

Multivariate optimization of a high brightness dc gun photoinjector

Ivan V. Bazarov* and Charles K. Sinclair†

Laboratory for Elementary Particle Physics, Cornell University, Ithaca, New York 14853, USA

(Received 1 February 2005; published 24 March 2005)

We have conducted a multiobjective computational optimization of a high brightness, high average current photoinjector under development at Cornell University. This injector employs a dc photoemission electron gun. Using evolutionary algorithms combined with parallel computing resources, the multivariate parameter space of the photoinjector was explored for optimal values. This powerful computational tool allows an extensive study of complex and nonlinear systems such as the space-charge dominated regions of an accelerator, and has broad areas of potential application to accelerator physics and engineering problems. In the present case, the optimized injector is simulated to deliver beam of very high quality (e.g., a rms normalized emittance of 0.1 mm mrad for 0.1 nC, and 0.7 mm mrad for 1 nC bunches). The field strengths of the active elements of the injector are moderate and technically practical. The relevance of these results to various novel linac-based accelerator proposals is pointed out.

DOI: 10.1103/PhysRevSTAB.8.034202

PACS numbers: 29.25.Bx

I. INTRODUCTION

The successful operation of a cw energy recovery linac (ERL) with moderate beam current for the Jefferson Laboratory free electron laser [1], combined with the demonstration of reliable high gradient operation of cw superconducting accelerator cavities at several laboratories [2,3], has led to great current interest in developing ERLs for a number of applications. These applications include electron cooling [4], very high power free electron lasers, linac-ring versions of an electron-ion collider [5], and the production of high brightness, short pulse synchrotron radiation x-ray beams [6]. All of these applications require the development of high average current, high brightness electron injectors. Average currents approaching an Ampere, with transverse and longitudinal emittances lower than the present state-of-the-art, are currently being discussed.

Prior to the development of ERLs, the need for high average current electron injectors was limited by the practical problems in accelerating such currents to even moderate beam energies. To date, most high average current electron injectors have employed dc electron guns delivering cw or long pulse beams from gridded thermionic emission cathodes. The beam from these guns is drift bunched, sometimes following rf chopping, subharmonic bunching, or prebunching systems, to produce a bunched beam for subsequent acceleration. In such injectors, the emittance is limited by the relatively high thermal emittance of the cathode, effects due to the grid, and emittance growth during the bunching process.

More recently, photoemission cathodes have been utilized for high brightness electron injectors. In conjunction with suitable short pulse lasers, photoemission cathodes can deliver a bunched beam directly from the electron gun.

The cathode current density can be significantly higher than that from a thermionic cathode, leading to the possibility of high beam brightness. The absence of a grid, and excellent cathode smoothness, eliminate these emittance degrading effects. Negative electron affinity photocathodes offer the possibility of a very small thermal emittance. Electron injectors employing photoemission cathodes have delivered average currents of 9.1 mA in a train of 0.122 nC bunches from a dc gun [7], and 32 mA in a train of 4.74 nC bunches from a 433 MHz rf gun [8].

Photoemission electron guns are operated very differently from thermionic guns. The photocathode is generally neither fully nor uniformly illuminated, in contrast to a thermionic gun, where the full cathode area delivers a relatively uniform current density in space-charge limited flow. Short duration pulses leaving a photocathode fill only a small fraction of the cathode-anode gap, and space-charge forces lengthen the bunch and degrade the emittance from the photocathode onward. Delivering even moderate bunch charges in a very short duration pulse from a small cathode area requires a substantial electric field at the photocathode at the time of photoemission. This reality has led to the development of rf photoinjectors, in which a photoemission cathode is mounted in the end wall of a rf accelerating structure, allowing a very high electric field to be applied to the photocathode at the time of photoemission. The discovery that certain focusing conditions following a photoemission electron gun could recover much of the space-charge induced emittance growth has led to the development of a number of electron injectors providing high peak brightness beams [9].

For true cw applications, room temperature rf guns have limited utility, since the cavity field strength is limited by power dissipation in the cavity walls. Superconducting rf (SRF) guns offer an obvious solution to this problem, and have been demonstrated [10–12], but their practical implementation to deliver high average current and high

*Electronic address: ib38@cornell.edu

†Electronic address: cks26@cornell.edu

brightness beams presents a number of technical challenges. An alternative for cw applications is a dc electron gun, in which field emission from the cathode electrode structure is the principal phenomenon limiting the cathode field strength. The required cathode field strength may be minimized by delivering a relatively long duration bunch from the photocathode, and using conventional drift bunching. By using a low thermal emittance photocathode, the transverse size of the illuminated spot at the cathode can be increased while meeting the total emittance goal. A longer bunch and a larger spot size both reduce the detrimental space-charge effects.

Determining the optimal parameter set for a high average current dc photoemission electron source with drift bunching is not a simple task. In general, one expects the optimal solution to be dependent on the bunch charge, the cathode field strength and gun voltage, the transverse and longitudinal profiles of the laser illumination at the photocathode, and the locations and field strengths of the focusing, bunching, and accelerating elements following the electron gun. There are many practical constraints involving the physical size of the elements, practical field strengths, and the realities of the vacuum system. The nonlinear nature of the space-charge force precludes obtaining meaningful analytic estimates. Present day codes allow good quality results to be obtained in tracking a bunch through a complete injector, but the large number of parameters and constraints involved makes a complete injector optimization formidable.

At Cornell University, we are planning the construction of an ERL-based synchrotron radiation x-ray source. The goals of this project are to produce x-ray beams offering exceptionally high brightness, high coherence, and very short duration x-ray pulses to allow pump-probe experiments. Ultimately, we would like to deliver diffraction limited x-ray beams at 1 Å wavelength, requiring a geometric emittance of 8 pm rad at 5 GeV beam energy. The quality of the beam delivered from the electron injector is a crucial aspect of this project. Our assessment of current technology led us to design an injector based on a very high voltage dc photoemission electron gun, illuminated with moderate duration optical pulses from a cw laser, and followed by conventional drift bunching and an SRF accelerator. We developed a design for such an injector that met our baseline goals of 77 pC per bunch at a bunch repetition rate of 1300 MHz (corresponding to 100 mA average current), with a normalized emittance below 2 mm mrad at 5 GeV, and a bunch length of 0.7 mm rms [13].

With a baseline injector design in hand, it is natural to investigate the extent to which the design might be improved. In this paper, we report the results of an extensive computational optimization of our original injector design. This optimization was conducted using large scale parallel processing, originally with a dedicated cluster computer having 128 2 GHz processors, and ultimately using the

available computational power on the many desktop computers on the laboratory network.

This paper is organized in the following fashion: in Sec. II we describe the development of evolutionary algorithms and their implementation in a parallel computing environment. Then we summarize the parameters used in our optimization problem. In Sec. III we present the improvements obtained in the simulated electron beam properties. In particular, we show how the use of multiobjective evolutionary algorithms helped us address the following questions: What is the optimal transverse and longitudinal shape of the laser pulse? How high should the gun voltage be for good injector performance? How does the thermal emittance of the photocathode affect the final emittance? What are the trade-offs between bunch length, emittance, and bunch charge? We conclude with a brief discussion of our results, and the prospects for future developments using this powerful computational technique.

