOI

10.1016/j.nima.2021.166294

Peer reviewed

X-ray FEL linear accelerator design via start-to-end global optimization

Ji Qiang

Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Abstract

An x-ray Free Electron Laser (FEL) prefers using an electron beam with low emittance, small energy spread, and a high core current to generate coherent radiation through an undulator. In order to attain such a high brightness beam, the linear accelerator beam dynamics design generally involves separate photoinjector optimization and linac optimization. In this paper, we propose a new beam dynamics design strategy based on global optimization with fast start-to-end simulations from the photocathode to the end of the accelerator. The new start-to-end model significantly reduces the simulation time and makes the global optimization practical. The global optimization method avoids the need to choose a single solution based on bunch length at the injector exit for the linac optimization and helps find the solution with unfavorable bunch length at the injector exit but better phase space distribution that can result in better final electron beam phase space distribution at the entrance of the undulator. Using the start-to-end global optimization, we showed in an application example, with a 100 pC beam that good transverse emittance and over kilo-Ampere final core current can be attained using a photoinjector that consists of a VHF gun and boosting RF cavities.

Preprint submitted to NIMA *December 18, 2021*

1. Introduction

Coherent radiation from an x-ray free electron laser (FEL) light source provides an important tool for scientific discovery in physics, chemistry, biology and other fields. To produce such a radiation at short x-ray wavelength, it is preferable to use a high brightness electron beam with a high core current, small energy spread, and small emittance through the FEL undulator. In most modern x-ray FEL light sources, the high brightness electron beam is produced through a linear accelerator beam delivery system [1, 2, 3, 4, 5, 6]. This type of accelerator typically consists of a photoinjector that generates an initial high brightness electron beam and a Radio Frequency (RF) linac that accelerates the beam to multiple GeV energy and compresses the beam to hundreds or thousands Ampere core current before entering the x-ray FEL radiation undulator section.

At present, the beam dynamics design of an x-ray FEL accelerator is generally divided into a photoinjector design and a RF linac design. In the photoinjector beam dynamics design, a multi-objective optimizer based on an evolutionary algorithm is used together with a beam dynamics simulation program to find optimal solutions at the exit of the injector [7, 8, 9, 10]. The beam dynamics program simulates the electron beam generating from the photocathode, accelerating, and transporting through the photoinjector. The multi-objective optimizer uses the simulation results at the exit of the photoinjector as objective function values and adjusts control parameters inside the injector to obtain a set of optimal solutions so that each optimal solution is not worse than the other feasible solutions. The control parameters of photoinjector typically involve laser pulse length, transverse spot size, RF gun and boosting cavity amplitudes and phases, and solenoid strengths. The objective functions typically involve transverse Root-Mean-Square (RMS) emittance and RMS bunch length at the injector exit. Some constraints such as final beam energy, energy spread, longitudinal current skewness and high-order nonlinearity of phase space are applied at the exit of the injector during the optimization process. After the photoinjector beam dynamics optimization is done, one specific solution is selected from the optimal solutions and passed to the linac group for the linac beam dynamics design.

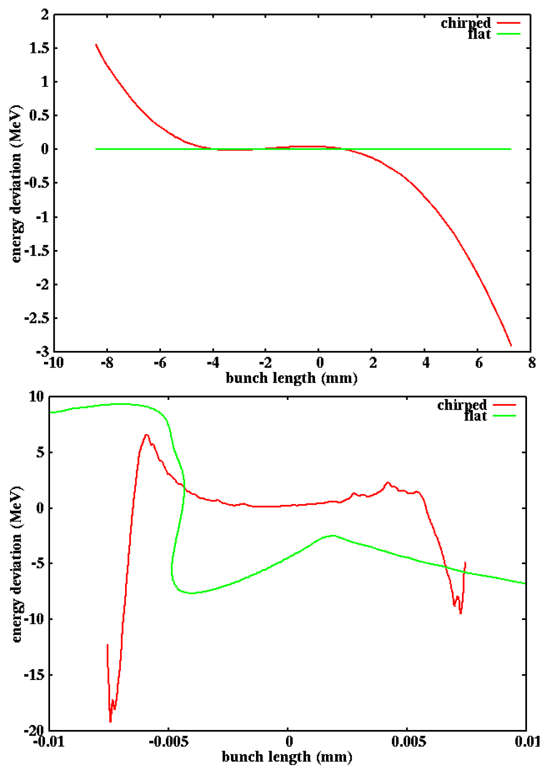

Figure 1: Initial longitudinal chirped (red) and flat (green) phase space distributions (top) with the same current profile and final longitudinal phase space distributions (bottom) from the initial chirped (red) and from the initial flat (green) phase space distributions after passing through the same linac settings.

Using the selected photoinjector solution as an initial beam distribution, the linac group will carry out beam dynamics optimization of the rest of the linear accelerator before sending the electron beam into the x-ray FEL radiation undulator. As the beam energy is sufficiently high inside the linac, the relative longitudinal positions among electrons will not change significantly through most part of the accelerator except the magnetic bunch compression regions, where the electron longitudinal bunch length is compressed in order to achieve a high core current. The transverse beam dynamics and the longitudinal beam dynamics can be optimized separately in the linac design. In the longitudinal beam dynamics design, the linac parameters such as RF cavity amplitudes and phases and bending angles inside the magnetic bunch compression regions are adjusted to attain a desired final peak current and longitudinal phase space distribution at the entrance of the undulator section [11]. The linac beam dynamics simulation program will track electron beam through kilometer long accelerator to obtain final longitudinal phase space distribution and peak core current. The optimizer will automatically adjust control parameters inside the linac to optimize the objective functions at the entrance of the undulator section. These objective functions can be final negative fraction of particles that measures the core current and RMS energy spread that measure the flatness of the longitudinal phase space distribution. After the longitudinal beam dynamics optimization, the transverse beam dynamics optimization can be done by tuning quadrupole settings inside the linac to minimize transverse emittance growth through the accelerator.

In the above design process, the final electron beam quality after the linac beam dynamics optimization will depend on the initial electron beam distribution at the exit of the photoinjector. A single solution selected from the optimal photoinjector solutions of transverse RMS emittance and RMS bunch length does not take into account the effect of electron beam phase space distribution at the exit of the injector on the final beam distribution at the end of the accelerator system. Figure 1 shows an example of two initial distributions with exactly the same current profile, i.e. RMS bunch length, but different longitudinal phase space distributions at the entrance of the linac and the final longitudinal phase space distributions from these two initial distributions. Here, one initial longitudinal phase space (energy) distribution has a nonlinear dependence (chirp) between the energy and the position, while the other phase space distribution is flat. It is seen that the initial chirped longitudinal phase-space distribution results in a flat final longitudinal phase-space distribution while the initial flat longitudinal phase space distribution results in a distorted final longitudinal distribution. This example suggests the importance of the initial longitudinal phase space distribution to the final phase space distribution even though both initial distributions have the same current profile. The use of a single optimal solution from the photoinjector optimization might miss a potential good longitudinal phase space distribution solution from the photoinjector simulation that results in a better final core peak current and longitudinal phase space distribution.

In this study, we propose to use an integrated start-to-end simulation in the x-ray FEL linear accelerator beam dynamics design optimization. This design will simultaneously optimize the photoinjector control parameters and the linac control parameters to attain optimal solutions of three objectives, transverse RMS emittance at the exit of the photoinjector, final core current, and energy spread at the entrance of the undulator section. Through the start-to-end simulation, there is no need to decide on a single optimal solution based on RMS bunch length at the exit of the photoinjector. The effect of the longitudinal phase space distribution at the exit of the photoinjector on the final electron beam longitudinal phase space distribution is automatically included through the start-to-end simulation. Such a start-to-end optimization design helps find the solution with an initial unfavorable RMS bunch length (or current) at the exit of the photoinjector, but could lead to better final electron beam longitudinal phase space distribution and core current at the entrance of the undulator section.

A global beam dynamics optimization based on the start-to-end simulation was tried in an LCLS-II design study [12]. In that study, a brute force start-to-end model based on fully three-dimensional injector and linac simulations was used in the optimization. Using a three-dimensional model through the linac substantially slowed down the computational speed and made the global optimization extremely time consuming. In this study, we took advantage of the separation of longitudinal beam dynamics and transverse beam dynamics inside the linac, and used a fast longitudinal beam dynamics model to simulate electron beam evolution through the linac beam delivery system. This resulted in a new start-to-end simulation model that significantly reduced the computational time and made the global optimization with the start-to-end simulation practical. In addition, we improved the original multi-objective optimization method and proposed a new design strategy by optimizing three objective function from the start-to-end simulation simultaneously.

In the following sections, after the Introduction, we present the start-to-end beam dynamics simulation model in Section II; We discuss the multi-objective global optimization method in Section III; We illustrate the multi-objective start-to-end global optimization with an application example in Section IV; and draw conclusions in Section V.

2. start-to-end beam dynamics simulation model

The start-to-end beam dynamics simulation of an electron beam through the x-ray FEL accelerator starts with generating a three-dimensional electron macroparticle distribution with given thermal emittances behind the photocathode following the laser pulse's longitudinal and transverse distributions. Here, each macroparticle represents a number of real electrons. These macroparticles are then moved out of the cathode during the given emission time. During the process of emission, space-charge forces among the macroparticles outside the cathode are included in the simulation together with external fields from the RF gun and the solenoid magnet inside the photoinjector. After that, the electron beam macroparticles will be further accelerated by boosting RF cavities through the injector.

In the beam dynamics simulation, the macroparticles inside the photoinjector are advanced self-consistently using a particle-in-cell approach with time as an independent variable [13]. The equations of motion for a macroparticle are given as:

$$\dot{\mathbf{r}} = \frac{\mathbf{p}}{m\gamma} \tag{1}$$

$$\dot{\mathbf{p}} = q(\mathbf{E} + \frac{\mathbf{p}}{m\gamma} \times \mathbf{B}) \tag{2}$$

where, $\mathbf{r}$ is spatial position vector, $\mathbf{p}$ is mechanic momentum vector, $\gamma = 1/\sqrt{1-\beta^2}$, $\beta^2 = \sum \beta_i^2$, $\beta_i = v_i/c$ with $i = x, y, z$, c is the speed of light in vacuum, m is the rest mass of particle, q is the charge of particle. The electric field, $\mathbf{E}$, and magnetic field, $\mathbf{B}$, include the contributions from both external focusing and accelerating fields and space-charge fields of intra-particle Coulomb interactions.

The equations of motion are solved using a second-order leap-frog algorithm: the particles are drifted half time step; the particles are collected and deposited onto a three-dimensional grid in the beam frame; the Poisson equation is solved in the beam frame; the electric and magnetic fields are obtained in the laboratory frame through the Lorentz transformation; the particle momenta are updated using both the space-charge fields and the external fields for one time step following Eq. 2; the particles are drifted another half time step. This procedure is repeated for many time steps until the beam exits the photoinjector.

To calculate the space-charge forces, we solve the three-dimensional Poisson equation in the beam frame. The solution of Poisson's equation can be written as:

$$\phi(x,y,z) = \frac{1}{4\pi\epsilon_0}\int\int\int G(x,x',y,y',z,z')\rho(x',y',z')\ dx'dy'dz' \quad (3)$$

where G is Green's function, ρ is the charge density distribution function. For the electron beam inside a photoinjector, an open boundary condition can be assumed for the solution of the Green's function in the above equation.

The computational domain containing the macroparticles has a range of $(0, L_x)$, $(0, L_y)$ and $(0, L_z)$, and each dimension is discretized using N_x, N_y and N_z points. The integral of the above equation in the entire computational domain can be written as a sum of integrals of individual cells. If we assume that the charge density is constant within each cell centered at the grid point (x_i, y_j, z_k), from Eq. 3, the electric potentials on the grid can be approximated as:

$$\phi(x_i,y_j,z_k) = \frac{1}{4\pi\epsilon_0}\sum_{i'=1}^{N_x}\sum_{j'=1}^{N_y}\sum_{k'=1}^{N_z}\bar{G}(x_i-x_{i'},y_j-y_{j'},z_k-z_{k'})\rho(x_{i'},y_{j'},z_{k'}) \quad (4)$$

where $x_i = (i-1)h_x$, $y_j = (j-1)h_y$, and $z_k = (k-1)h_z$, and the effective Green function $\bar{G}$ is defined as:

$$\bar{G}_(x_i-x_{i'},y_j-y_{j'},z_k-z_{k'}) = \int_{x_{i'}-h_x/2}^{x_{i'}+h_x/2}dx'\int_{y_{j'}-h_y/2}^{y_{j'}+h_y/2}dy'\int_{z_{k'}-h_z/2}^{z_{k'}+h_z/2}dz'G(x_i-x',y_j-y',z_k-z') \quad (5)$$

where h_x, h_y, and h_z are cell size in each dimension respectively. The above integral can be calculated analytically in a closed form for the Green function $G = 1/R$ [14]:

$$\int\int\int \frac{1}{\sqrt{x^2+y^2+z^2}}dxdydz \doteq yz\ln(x+\sqrt{x^2+y^2+z^2}) + xz\ln(y+\sqrt{x^2+y^2+z^2}) + xy\ln(z+\sqrt{x^2+y^2+z^2}) - \frac{z^2}{2}\arctan(\frac{xy}{z\sqrt{x^2+y^2+z^2}}) - \frac{y^2}{2}\arctan(\frac{xz}{y\sqrt{x^2+y^2+z^2}}) - \frac{x^2}{2}\arctan(\frac{yz}{x\sqrt{x^2+y^2+z^2}}) \quad (6)$$

After the electron beam exits the photoinjector, a fast longitudinal beam dynamics model with position as an independent variable is used to track those macroparticles through the linac and beam transport lattice to the entrance of the undulator section [15]. Using such a simplified model is based on the observation that for an electron beam with an energy over multiple MeVs, transverse focusing does not significantly affect longitudinal phase space distribution of the beam. The longitudinal and the transverse beam dynamics designs through the rest of the linear accelerator can be done separately. This dramatically improves the computational speed to track the macroparticles through thousands of beam line elements of the accelerator system.