II. ALGORITHMS

The large number of variables and constraints in the problem under consideration (the field strengths of various elements in the injector, their physical size and position along the beam line, the phases of time dependent fields, and initial beam distributions at the photocathode) and the fact that each simulation through the injector is time consuming poses a challenge for performing any meaningful optimization in a reasonable amount of time. As it is pointed out in [14], there are three ways of doing a job faster, namely, (1) work harder, (2) work smarter, and (3) get help. In the world of computers these translate into (1) higher processor speed, (2) better algorithms, and (3) parallel processing. The fact that the injector optimization involves conflicting objectives (e.g., minimize emittance, maximize bunch charge, minimize bunch length, etc.), and the need for parallelization to increase the throughput make multiobjective evolutionary algorithms a suitable choice for the task. The primary algorithm used to obtain the results presented in this paper is a modified version of SPEA2 (strength Pareto evolutionary algorithm 2) [15]. NSGA-II (nondominated sorting genetic algorithm II) [16] was also used in these optimizations and demonstrated similar performance to SPEA2 when specifying two objective functions. However, it did not give as good a spread of solutions in higher dimensions of objective space.

For the convenience of readers not familiar with evolutionary algorithms, we provide a brief overview of their most salient features and then describe our particular implementation. A vast literature on multiobjective evolutionary algorithms is available. The interested reader is referred to [17,18] for more details on the subject.

A. Multiobjective evolutionary algorithms

We begin by providing several definitions important for describing the way an evolutionary algorithm performs,

and for visualization of optimization results. A multiobjective optimization problem can be formulated as follows:

$$\left. \begin{array}{l} \text{maximize} \quad f_m(x_1, x_2, \dots, x_n), \quad m = 1, 2, \dots, M; \\ \text{subject to} \quad g_j(x_1, x_2, \dots, x_n) \geq 0, \quad j = 1, 2, \dots, J; \\ \quad \quad \quad x_i^{(L)} \leq x_i \leq x_i^{(U)}, \quad i = 1, 2, \dots, n. \end{array} \right\}$$

Objective functions $f(\mathbf{x}) \equiv (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_M(\mathbf{x}))$ map n -dimensional *decision variable space* \mathcal{D} of vector $\mathbf{x} \equiv (x_1, x_2, \dots, x_n)$, bounded with lower and upper limits $\mathbf{x}^{(L)}$ and $\mathbf{x}^{(U)}$, respectively, into M -dimensional objective space \mathcal{Z} . The goal is to find all (or as many as possible) solutions \mathbf{x} that maximize each objective function $f_m(\mathbf{x})$ subject to J inequality constraints $g_j(\mathbf{x}) \geq 0$. Here we do not consider equality constraints, which are more difficult to deal with in practice. A solution \mathbf{x} that satisfies all constraints and variable bounds is called *feasible* (as opposed to *infeasible*) and the set of all feasible solutions represents the *feasible region* S in the decision variable space \mathcal{D} .

An important concept that allows one to compare feasible solutions with multiple objectives is that of *domination*, which is defined as follows:

Definition 1. A feasible solution \mathbf{x}_a is said to dominate another feasible solution \mathbf{x}_b , or $\mathbf{x}_a \succ \mathbf{x}_b$, if the solution \mathbf{x}_a is not worse than \mathbf{x}_b in all objectives and \mathbf{x}_a is strictly better than \mathbf{x}_b in at least one objective. In other words, $\forall m \in 1, 2, \dots, M : f_m(\mathbf{x}_a) \geq f_m(\mathbf{x}_b)$ and $\exists m' \in 1, 2, \dots, M : f_{m'}(\mathbf{x}_a) > f_{m'}(\mathbf{x}_b)$.

The dominance relation is transitive (i.e. $\mathbf{x}_a \succ \mathbf{x}_c$ if $\mathbf{x}_a \succ \mathbf{x}_b$ and $\mathbf{x}_b \succ \mathbf{x}_c$) and this property allows a binary relation as an ordering operator (since the dominance relation is not reflexive, i.e., no solution dominates itself, a set that is ordered in this fashion represents an instance of what is known in mathematical literature as *quasiorder*). Thus, the concept of dominance allows sorting of trial solutions during multiobjective optimization. In particular, one is interested in finding a *nondominated set*:

Definition 2. Among a set of solutions \mathcal{P} , the *nondominated subset of solutions* \mathcal{P}' are those that are not dominated by any member of the set \mathcal{P} .

The better solutions will comprise the nondominated set, and provide direction for the optimization search. When the set \mathcal{P} is the entire search space, the resulting nondominated set is called the *Pareto-optimal set*. Clearly, if one is able to find the Pareto-optimal set of solutions, the multiobjective optimization problem is solved.

In Definition 1 the two solutions \mathbf{x}_a and \mathbf{x}_b already belong to the feasible region S . In general, however, finding the region of feasible solutions can be a challenge in itself. Using the definition of *constrain-domination* allows classification of trial solutions independent of whether the solutions are feasible or infeasible.

Definition 3. A solution \mathbf{x}_a is said to *constrain-dominate* a solution \mathbf{x}_b , or $\mathbf{x}_a \succ_c \mathbf{x}_b$, if any of the following conditions are true:

1. Solution \mathbf{x}_a is feasible and solution \mathbf{x}_b is infeasible.
2. Solutions \mathbf{x}_a and \mathbf{x}_b are both infeasible, but solution \mathbf{x}_a has smaller constraint violation, i.e., dominates \mathbf{x}_b with respect to constraint functions for $\forall j' \in 1, 2, \dots, J : g_{j'}(\mathbf{x}_a) < 0$ or $g_{j'}(\mathbf{x}_b) < 0$.
3. Solutions \mathbf{x}_a and \mathbf{x}_b are both feasible and \mathbf{x}_a dominates \mathbf{x}_b according to Definition 1.

In the early stages of optimization, the constrain-dominance relation tends to emphasize feasible solutions over infeasible solutions, while at a later stage, when most trial solutions are found in the feasible region S , the dominance relation becomes more prominent, favoring solutions with better objective functions.

Evolutionary (or genetic) algorithms vary widely in their particulars, but they share several common properties. These include working with a population of solutions rather than a single solution, and using stochastic operators to both emphasize better solutions and to create new trial solutions. The former makes these methods well-suited for optimizations with more than one objective, i.e., a multiobjective optimization problem, while the latter makes them generally applicable to solving a broad range of problems, since no particular structure of the problem is assumed in the algorithms. Furthermore, these methods do not use any gradient information, avoiding the potentially noisy calculation of derivatives, and allowing their use in cases where gradient information is not readily available. Finally, because these algorithms work with a population of solutions, they can be easily adapted for parallel processing.

A typical evolutionary algorithm works as follows. A large number (e.g., a few hundred) of initial trial solutions is generated by randomly assigning vectors \mathbf{x} in space \mathcal{D} . Objective functions $f(\mathbf{x})$ and constraints $g(\mathbf{x})$ are evaluated for this population. A fitness value is assigned to each member of the population based on a comparison relation such as constrain-domination. A selection operator is then applied to create a mating pool of some fixed size with a preference for solutions with better fitness. Thus, better fitness solutions tend to have more copies in the mating pool than less desirable solutions. A crossing operator then generates new offspring solutions from parent solutions in the mating pool. An optional mutation operator may be used to diversify the offspring by perturbing the new offspring solutions in some fashion. The offspring solutions generated by these selection, crossing, and mutation operations are carried over to the next generation and the process is repeated. The ideas underlying these operations are that solutions from the full decision space will be adequately sampled, and that better fitness solutions will be selected, tending to produce more successful offspring and moving the solutions toward the true Pareto-optimal front.