In the fast longitudinal beam dynamics model, each electron macroparticle has longitudinal coordinates $(z, \Delta\gamma)$ with respect to the reference particle (s_0, γ_0) and charge weight w. Here, $z = s - s_0$ is the bunch length coordinate (z_{max} corresponds to the bunch head and z_{min} the bunch tail), $\Delta\gamma = \frac{E-E_0}{mc^2}$, E is the total energy of the particle, and E_0 is the total energy of the reference particle. For the longitudinal beam dynamics study, we include only drifts, RF cavities, and magnetic compression chicanes as the beam line elements of the x-ray FEL linear accelerator. The other focusing elements such as quadrupoles are treated as drifts.

For a macroparticle transporting through a lumped RF cavity element with total length L_{acc}, its longitudinal coordinates will be updated following a leap-frog approximation:

$$\begin{aligned}
z^+ &= z_1 + \frac{L_{acc}}{2}\Delta\gamma_1/(\gamma_{01}\beta_{01})^3 &(7)\\
\gamma_0^+ &= \gamma_{01} + \frac{L_{acc}}{2}\frac{qV_{acc}}{mc^2}\cos(\phi_0) &(8)\\
\Delta\gamma_2 &= \Delta\gamma_1 + L_{acc}\frac{qV_{acc}}{mc^2}(\cos(\phi_0 - kz^+) - \cos(\phi_0)) &(9)\\
z_2 &= z^+ + \frac{L_{acc}}{2}\Delta\gamma_2/(\gamma_0^+\beta_0^+)^3 &(10)\\
\gamma_{02} &= \gamma_0^+ + \frac{L_{acc}}{2}\frac{qV_{acc}}{mc^2}\cos(\phi_0) &(11)
\end{aligned}$$

where subscript 1 and 2 denote the quantity before and after the lumped cavity element respectively, $V_{acc} = qV_{rf}/L_{acc}$ is the accelerating gradient amplitude, k is the RF wave number, and ϕ_0 is the RF cavity design phase.

The magnetic bunch compression chicane is modeled as a thin lens element. The particle longitudinal position after the chicane is given by [16]:

$$z = z + R_{56}\frac{\Delta\gamma}{\gamma_0} + T_{566}(\frac{\Delta\gamma}{\gamma_0})^2 + U_{5666}(\frac{\Delta\gamma}{\gamma_0})^3 \qquad (12)$$

where

$$\begin{aligned}
R_{56} &\approx 2\theta^2(L_{db} + \frac{2}{3}L_b) &(13)\\
T_{566} &\approx -\frac{3}{2}R_{56} &(14)\\
U_{5666} &\approx 2R_{56} &(15)
\end{aligned}$$

where θ is the bending angle of one of dipole magnets (assuming that all four dipoles have the same bending angle amplitude), L_b is the length of the dipole magnet, and L_{db} is the distance between the first and the second (or between the third and fourth) dipole bending magnets.

Collective effects such as the longitudinal space-charge effect, structure and resistive wall wakefields, and the coherent synchrotron radiation play an important role in the longitudinal beam dynamics and are included in this model. For the longitudinal space-charge effect, instead of using the space-charge impedance model in the frequency domain, we assume that the electron beam is a round cylinder with separable uniform transverse density distribution and longitudinal density distribution. The longitudinal space-charge field on the axis is given as:

$$E_z^{sc}(0,0,z) = \frac{1}{4\pi\epsilon_0}\frac{2}{a^2}\int\int \frac{\gamma_0(z-z')\rho(z')}{(\gamma_0^2(z-z')^2+r'^2)^{3/2}} r' dz' dr' \quad (16)$$

After integrating with respect to the transverse radial dimension, the longitudinal space-charge field on the axis can be written as:

$$E_z^{sc}(z) = \frac{1}{4\pi\epsilon_0}\frac{2}{a^2}\Big(\int_{z_{min}}^{z} \rho(z')dz' - \int_{z}^{z_{max}} \rho(z')dz' - \int_{z_{min}}^{z_{max}} \frac{\gamma_0(z-z')\rho(z')}{\sqrt{\gamma_0^2(z-z')^2+a^2}} dz'\Big) \quad (17)$$

where a is the radius of the cylinder, z_{min} and z_{max} denote the minimum and the maximum longitudinal bunch length positions, and ρ is the electron beam longitudinal charge density distribution. The above convolution can be computed efficiently using an FFT based method [17, 18].

The longitudinal wakefields from both the structure wakefields of RF cavities and the resistive wall wakefields of conducting pipes are included in the model. The coherent synchrotron radiation effects through a bending magnet are also included in the model using a one-dimensional model [19]. This fast longitudinal beam dynamics model was benchmarked with the three-dimensional simulation using the IMPACT code [20, 21] in the previous study. The benchmark between the above 1D longitudinal beam dynamics model using lumped elements and the 3D element-by-element multi-particle simulation shows good agreement between these two models. The use of the 1D model is based on the assumption of decoupling between the transverse and the longitudinal beam dynamics, which is normally valid in x-ray FEL electron linacs. In an electron accelerator where the coupling between the transverse and the longitudinal dynamics is significant, the 3D simulation will be needed.

3. Multi-Objective Global Start-to-End Optimization

In accelerator design, there could be multiple physical objectives (e.g. emittance and current) that need to be optimized simultaneously. The multi-objective optimization can be written in the general mathematical form as:

$$min \begin{cases} f_1(\vec{x}) \\ \dots \qquad \textit{subject to constraints} \\ f_n(\vec{x}) \end{cases} \tag{18}$$

where $f_1, \cdots, f_n$ are n objective functions to be optimized and $\vec{x}$ is a vector of control parameters. The goal of multi-objective optimization is to find the Pareto front in the feasible objective solution space. The Pareto front is a collection of non-dominated solutions in the entire feasible solution space. Any other solutions in the feasible solution space will be dominated by those solutions on the Pareto optimal front. In the multi-objective optimization, a solution A dominates a solution B if all components of A are at least as good as those of B (with at least one component strictly better). Here, a component of A corresponds to one objective function value in the optimization problem, i.e. $A_i = f_i(\vec{x})$. The solution A is non-dominated if it is not dominated by any solutions.