In elitist evolutionary algorithms, such as NSGA-II and SPEA2, the selection operator is elite-preserving, i.e., the

best solutions (e.g., the current nondominated set) from among both parents or some externally maintained archive, and the offspring are carried over to the subsequent generations. Elitism ensures that good solutions found previously do not get lost in later generations and always get a chance for representation in the mating pool, improving the overall performance of the evolutionary algorithm. Furthermore, to ensure that the final nondominated set contains solutions that are well distributed in objective Z or solution \mathcal{D} spaces, the selection operator is chosen so that solutions from less crowded regions, which are otherwise equivalent with respect to the dominance relation, are emphasized over ones from more crowded regions.

B. Implementation

The computational bottleneck in injector optimization is the calculation of objective and constraint functions, as this step requires particle tracking through the injector with space charge included. To reduce the wall-clock time for these calculations to a reasonable value, parallel processing is used.

Evaluation of objectives and constraints of a population of solutions from a single generation is ideal for parallel processing. Computation of the objective and constraint functions of a particular trial solution is done on a single processor. The message passing between various processes occurs only at the end of each generation, when selection and crossing of the individuals in the whole population takes place. Thus, parallelization of evolutionary algorithms does not require high-bandwidth low-latency interconnections of the various nodes in a cluster, and even computers that are part of a usual network form an effective parallel environment for doing optimizations.

A single evaluation of a new trial solution requires a variable length of time, due, for example, to the varying complexity of its data set, or to the processor speed or different work loads on individual computers. It is thus inevitable that near the end of the computations for each generation, a considerable fraction of the computing nodes would be waiting for the last exchange of information, from those few evaluations taking the longest time, with the main processor responsible for performing the evolutionary algorithm operations. To circumvent this undesirable situation, the algorithm begins running a new generation by applying a premature selection from a subset of the evaluated fraction of the population, sending the new offspring trial solutions to inactive nodes for evaluation until no idle node remains. This way, more trial solutions than the number of available nodes are generated in the beginning, and the solutions which have not been evaluated are kept in a waiting queue. As soon as a node finishes its previous job and becomes available, a trial solution from this queue is sent to it for evaluation. When the waiting queue becomes empty, the algorithm performs evolutionary operations (selection, crossing, mutation, etc.) on the

population that now includes the newly evaluated solutions. This completes the cycle and a new generation is started. This way the idle time in the system is dramatically reduced, and the longer time required for the evaluation of some trial solutions does not retard the overall performance.

Parallel implementation of the evolutionary algorithms was realized on two 64 and 32 dual-processor cluster computers, as well as nearly 100 desktop computers within the laboratory's network. The latter were utilized as a latent computational resource, performing trial solution evaluations in the background when their normal work load was minimal. A "master-slave" model for algorithm execution was used: the master processor performs all the evolutionary operations on the ordered population of evaluated trial solutions, and sends the trial solutions to the slaves for evaluation.

It should be noted that in doing this parallelization, the association of a particular solution with a particular generation is lost, though this is without consequence to the final optimization.

A Platform and Programming Language Independent Interface for Search Algorithms (PISA) library [19] provided the framework for using existing evolutionary algorithms as well as for developing new ones. The algorithms of the library were modified to include constraint handling through the constrain-domination relation. The underlying feature of PISA is to separate algorithms into two loosely coupled parts: variator and selector. The selector contains algorithm specific operations, such as assigning fitness to an entire population and creating a new mating pool with better solutions from this population based on a specific algorithm. The variator, on the other hand, contains problem specific implementation (e.g., provides objectives and constraints for new decision variables) as well as crossing and mutation operators acting on the individuals chosen by the selector and passed to the variator. Parallelization involves modifying the variator only, which can then be combined with any particular version of the selector to form a working optimization algorithm. This way one is able to apply various evolutionary algorithms to the problem without the need to reprogram.

We used the program ASTRA [20] as the space-charge code to evaluate the objective and constraint functions. This choice was motivated by the good trade-off between speed and performance available in ASTRA, as well as the "low-maintenance" nature of the code (e.g., automatic mesh calculation).

C. Problem setup

The injector system under study is described in [13]. A schematic layout of the various beam line elements is shown in Fig. 1. The high-voltage dc gun is followed by two focusing solenoids, one before and one after a single-

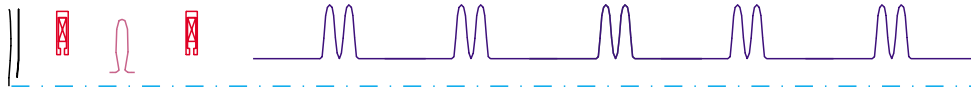


FIG. 1. (Color) Schematic of the dc gun injector layout.

cell, normal conducting buncher cavity. A cryomodule containing five 2-cell superconducting cavities [21], capable of accelerating the beam to energies as high as 15 MeV, is followed by a 3 m drift region. The fundamental rf frequency is 1.3 GHz. The entire injector is assumed to be axially symmetric.

For the injector optimizations reported here, we have used a total of 22 decision variables. In all cases, the longitudinal separation between elements has been constrained to allow for the assembly of a physically realistic vacuum system, and for realities such as the transition from room temperature to 2 K at the entrance and exit of the cryomodule. Longitudinal positions are measured with respect to the cathode, and the cathode-anode separation is fixed in all simulations. Four variables specify the longitudinal position of the solenoids, the buncher, and the first superconducting cavity. The distances between subsequent cavities are determined by the thermal load on the 2 K refrigerator and the need to minimize the coupling to the higher order mode (HOM) loads at the fundamental frequency [22]. The accelerating gradient and phase of the first cavity in the cryomodule were specified separately from the gradient and phase of the remaining four cavities, which were fixed with respect to one another. Four variables specified the dc gun voltage, the two solenoid fields and the buncher cavity gradient. The longitudinal field profiles in all elements were determined with suitable codes (POISSON [23] for static fields and SLANS [24] for rf fields). Two variables specified the spot size and duration of the laser pulse at the cathode. Two pairs of three parameters defined the transverse and longitudinal laser profiles (see Sec. III C for a full description). Finally, two parameters represented the bunch charge and the effective thermal energy E_{th} of the photocathode. The thermal emittance of the photocathode is $\epsilon_{n,rms} = \sigma\sqrt{E_{th}/mc^2}$, with σ and mc^2 being the rms laser spot size and the rest mass of electron, respectively.

The objective functions can include any of the parameters obtained from the ASTRA simulations, such as the emittance, bunch length, and energy spread at the end of the beam line (or any other longitudinal position). Because of the difficulty of visualizing optimization results with a large number of objectives, the number of objectives in each optimization run was limited to three at most.

Constraint functions can represent additional requirements, such as an upright longitudinal phase space orientation (by bounding the correlated energy spread), or desired beam matching conditions at the end of the injector, or can impose restrictions to map out a certain parameter space of interest.

In the initial optimization stages, the bunch is populated with a relatively small number of macroparticles (several thousand), and the optimization is run for several hundred generations. The small number of macroparticles (and the correspondingly coarser space-charge mesh) does not allow accurate solutions for the beam properties through the injector, but it does retain their general characteristics, and considerably speeds up complete optimization.