For global optimization of accelerator designs, we adopted an evolutionary algorithm. An evolutionary algorithm is a stochastic global optimization method.

It uses a group of solutions as a population, evolves these solutions from one generation to the next generation like biological evolution. The number of solutions in the group is called population size. Different evolutionary algorithms use different strategies to generate next generation solutions (offspring population), from the present solutions (parent solutions). The mutation strategy is a method to generate a new solution from the parent solutions, which will be discussed in the following sections.

The genetic evolutionary optimization algorithm has been used in accelerator community [22, 23, 24, 25]. In this study, we extended a recently developed multi-objective differential evolution algorithm for the x-ray accelerator start-to-end global design optimization. This algorithm varies population size of each generation and uses an external storage to save all non-dominated solutions [12]. The use of variable population from generation to generation is based on the observation that during the early stage of evolution, the number of nondominated solutions is small. There is no need to keep many dominated solutions in the parent population. As the search evolves, more and more nondominated solutions are obtained. These nondominated solutions are stored in an external storage so that they can be used to select the new parent population. The variable population size with the external storage helps reduce the number of objective function evaluations and improve the speed of convergence. This algorithm is summarized in the following steps:

- Step 0: Define the minimum parent population size, NP_{min} and the maximum size, NP_{max} of the population. Define the maximum size of the external storage, NP_{ext}.

- Step 1: An initial NP_{ini} population of control parameter vectors are sampled quasi-randomly to cover the entire feasible parameter space.

- Step 2: Generate the offspring population using a unified differential evolution algorithm.

- Step 3: Check the new population against the constraints.
- Step 4: Combine the new population with the existing parent population from the external storage. NP_{ndom} Non-dominated solutions are obtained from this group of solutions and $min(NP_{ndom}, NP_{ext})$ of solutions are put back to the external storage. Pruning is used if $NP_{ndom} > NP_{ext}$. NP parent solutions are selected from this group of solutions for next generation production. If $NP_{min} \leq NP_{ndom} \leq NP_{max}$, $NP = NP_{ndom}$. Otherwise, $NP = NP_{min}$ if $NP_{ndom} < NP_{min}$ or $NP = NP_{max}$ if $NP_{ndom} > NP_{max}$. The elitism is emphasized through keeping the non-dominated solutions while the diversity is maintained by penalizing the over-crowded solutions through pruning by removing the solution with least distance to the other solutions.
- Step 5: If the stopping condition is met, stop. Otherwise, return to Step 2.

The differential evolution algorithm is a simple but powerful method that uses the differences of parent solutions to generate new candidate solutions in global optimization [26, 27, 28]. It generates new offspring using two operations: mutation and crossover. In the mutation operation, for each population member (target vector) $\vec{x}_i$, $i = 1, 2, 3, \cdots, NP$ at generation G, a new mutant vector $\vec{v}_i$ is generated by following a mutation strategy. A number of mutation strategies have been proposed in the standard differential evolution algorithm. A single unified mutation strategy that contains most standard mutation strategies can be written as [29]:

$$\vec{v}_i = \vec{x}_i + F_1(\vec{x}_b - \vec{x}_i) + F_2(\vec{x}_{r_1} - \vec{x}_i) + F_3(\vec{x}_{r_2} - \vec{x}_{r_3}) \tag{19}$$

where $\vec{x}_b$ is the best solution among the parent solutions, $\vec{x}_{r_1}$, $\vec{x}_{r_2}$ and $\vec{x}_{r_3}$ are three randomly selected pareent solutions, and the three parameters F_1, F_2, and F_3 are the weights from each difference of parent solutions. This unified expression represents a combination of exploitation (using the best found solution) and exploration (using randomly chosen solutions) when generating the new mutant solution. Using the equation (19), the multiple mutation strategies of the standard differential evolution algorithm can be included in a single expression. For example, a standard differential evolution algorithm can be attained by setting $F_1 = 0$, $F_2 = 1$ and $F_3 = 1$. This new expression provides an opportunity to explore more broadly the space of mutation operators. Using a different set of parameters F_1, F_2, F_3, a new mutation strategy can be achieved. Moreover, by adjusting these parameters during the evolution, the multiple mutation strategies and their combinations can be used during different stages of optimization. In this study, these parameters are randomly sampled from a uniform distribution between zero and one at each generation to cover a wide range of mutation strategies. Using a range of mutation strategies improves the diversity of the next generation solutions.

A crossover operation between the new generated mutant vector $\vec{v}_i$ and the target vector $\vec{x}_i$ is used to further increase the diversity of the new candidate solution. This operation combines the two vectors into a new trial vector $\vec{U}_i, i = 1, 2, 3, \cdots, NP$, where the components of the trial vector are obtained from the components of $\vec{v}_i$ or $\vec{x}_i$ according to a crossover probability Cr. In the binomial crossover scheme for a D dimensional control parameter space, the new trial vector $\vec{U}_i$, $i = 1, 2, \cdots, NP$ is generated using the following rule:

$$\vec{U}_i = (u_{i1}, u_{i2}, \cdots, u_{iD}) \tag{20}$$

$$u_{ij} = \begin{cases} v_{ij}, & \text{if } \mathrm{rand}_j \leq Cr \text{ or } j = \mathrm{mbr}_i \\ x_{ij}, & \text{otherwise} \end{cases} \tag{21}$$

where rand_j is a randomly chosen real number in the interval $[0, 1]$, and the index mbr_i is a randomly chosen integer in the range $[1, D]$. This ensures that the new trial vector contains at least one component from the new mutant vector. The cross probability Cr varies from generation to generation following a uniform distribution between 0.5 and one.

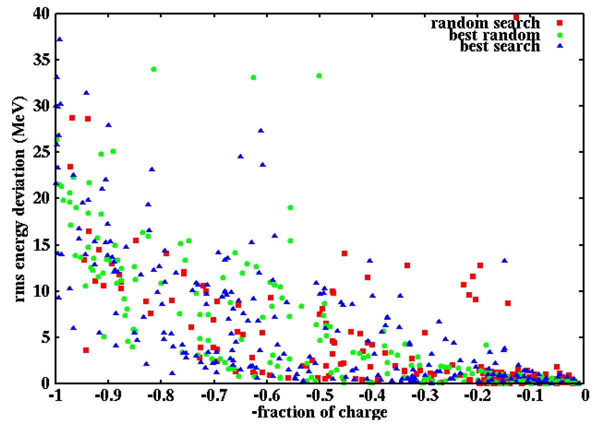

Figure 2: Projected Pareto front of final RMS energy deviation versus negative fraction of charge from the three objective function optimization using *random search* (red), *best random* (green), and *best search* (blue) for the linac sub-optimization.