After this initial run is completed, most of the solutions in the last generation already belong to the feasible region and retain the coarse features found in the true Pareto-optimal front set. This set of decision variables is carried over to the second stage of the optimization with a similar number of generations, in which the bunch is populated with a larger number of macroparticles, and the solutions are further refined. The results presented in this paper correspond to 3×10^4 macroparticles in the bunch, with the number of longitudinal and radial mesh divisions set to 75 and 37, respectively. The integration step was set to 0.3 ps near the photocathode and 3 ps downstream.

III. RESULTS

The ERL will be operated over a wide range of parameters to meet the diverse needs of x-ray scientists. In particular, for experiments requiring high (transverse) coherence one seeks to maximize the average brilliance of the beam, while pump-probe experiments primarily require a large photon flux during a short (subpicosecond) duration pulse. The former requirement favors lower bunch charges at maximum repetition rates (GHz), while the latter requires larger bunch charges at a more moderate repetition rate (MHz). We present most results of our optimization for two choices of bunch charge: 80 pC and 0.8 nC. At the end of this section we show the calculated performance of the optimized injector over the range of bunch charges from 10 pC to 1 nC.

A. Emittance versus bunch length

For the production of synchrotron radiation x rays from an undulator, the trade-off between emittance and bunch length is not *a priori* clear. While the space-charge forces are less severe as the bunch becomes longer, the energy spread in the linear accelerator is worse for longer bunches, decreasing the beam monochromaticity and the effectiveness of an undulator with a large number of periods [25]. Knowing the dependence of emittance at the end of the injector for different bunch lengths allows one to make an optimal choice of this parameter for a particular applica-

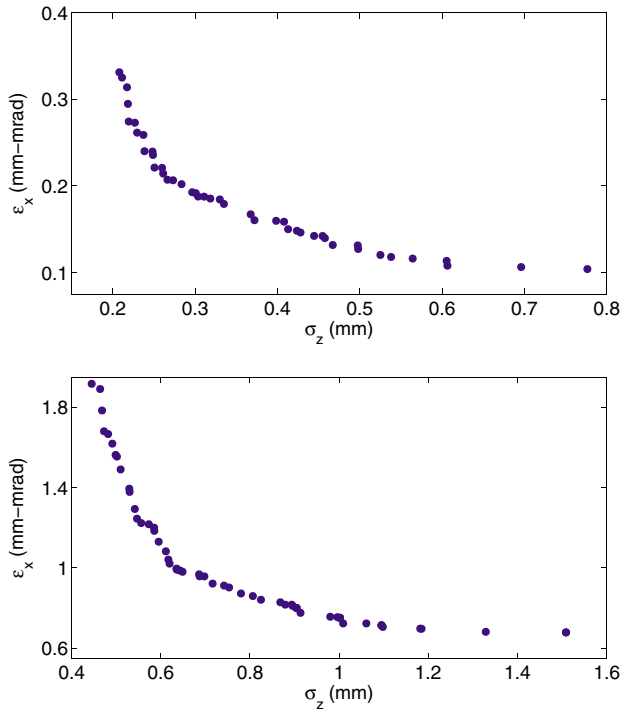


FIG. 2. (Color) Normalized rms emittance versus bunch length at the end of the injector for (top) 80 pC and (bottom) 0.8 nC bunch charges.

tion by taking into account emittance-diluting effects downstream of the injector.

Figure 2 shows the minimization of two objectives: the transverse emittance and bunch length at the end of the injector for 80 pC (upper plot) and 0.8 nC (lower plot) bunch charge. An additional constraint required that the longitudinal phase space ellipse be upright. Figures 3 and 4 show the beam evolution along the injector beam line corresponding to one of the cases in Fig. 2 for either charge.

The average dc gun voltage in these optimizations was 610 and 810 kV for 80 pC and 0.8 nC, respectively. The effective thermal energy of the cathode here and in all other simulations in this paper (unless specified otherwise) is 35 meV—the measured value for room temperature GaAs cathodes illuminated with near band gap energy photons [26]. The average laser pulse duration at the cathode was 19 and 15 ps rms and the average energy of the beam at the end of the injector was 12.6 and 13.2 MeV for the low and high charge cases, respectively. The average values of dc gun voltage and initial bunch duration are affected by the choice of the initial seed of successful solutions found in previous runs and used in these optimizations. We point out in passing that the brightness of the beam is several times higher in the low charge per bunch case.

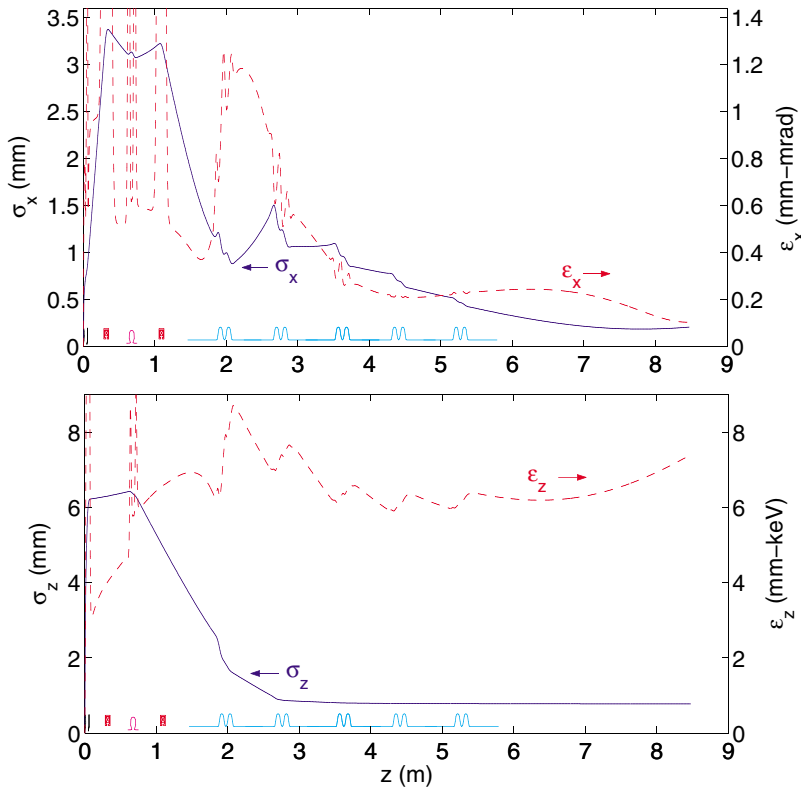


FIG. 3. (Color) Beam evolution in the injector for 80 pC bunch charge: normalized transverse (top) and longitudinal (bottom) rms emittances (dashed line) and sizes (solid line) versus position in the injector.