The objective function values in the above optimizer result from the start-to-end simulation of the x-ray linear accelerator system. In the start-to-end simulation, the computational time spent in the photoinjector section is an order of magnitude longer than that in the rest of the accelerator due to the use of the fast longitudinal beam dynamics model inside that section. In order to enhance exploration of the linac control parameter space, sub-optimization of the linac is performed during each objective function evaluation of the global optimization. That is, during the start-to-end simulation, the longitudinal phase space distribution at the exit of the photoinjector is used for a few extra iterations of the linac control parameters before returning the objective function values of the linac to the global optimizer. In this study, we tested three sub-optimization methods for the linac parameter exploration. These methods include a *random search* method, a *best random* method, and a *best search* method. In the *random search* method, a number of random solutions are obtained from the randomly sampled control parameters in the feasible range of the linac parameters. The best solution (e.g. based on the final core current) is selected from these random solutions and used as the linac objective functions of the start-to-end simulation. In the *best random* method, a best solution is selected from an initial group of random solutions. A few iterations are applied to these solutions following a mutation strategy similar to the *best random* mutation strategy of the differential evolution method. That is, the next generation solution for the i^{th} solution will be:

$$\vec{v}_i = \vec{x}_i + 2\vec{R}_1(\vec{x}_b - \vec{x}_i) + \vec{R}_2(\vec{x}_{r_i} - \vec{x}_{r_j}) \tag{22}$$

Here, R_1 and R_2 are two random number vectors between zero and one from the uniform distribution, $\vec{x}_b$ is the best solution from the previous generation, $\vec{x}_{r_i}$ and $\vec{x}_{r_j}$ are two random solutions of the previous generation. There is no crossover operation in the iteration. The use of the difference of two previous random solutions increases the diversity of the new solution. In the *best search* method, only the best solution of the previous group of random solutions is used to guide the production of next generation solutions, that is:

$$\vec{v}_i = \vec{x}_i + 2\vec{R}_1(\vec{x}_b - \vec{x}_i) \tag{23}$$

There is no use of the difference of two previous solutions to increase diversity of the new solutions.

As a comparison of the above three methods, we ran the three-objective function optimization using the application example in the next section for 11 hours of computing time. Figure 2 shows the projected two-dimensional Pareto front after that computing time using these three linac sub-optimization methods. It is seen that the *best search* method shows better performance than the other two sub-optimization methods and generates more solutions towards larger fraction of charge inside the core with smaller energy spread. In the next section, we will use this method for the linac sub-optimization during the electron beam start-to-end global optimization. When the optimization process approaches convergence, the linac sub-optimization is switched to a local Simplex optimization method for several iterations.

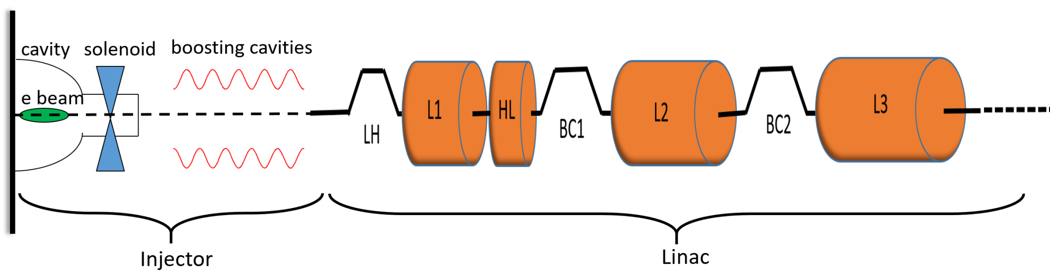

Figure 3: Schematic layout of an x-ray FEL linear accelerator that consists of an injector section and a linac section.

4. An application example

The above beam dynamics design via global start-to-end optimization is illustrated using an x-ray FEL linear accelerator application example. A schematic of the linear accelerator system is shown in Figure 3. This accelerator consists of a photoinjector and a LCLS-II like linac that accelerates a 100 pC electron beam to over 4 GeV final energy [30, 31]. The photoinjector consists of a normal conducting Very-High-Frequency (VHF) RF gun operating at 187 MHz with 20 MV/m accelerating gradient at the photocathode [32], a solenoid magnet, and eight superconducting RF cavities operating at 1.3 GHz frequency with a 32 MV/m maximum electric field on the axis. The linac consists of a laser heater (LH) section, a linac section one (L1) with 16 1.3 GHz RF superconducting cavities and 16 3.9 GHz RF superconducting cavities as harmonic linearizers (HL), a bunch compression section one (BC1), a 1.3 GHz superconducting linac section two (L2), a bunch compression section two (BC2), another 1.3 GHz superconducting linac section three (L3), and a beam transport section to the entrance of the x-ray FEL radiation undulator section. The photoinjector generates and accelerates the electron beam to more than 90 MeV and the linac accelerates the beam to over 4 GeV.

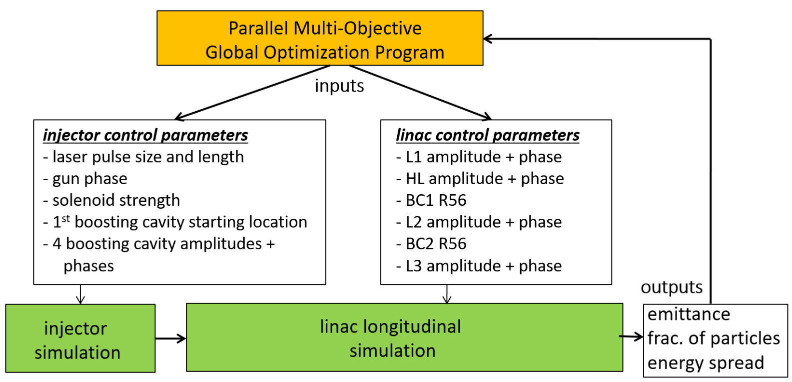

Figure 4: A schematic of flow diagram of global accelerator beam dynamics optimization.

In global start-to-end optimization, there are three objective functions to be optimized. These three objectives are transverse RMS projected emittance at the exit of the photoinjector, negative fraction of particles inside a core window to measure beam peak current, and RMS correlated energy spread inside the window to measure longitudinal phase space flatness at the entrance of the undulator section. Figure 4 shows a flow diagram of the global optimization including both injector control parameters and linac control parameters in the start-to-end beam dynamics optimization. The optimizer calls the start-to-end beam dynamics simulation to obtain objective function values by passing the injector control parameters and the linac control parameters into the subroutines for objective function evaluations. There are total 23 control parameters. A list of these parameters and their ranges is given in Table 1 and Table 2.