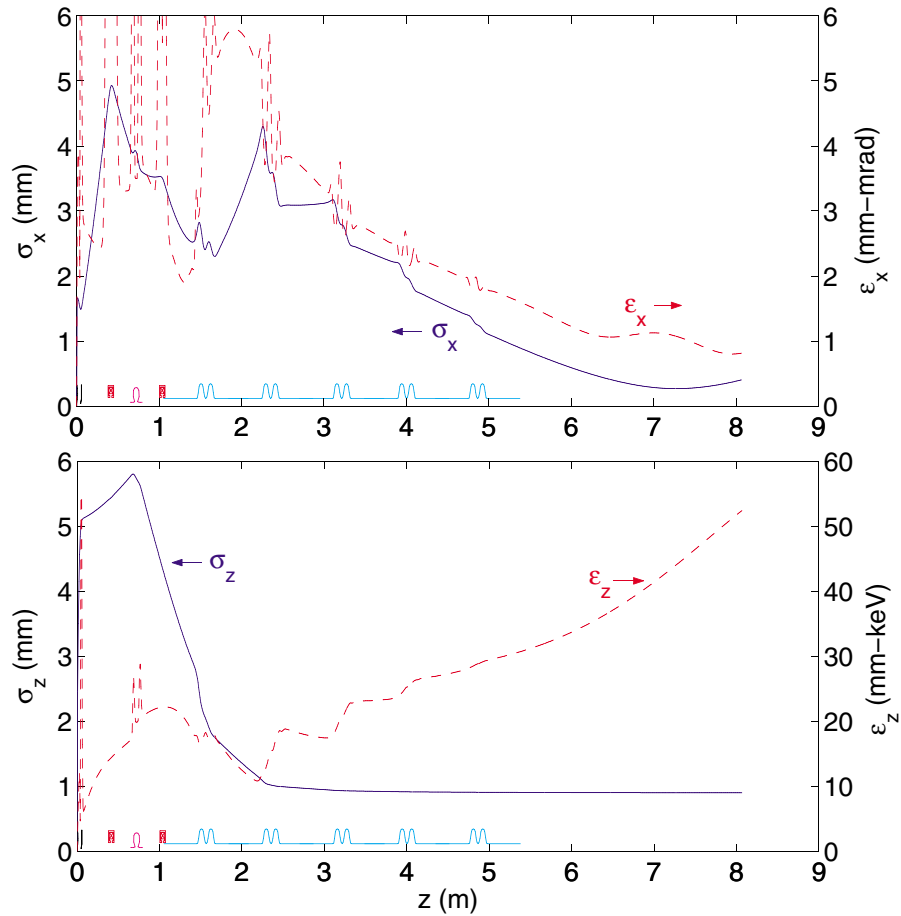


FIG. 4. (Color) Beam evolution in the injector for 0.8 nC bunch charge: normalized transverse (top) and longitudinal (bottom) rms emittances (dashed line) and sizes (solid line) versus position in the injector.

The beam evolution through the injector exhibits several interesting features [27]. While the bunch downstream of the dc gun has a nearly symmetrical ellipsoidal shape, during compression the bunch begins to develop a teardrop-like shape with a more tightly focused tail. After the beam exits the cryomodule, the slice emittances in the larger diameter region of the bunch are aligned but are mismatched with those in the bunch tail. The final alignment of the slice emittances takes place in the drift region following the cryomodule.

B. Gun voltage

It is widely believed that a very high gun voltage and electric field at the photocathode is critical to achieving good beam properties. We carried out a two-objective optimization, in which a minimum transverse emittance was sought while the gun voltage was being minimized at the same time. The cathode-anode geometry was kept fixed during this optimization, so an increased gun voltage results in a similarly increased cathode electric field strength. An additional constraint picked only those solutions that had the rms bunch length at the end of the injector less than

0.9 mm. The resulting nondominated front is shown in Fig. 5.

The longitudinal electric field on the axis of the gun is shown in Fig. 6. The gun has a 25° angle Pierce-like cathode electrode to provide transverse focusing.

The gaps in Fig. 5 (upper plot) are algorithm specific and reflect the following fact. Optimum values of most decision variables depend strongly on the gun voltage, and a good spread in this objective (i.e., the voltage) implies a similar spread in the decision space. On the other hand, crossing operators work best when a successful offspring does not differ dramatically from the parents, thus making it difficult for the algorithm to uniformly fill the objective space. To reduce this effect one has to increase the population size and the number of generations.

The salient feature of Fig. 5 is clear—a higher gun voltage is important to obtain low emittance only up to a certain point, after which the dependence on gun voltage is relatively small. We note that the gun voltage required to obtain small emittances is well within the specifications of our injector system. Our injector will use a gun voltage between 500 and 750 kV, which is above the values presently achieved with dc photoemission guns. We have in-

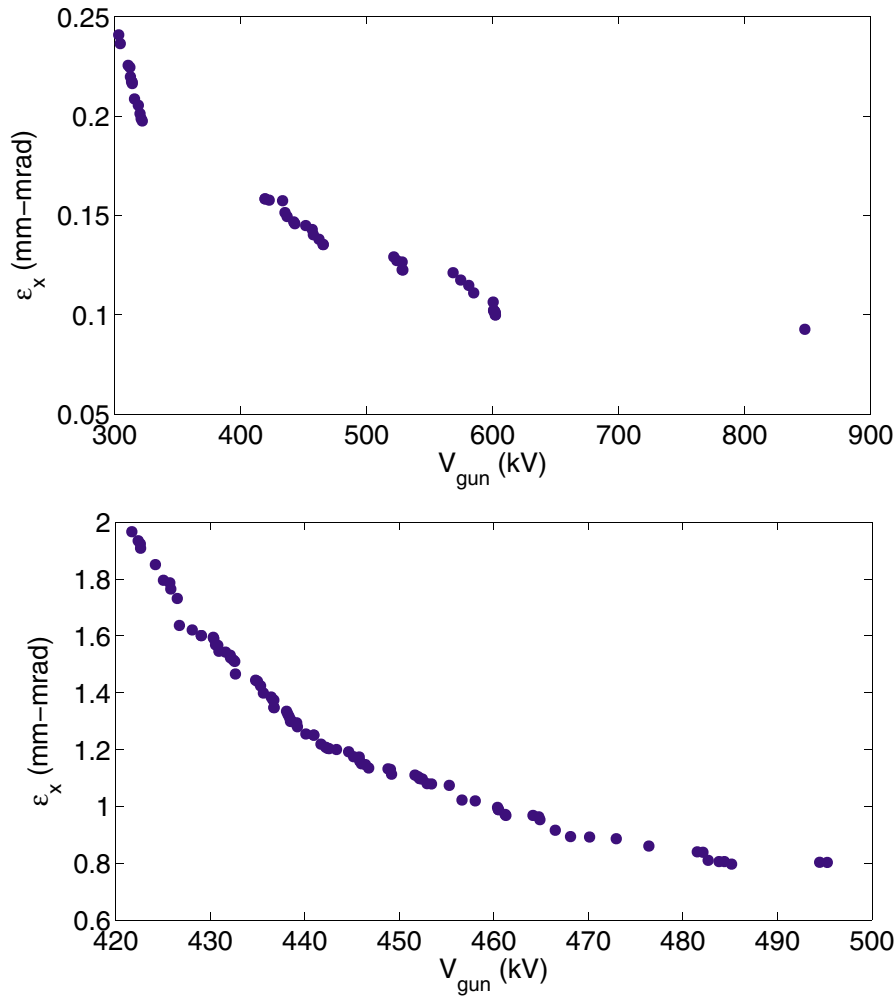


FIG. 5. (Color) Normalized rms emittance versus voltage in the gun for (top) 80 pC and (bottom) 0.8 nC bunch charges. The average bunch length corresponding to these calculations was (top) 0.8 mm and (bottom) 0.9 mm.