Table 1: Injector Control Parameters

Parameter	**value**	**Range**
Laser trans. size (mm)	0.35	0.14 – 0.56
Laser pulse length (ps)	50.5	20.2 – 80.9
Gun phase (deg)	340	306 – 374
solenoid peak field (T)	0.05	0.02 – 0.08
cryomodule starting loc. (m)	1	0.5 – 2
$1^{st} - 4^{th}$ cavity accel. gradient (MV/m)	12.5	0 – 16
$1^{st} - 4^{th}$ cavity phase (deg)	180	0 – 360

Table 2: Linac Control Parameters

Parameter	**value**	**Range**
Linac1 accel. gradient (MV/m)	13	9.8 – 16.0
Linac1 phase (deg)	-14	-28 – -7
Harmonic linearizer gradient (MV/m)	11	8.5 – 13.5
Harmonic linearizer phase (deg)	-150	-195 – -105
BC1 bending angle (rad)	-0.1	-0.12 – -0.08
Linac2 accel. gradient (MV/m)	13	11.7 – 16.0
Linac2 phase (deg)	-20	-40 – 0
BC2 bending angle (rad)	0.044	0.035 – 0.057
Linac3 accel. gradient (MV/m)	15.5	14.0 – 17.0
Linac3 phase (deg)	1	-3 – 3

The 13 control parameters inside the injector are laser transverse size, laser pulse length, VHF gun RF phase, solenoid strength, first four boosting cavity amplitudes and phases, and starting location of the boosting cavity cryomodule. The rest four boosting RF cavities were assumed to run with 32 MV/m maximum field on the axis and zero degree RF design phase. The 10 control parameters inside the linac are the linac section one 1.3 GHz superconducting RF cavity amplitude and phase, the 3rd harmonic cavity amplitude and phase, the bending angle in bunch compressor one, the linac section two RF cavity amplitude and phase, the bending angle in bunch compressor two, and the linac section three RF cavity amplitude and phase. A number of constraints are applied at the exit of the photoinjector. These constraints include the electron beam energy (> 91 MeV), electron RMS energy spread ($< 2\%$), electron beam transverse RMS size (< 0.4 mm), and transverse RMS emittance (< 0.3 mm mrad). At the exit of the photoinjector, the electron beam transverse RMS emittance is recorded as the first objective function value of the optimizer. This emittance sets the limit of the final transverse emittance that can be achieved at the entrance of the undulator section. With a careful design of the linac quadrupole settings to minimize the emittance growth due to the space-charge effects and the CSR effects, the initial emittance through the linac can be reasonably preserved. The electron beam longitudinal phase space distribution at the injector exit is fed into the longitudinal beam dynamics model to track the beam through the superconducting RF linac and transport beam line elements. All simulations were done using about 200 thousands macroparticles with $64 \times 64 \times 64$ grid points through the injector and 64 grid points through the linac. The final negative fraction of particles inside a longitudinal phase space region of $[-7.5 : 7.5]$ μm and $[-5.11 : 5.11]$ MeV, and the RMS energy spread of these partilces are recorded as the second and the third objective function values of the optimizer. The larger fraction of charge inside the window, the higher core current will be. The smaller RMS energy spread inside the window, the flatter longitudinal phase space will be. A high core current and flat longitudinal phase space helps improve the x-ray FEL radiation power and reduce the radiation band- width. Besides constraints at the exit of the injector, there are constraints at the linac and beam delivery system exit such as electron beam energy (> 4 GeV), fraction of particles inside the window (> 0.4), and RMS energy spread (< 0.51 MeV).

Instead of starting with an initial 512 random population of the entire accelerator control parameter space, we first run several generations of the photoinjector optimization only. This ensures that there will be sufficient solutions out of the photoinjector with a reasonable beam energy satisfying the injector energy constraint. Those solutions at the exit of the injector that do not satisfy that constraint will be automatically excluded during the global start-to-end optimization. The photoinjector control parameter solutions are then combined with 512 quasi-random samples of the linac control parameter space to form an initial population in the entire 23 accelerator control parameter space.

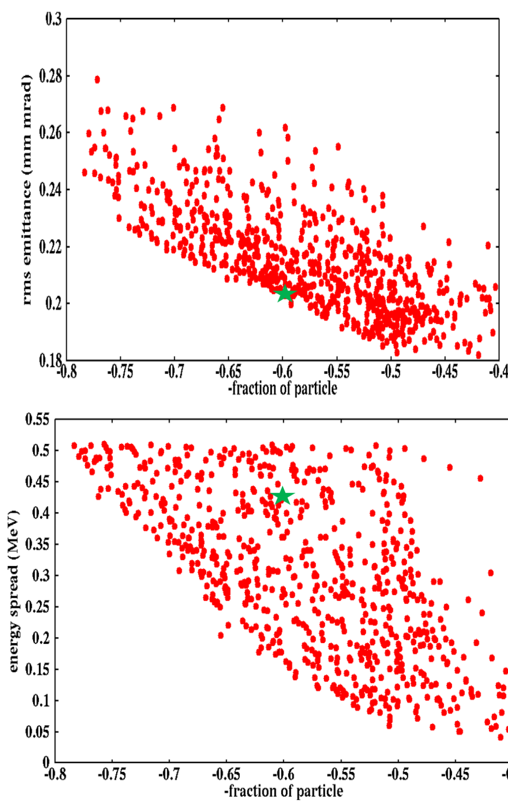

Figure 5: Pareto front of two-dimensional projection of the three objective functions onto the emittance-negative fraction of particle plane (left) and the energy spread-negative fraction of particle plane (right). The green star is the illustrative example solution.

Figure 5 shows Pareto front of the three objectives projected onto a two-dimensional plane for the purpose of better visualization. The transverse RMS emittances at the exit of the injector are below 0.3 mm mrad for all these solutions. These emittances are comparable to those at the exit of LCLS-II photoinjector including a RF buncher cavity [9]. The final RMS energy spreads are also below about 0.5 MeV. There is a correlation between the injector transverse emittance and the final fraction of particles inside the core, and a correlation between the final RMS energy spread and the fraction of particles inside the core. The final higher current with a larger fraction of particles inside the core shows larger transverse RMS emittance at the exit of the photoinjector and larger RMS energy spread at the entrance of the undulator section. The higher core current is probably due to the higher initial peak current at the injector exit and the larger longitudinal compression through the linac. The higher peak current at the exit of the photoinjector is correlated with larger transverse emittance at the exit of the injector. The larger longitudinal compression inside the linac can result in a larger final RMS energy spread at the entrance of the undulator section.