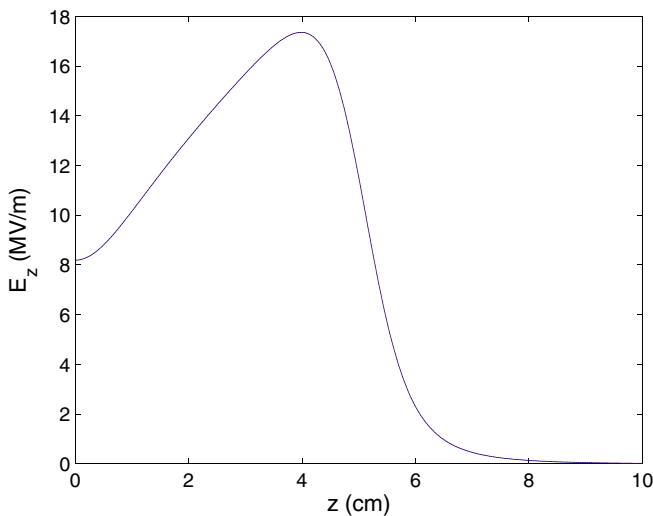


FIG. 6. (Color) Axial electric field in the dc gun for 750 kV voltage.

creased the physical electrode dimensions over those in use with the Jefferson Lab gun [7] to make operation at 500 kV comparable to their present operation at 350 kV. In addition, we anticipate using dielectric coatings on the electrode surfaces to greatly reduce field emission, which is the primary limiting phenomenon in very high voltage guns [28].

It is interesting to observe that the thermal emittance in Fig. 5 on average accounts for 77% and 58% of the final emittance for 80 pC and 0.8 nC bunch charges, respectively (cf. Sec. III D).

It is important to point out, however, that the optimized distances between the elements before the cryomodule are rather crowded at lower gun voltages. For example, the distance from the photocathode to the center of the first SRF cavity is only 1.25 m for 80 pC at 300 kV gun voltage. The spacing of the elements becomes more relaxed in an almost linear fashion as the gun voltage is increased, e.g., the same distance to the first SRF cavity is 2.3 m for 850 kV.

C. Optimal initial distribution

Proper shaping of the laser pulse illuminating the photocathode, both transversely and longitudinally, is known to be important for obtaining a minimum emittance [29–31]. While basic considerations might lead one to believe that the desired electron distribution should make the space-charge forces linear, the fact that the bunch evolves in a nonlinear fashion in the nonrelativistic region near the cathode under the influence of the image space charge, and during drift bunching downstream of the bunching cavity, make it difficult to predict the optimal shape of the laser distribution incident on the photocathode. On the other hand, the question is well posed for optimization such as we perform in this study.

Six decision variables over the interval $[0, 1]$ specified the transverse and longitudinal laser profiles, viz., a tail parameter specified the fraction of the total width of the profile occupied by tails as opposed to a flattop region (0 corresponding to uniform, and 1 to Gaussian shapes); a dip parameter allowed creating profiles depleted in the center; and an additional shape parameter, which, when set to 1 corresponds to a half-circle or, when set to 0, corresponds to a top hat. The full profile is obtained by multiplication of the three respective parts.

Figure 7 shows typical optimum longitudinal and transverse profiles obtained from an emittance minimization for the two charge cases. It is interesting to note the central depletion region in the longitudinal profile for the 80 pC case and the fact that the transverse profile is intermediate between elliptical and flattop shapes in both cases.

Of the two profiles, the longitudinal shape is technically the more difficult to produce. Figures 8 and 9 show the sensitivity of the emittance to the tail parameter of the longitudinal distribution for the 80 pC and 0.8 nC bunch charge cases, respectively. The final bunch length was about 0.9 mm for both cases. The solid line shows the emittance change as the longitudinal profile is changed

with all other parameters in the injector kept fixed. The two arrows indicate the minimum emittance values obtained with flattop transverse and longitudinal profiles, and with a flattop longitudinal profile and the transverse shape shown in Fig. 7. The data points show the minimum emittance obtained after the injector settings were optimized for each particular longitudinal tail parameter value. These results show that while there are longitudinal and transverse distributions at the photocathode that produce a minimum emittance, different injector settings can give similarly small emittances until the longitudinal tails become quite significant. Similar results are expected for deviations from the transverse profile that gives a minimum emittance.

D. Thermal energy of the photocathode

The choice of the most suitable photocathode involves many important considerations, such as the cathode operational longevity, its sensitivity to the vacuum environment, its temporal response, and the effective thermal energy of electron distribution leaving the cathode. The latter parameter combined with the transverse size of the illuminated area on the photocathode determines the minimum transverse emittance.

Figure 10 shows the emittance from our injector as a function of the effective thermal energy at the photocathode. The bunch length was constrained to be less than 0.9 mm in these results. The gun voltage varied between 607 and 622 kV for the lower charge and between 808 and 847 kV for the higher charge per bunch. These simulations emphasize the importance of the effective thermal energy of the photocathode to the ultimate beam brightness. In particular, it is often stated that in the case of a large charge per bunch, where the emittance is dominated by space-charge effects, the thermal emittance from the photocathode is unimportant. These results clearly show this statement is not so. The underlying reason is that a smaller

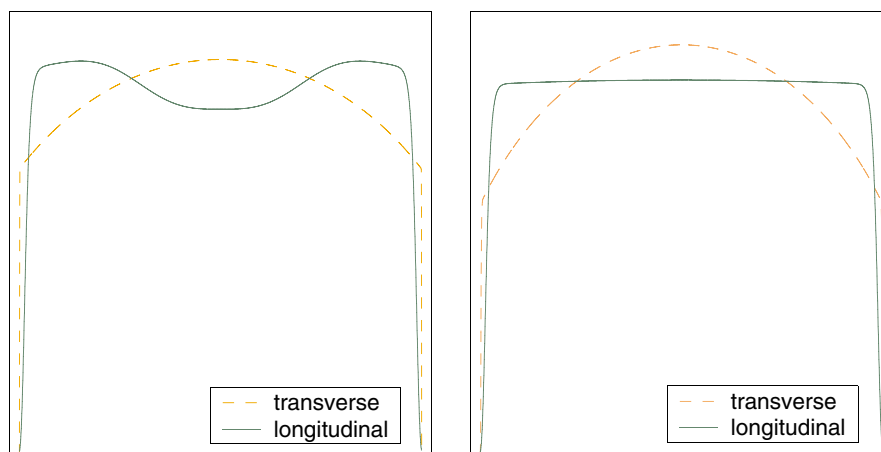


FIG. 7. (Color) Initial distribution profiles corresponding to minimum emittance at the end of the injector for (left) 80 pC and (right) 0.8 nC cases.

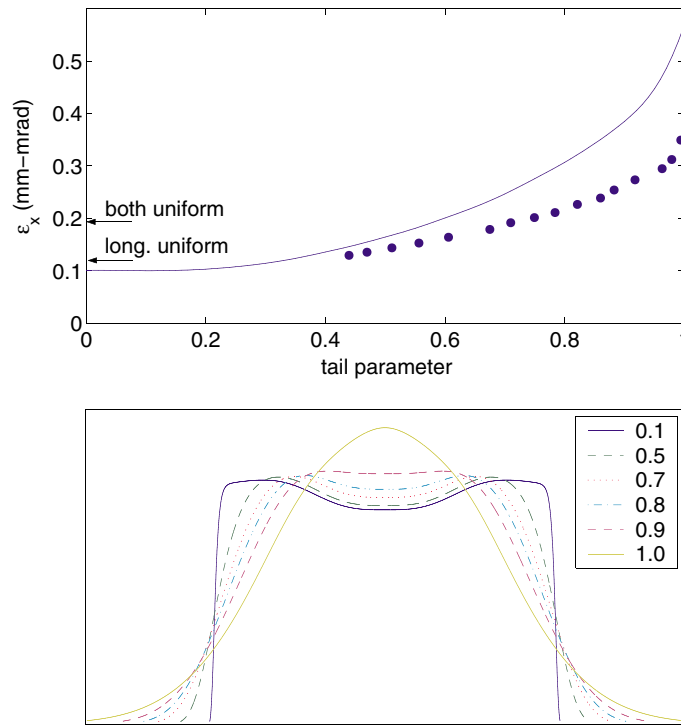


FIG. 8. (Color) 80 pC: emittance sensitivity (solid curve) to the longitudinal profile changes (top) and the corresponding profile shapes (bottom). Data points show the minimum normalized rms emittance obtained after the injector settings were reoptimized for each particular longitudinal tail parameter value.