As an illustration, we selected a solution from these Pareto front solutions. Figure 6 shows the electron beam kinetic energy evolution through the photoinjector. The energy of electron beam at the exit of the injector is over 110 MeV. The energy gain through the first boosting cavity is less than the rest of the boosting cavities. The lower energy gain through this cavity helps longitudinal velocity bunching through the cavity. Figure 7 shows the horizontal RMS size and longitudinal RMS size evolution through the injector. The electron beam out of the VHF gun is transversely focused by the solenoid field before entering the boosting RF cavity cryomodule. Inside the first boosting cavity, the beam is longitudinally focused down to around 2 mm through velocity bunching before entering the second boosting cavity. After the second boosting cavity, the longitudinal bunch length stays nearly constant due to the fast acceleration of the electron beam to more than 20 MeV energy. The transverse beam size is further focused by the RF fields inside the second boosting cavity and continues decreasing through the boosting cavity cryomodule. Figure 8 shows the transverse emittance evolution through the injector. The transverse emittance shows little change after the second boosting cavity and stays about 0.2 mm mrad till the exit of the injector. The initial transverse thermal emittance was assumed 1 mm mrad per mm RMS transverse size in all simulations. Figure 9 shows the longitudinal current profile and phase space distribution at the exit of the injector. Here, the relative energy deviation in this figure denotes the individual electron energy deviation with respect to the average beam energy divided by that energy at the exit of the injector. The peak of current is about 6 Ampere while the relative RMS energy spread is less than 1%. This current is much lower than the 12 Ampere peak current at the exit of the LCLS-II injector [33]. However, with the appropriate choice of linac parameters through the start-to-end global optimization, a reasonable final core current can still be achieved. Figure 10 shows the final longitudinal current profile and phase space distribution at the entrance of the undulator section. It is seen that over kilo-Ampere core current is attained with a relatively flat longitudinal phase space distribution. Such a high brightness electron beam can be used for the generation of coherent x-ray FEL radiation.

5. Conclusions

In this paper, we proposed a beam dynamics design method of the x-ray FEL linear accelerator based on multi-objective global optimization with start-to-end simulations. The start-to-end simulation involves three-dimensional self-consistent beam dynamics simulation of electron beam evolution from the pho-

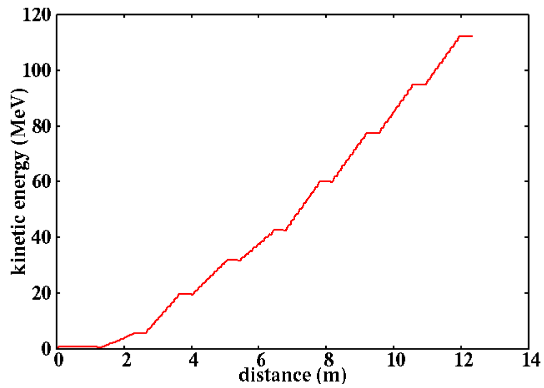

Figure 6: Electron beam kinetic energy evolution through the photoinjector.

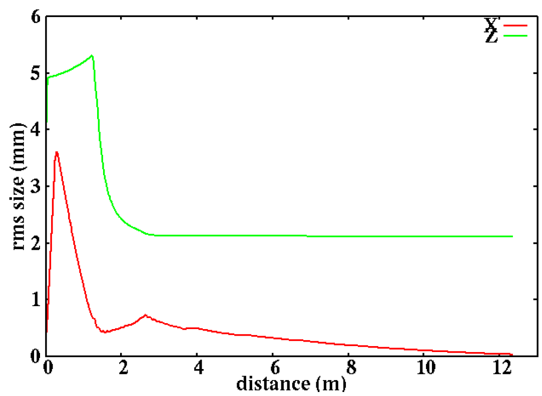

Figure 7: Electron beam transverse RMS size (red) and longitudinal RMS size (green) evolution through the photoinjector.

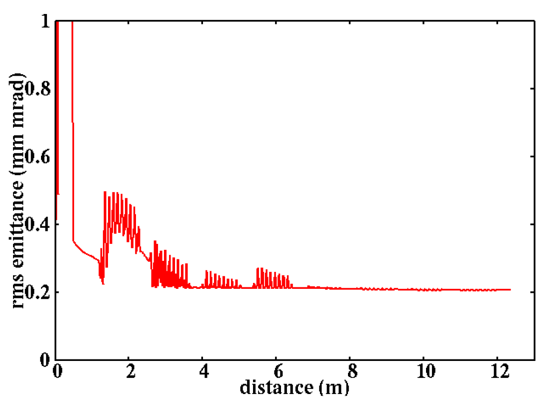

Figure 8: Electron beam transverse RMS projected emittance evolution through the photoinjector.

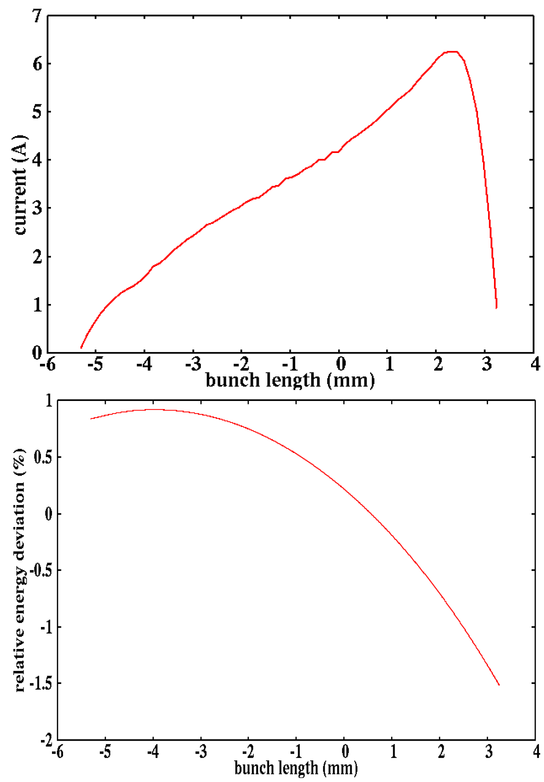

Figure 9: Electron beam current profile (top) and longitudinal phase space distribution (bottom) at the exit of the photoinjector.

tocathode to the end of the injector and fast longitudinal beam dynamics simulation through the RF linac and beam transport system. The global optimization employees control parameters inside both the photoinjector and the RF linac. The objectives of optimization include transverse emittance at the injector exit and the core current and RMS energy spread at the undulator entrance. Using the start-to-end simulation avoids the need to choose a specific solution with a given current and longitudinal phase space distribution at the photoinjector exit for the linac optimization. The impact of the current profile and longitudinal phase space at the injector exit on the final electron beam current and longitudinal phase space is automatically included through the start-to-end simulation. The multi-objective global optimization uses a variable population

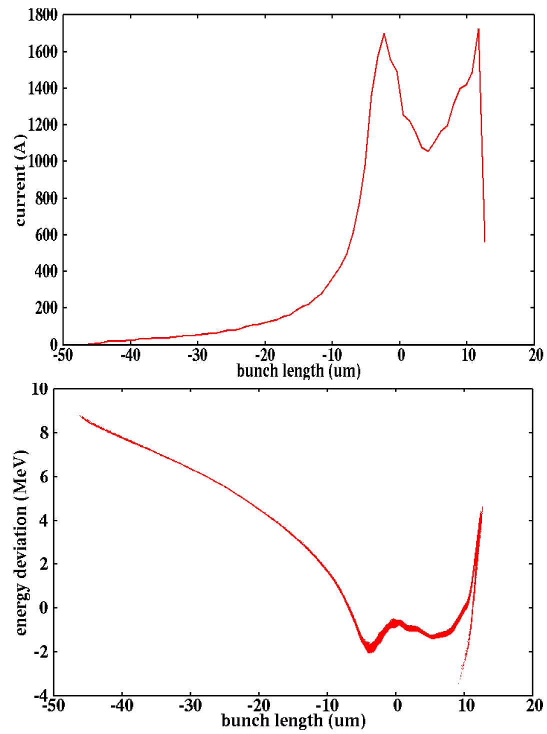

Figure 10: Electron beam current profile (top) and longitudinal phase space distribution (bottom) at the end of the linear accelerator beam delivery system.

size with external storage and a unified differential evolution algorithm with the linac sub-optimization to speed up the search for optimal solutions.

The above global optimization with start-to-end simulations was illustrated using an application example that consists of a photoinjector and a LCLS-II like RF linac. Good solutions with small electron beam transverse emittance, a high final core current, and relatively flat longitudinal phase space distribution were obtained through the start-to-end global optimization. In one solution, over thousand Ampere final core current was achieved with a 6 Ampere current and 0.2 mm mrad transverse emittance electron beam at the exit of the injector without the use of RF buncher cavity in the photoinjector. This suggests that a useful final high brightness electron beam could be obtained for an initial unfavorable solution at the linac entrance through the start-to-end global optimization.

In this study, we used an improved multi-objective optimizer based on the differential evolution method in the accelerator global beam dynamics design optimization. In the future study, we would like to further improve the computational speed in the global design optimization by exploring methods to include surrogate models in the optimizer [34, 35].

ACKNOWLEDGEMENTS

We would like to thank the LCLS-II physics design team for some parameters in the application example. This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and used computer resources at the National Energy Research Scientific Computing Center.

References

[1] P. Emma, et al., Nature Photon. 4 (2010) 641.

[2] E. Allaria, et al., Nature Photon. 6 (2012) 699.

[3] H. Weise and W. Decking, in Proceedings of the 2017 Free Electron Laser Conference, (JACoW, Geneva, 2017), pp. 9-13.

[4] H. Tanaka et al., Nat. Photonics 6, 540 (2012).

[5] H.-S. Kang et al., PAL-XFEL technical design report, Pohang Accelerator Laboratory, 2014.

[6] C. Milne et al., Appl. Sci. 7, 720 (2017).

[7] I. V. Bazarov and C. K. Sinclair, Phys. Rev. ST Accel. Beams 8, 034202 (2005).

[8] C. F. Papadopoulos et al., in Proc. of FEL2010, Malmo, Sweden, p. 479.

[9] C. Mitchell et al., in Proc. IPAC2016, Busan, Korea, p. 1699 (2016).

[10] N. Neveu et al., Phys. Rev. Accel. Beams 2019, 22, 054602.

[11] L. Wang et al., in Proc. FEL2014, Basel, Switzerland (2014), p. 763.

[12] J. Qiang, Int. J. Mod. Phys. A, 1942016 (2019).

[13] J. Qiang, S. Lidia, R. D. Ryne, and C. Limborg-Deprey, Phys. Rev. ST Accel. Beams **9**, 044204, (2006).

[14] J. Qiang, S. Lidia, R. D. Ryne, and C. Limborg-Deprey, Phys. Rev. ST Accel. Beams **10**, 129901 (2007).

[15] J. Qiang, Phys. Rev. Accel. Beams, vol. 22, 094401, 2019.

[16] M. Dohlus, T. Limberg, P. Emma, "Bunch compression for linac-based FELs," in ICFA Beam Dynamics Newsletter, No. 38, p.15, 2005.

[17] R.W. Hockney, J.W. Eastwood, Computer Simulation Using Particles, Adam Hilger, New York, 1988.

[18] W. H Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical recipes in Fortran: the art of scientific computing, 2nd ed. Cambridge University Press, 1992.

[19] E.L. Saldin, E.A. Schneidmiller, M.V. Yurkovb Nuclear Instruments and Methods in Physics Research Section A398 (2012) 373.

[20] J. Qiang, R. D. Ryne, S. Habib, V. Decyk, J. Comput. Phys. **163**, 434, 2000.

[21] J. Qiang, R. D. Ryne, M. Venturini, A. A. Zholents, and I. V. Pogorelov, Phys. Rev. ST Accel. Beams, vol. 12, 100702, 2009.

[22] R. Hajima, N. Takeda, H. Ohashi, and M. Akiyama, Nucl. Instrum. Methods Phys. Res., Sect. A 318, 822 (1992).

[23] L. Emery, in Proceedings of the 21st Particle Accelerator Conference, Knoxville, 2005, pp. 2962-2964.

[24] L. Yang, Y. Li, W. Guo, and S. Krinsky, Phys. Rev. ST Accel. Beams **14**, 054001 (2011).

[25] A. Hofler, B. Terzic, M. Kramer, A. Zvezdin, V. Morozov, Y. Roblin, F. Lin, and C. Jarvis, Phys. Rev. ST Accel. Beams **16**, 010101 (2013).

[26] R. Storn and K. Price, Journal of Global Optimization 1a1:341-359, (1997).

[27] M. M. Ali and A. Torn, Computers and Operations Research, Elsevier, no. 31, p. 1703, 2004.

[28] K. Price et al., Differential Evolution - A Practical Approach to Global Optimization, Springer, Berlin, 2005.

[29] J. Qiang, C. Mitchell, and A. Qiang, in Proc. CEC2016, Vancouver, p. 4061, 2016.

[30] T. O. Raubenheimer et al., in Proceedings of the 2015 Free Electron Laser Conference, Daejeon, Korea, (JACoW, Geneva, 2015), pp. 618-624.

[31] LCLS-II Conceptual Design Report, SLAC-R-1092, http://slac.stanford.edu/pubs/slacreports/reports09/slac-r-1092.pdf.

[32] K. Baptiste et al., Nuclear Instruments and Methods in Physics Research A 599, 9 (2009).

[33] F. Zhou et al., in Proc. IPAC2019, Melbourne, Australia, p. 2171, 2019.

[34] M. Kranjcevic, B. Riemann, A. Adelmann, and A. Streun, Phys. Rev. Accel. Beams **24**, 014601, 2021.

[35] R. Roussel, A. Hanuka and A. Edelen, Phys. Rev. Accel. Beams **24**, 062801, 2021.