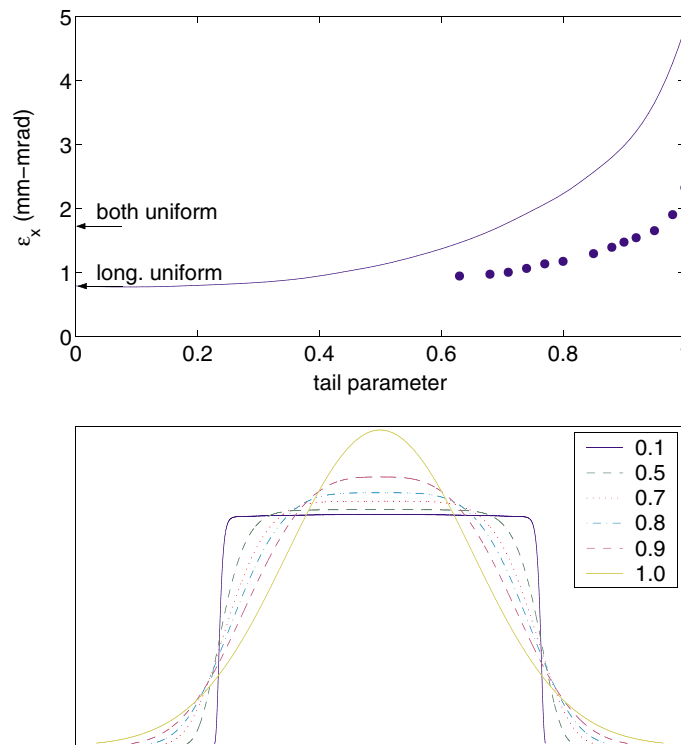


FIG. 9. (Color) 0.8 nC: emittance sensitivity (solid curve) to the longitudinal profile changes (top) and the corresponding profile shapes (bottom). Data points show the minimum normalized rms emittance obtained after the injector settings were reoptimized for each particular longitudinal tail parameter value.

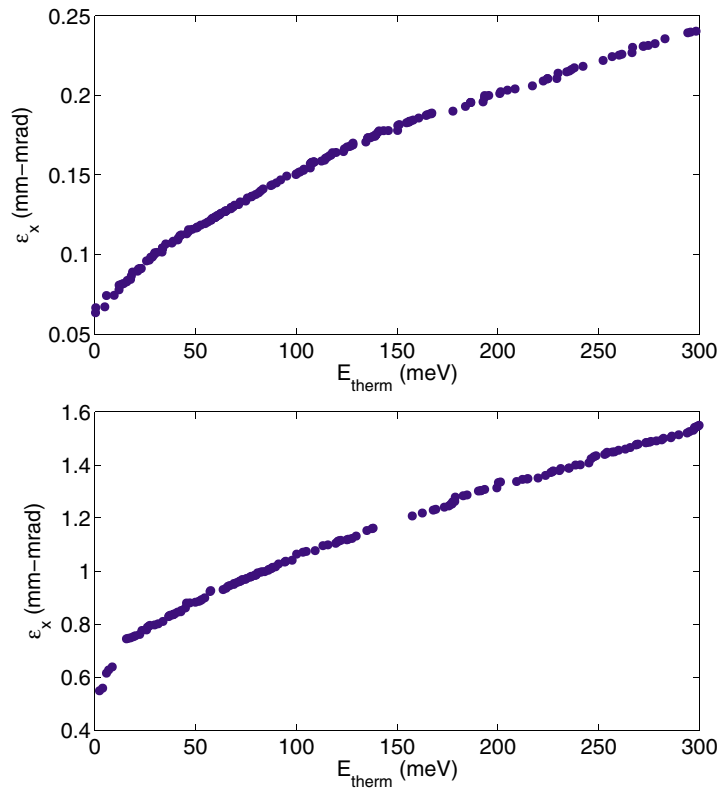


FIG. 10. (Color) The minimum transverse normalized rms emittance from injector as a function of the effective thermal energy from the photocathode for (top) 80 pC and (bottom) 0.8 nC bunch charge, respectively.

effective thermal energy allows the illuminated spot size at the photocathode to be larger, reducing the effects of space charge.

E. Emittance versus bunch length versus charge

Choosing the optimal parameters for an injector delivering a particular charge per bunch is a complex task. Often the exact trade-offs affecting the final beam quality are not known, and frequently (as in the case of an injector for an ERL light source) the beam quality requirements depend on the particular application. Knowing the Pareto-optimal front provides one with invaluable information for choosing those injector parameters that will be fixed by design, and provides guidance for selecting those parameters that are variable.

We performed multivariate optimizations of our injector design with three objectives: emittance minimization, bunch length minimization, and bunch charge maximization. The result of these optimizations is shown in Fig. 11. The population size for this optimization had to be increased because of higher dimensionality of the problem (3 objective functions). The optimization required a total of about 10^5 injector simulations. The data in Fig. 11 can be well approximated by $\epsilon_x = q(0.73 + 0.15/\sigma_z^{2.3})$, where ϵ_x is in mm mrad, q is the bunch charge in nC, and σ_z is the rms bunch length in mm. Figure 12 shows the longitudinal emittance for the same data.

In these optimizations the positions of various elements were allowed to change (with the shortest distance between the photocathode and the first SRF cavity limited to no less than 1 m). The general tendency is that for higher charge per bunch the optimal beam element locations move closer to the gun and the spot size at the photocathode increases, as can be intuitively expected, to minimize space-charge

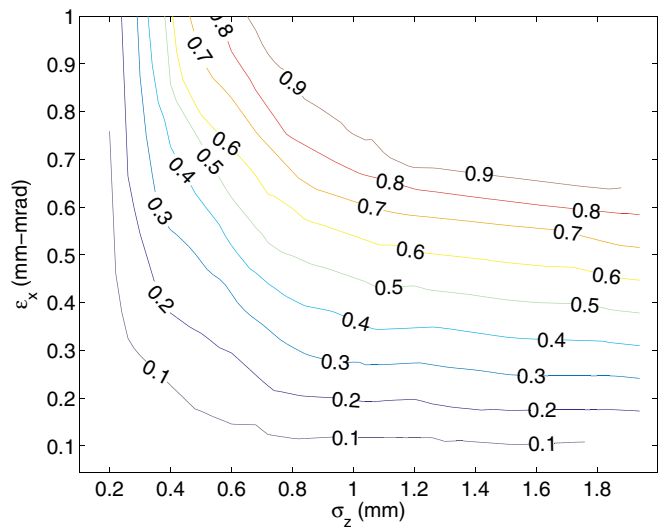


FIG. 11. (Color) Transverse normalized rms emittance versus bunch length for various charges in the injector (nC).

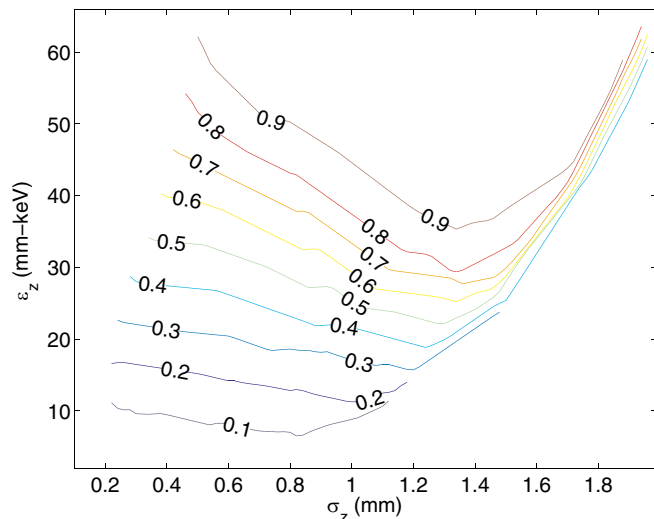


FIG. 12. (Color) Longitudinal rms emittance versus bunch length for various charges in the injector (nC).

effects at low energy. When the positions were kept fixed at the locations corresponding to the best emittance values for 0.1 nC (the distance to the first SRF cavity is 1.9 m) and the gun voltage was fixed to the design maximum value of 750 kV, the optimizer nonetheless was able to find good solutions for higher bunch charges. The emittances obtained were at most only about 30% larger than in the case where these elements were allowed to move. This suggests that the injector should perform well over a wide range of charge per bunch with fixed element locations.

IV. DISCUSSION AND OUTLOOK

It is interesting to point out several salient features of optimized injector settings. It is intuitive that the optimizer would seek to minimize contributions to the final emittance coming from the three major sources: thermal energy of the photocathode, space charge, and rf fields.

Table I provides a comparison between the parameters of the original injector design reported earlier in [13] and this work. The electric field distribution in the gun is different in the two cases, and the anode-cathode gap is smaller by 40% in this work. The laser distribution used in [13] was a temporal Gaussian, and a $\beta = 15$ Fermi-Dirac shape transversely, while the present work has the optimum distributions found. In comparing the two sets of injector parameters, we see that the illuminated spot on the cathode is smaller and the laser pulse duration is shorter in the present case. The focusing provided by the cathode electrode in the gun and by the first solenoid is stronger. Finally, the first superconducting cavity is operated at lower gradient, and well off crest in the bunching direction. This latter change reduces the rf emittance degradation from the first cavity, and results in a more adiabatic bunch-

TABLE I. Comparison of the injector parameters between [13] and this work.

Parameter	Ref. [13]	This work	Units
Charge	80	80	pC
Laser spot size (rms)	0.6	0.3	mm
Laser pulse duration (rms)	20	11	ps
dc gun voltage	500	750	kV
Buncher voltage	116	126	kV
SRF cavity 1 gradient	9.8	5.5	MV/m
SRF cavities 2–5 gradient	7.2	10.6	MV/m
SRF cavity 1 phase	10	43	°
Solenoid 1 peak field	0.058	0.077	T
Solenoid 2 peak field	0.040	0.043	T
Solenoid 1 position	0.29	0.26	m
Solenoid 2 position	1.00	1.12	m
Buncher position	0.80	0.57	m
SRF cavity 1 position	1.80	1.90	m
Transverse emittance (rms)	0.82	0.14	mm mrad
Bunch length (rms)	0.80	0.78	mm
Longitudinal emittance (rms)	8.7	6.2	mm keV
Kinetic energy	10.6	12.6	MeV

ing through the injector. Overall, there is no single improvement factor which stands out as the major reason for the reduced emittance.

As one might expect, the optimization produces a pronounced beam waist at the center of the first accelerating cavity, while the beam size at the buncher cavity is larger. The rf emittance contribution from the buncher cavity is comparable to, though smaller than, the contribution from the first accelerating cavity. In our optimization, we forced the last four accelerating cavities to have identical gradients and fixed phases with respect to each other. Thus, these last four cavities are operated very close to on crest. One can speculate that slightly better performance might be obtained by allowing these cavities to have independent gradients and phases to allow the bunching to be even more gentle.

The energy at the end of the injector is still relatively low, and the projected emittance is not fully degenerate at the location we have defined to be the “end” of the injector. Careful matching to the following accelerator section will be important to preserve the low emittance these simulations show to be possible from the injector.

Our optimizations demonstrate that an injector based on drift bunching of the beam from a photoemission cathode in a dc gun is an excellent choice to achieve a very high brightness, high average current electron beam. Excellent beam quality is available from this injector over a broad range of bunch charge. These optimizations further demonstrate that the thermal emittance available from the photocathode is important at all bunch charges. This makes the use of a negative electron affinity (NEA) photocathode, with its very low thermal emittance, an important advan-

tage. By cooling the negative electron affinity photocathode to liquid nitrogen temperatures, it should be possible to improve the beam emittance for a given bunch charge even further [32]. The interplay between the spot size at the photocathode, the thermal emittance, and the unrecoverable degradation from space-charge effects in the region immediately following the photocathode is complex, and not amenable to analytic calculations or estimates. It is here that the optimization calculation may lead to better results by exploring operating parameters that could otherwise be missed. The use of a dc gun allows a relatively long duration pulse from the photocathode, while the use of a low thermal emittance NEA photocathode may result in a larger optimum illuminated spot size at the cathode for a given final emittance. The longer pulse and larger spot size may reduce the degradation from the strong space-charge effects close to the photocathode. The temporal response of NEA photocathodes depends on the thickness of the cathode and the illumination wavelength and is long compared to the temporal response of alkali antimonide and alkali telluride photocathodes. The values for the temporal response from this optimization are comparable to those reported for bulk GaAs [33], but it will be necessary to explore these parameters experimentally to be fully confident in the results.

The injector design we have optimized employs practical field strengths for all elements, and has specifically incorporated realistic physical constraints to allow the vacuum system to be assembled, and to account for the transition between room temperature and the 2 K environment of the superconducting cavities. We have examined the sensitivity of the final emittance and bunch length to the strengths and locations of the various elements, and find that no unrealistic tolerances are required. The optimization is quite robust, in that when we have constrained various parameters, such as the physical separation of key elements, or moved away from the optimum laser profile at the photocathode, we have been able to reoptimize and obtain results generally close to those of the original optimization.

A question that can be raised with regard to the results presented here is whether the optimum solutions can be reproduced with another space-charge code. This question bears on the more significant issue of the accuracy of the available space-charge codes. Any optimization can only be as good as the model that is being used to evaluate the objectives. To address this issue in a very limited fashion we have compared several of our optimal solutions with results from PARMELA [34]. All of the variable parameters and the field maps corresponding to our optimal performance simulated with ASTRA tracking were directly introduced into PARMELA input files. We find that the agreement in transverse emittance is better than a factor of 2 for all these cases. We believe that much of the remaining difference can be attributed to slight differences in the way the

two codes treat field maps and perform the meshing for the space-charge calculation.

Based on these optimizations, we believe this injector design is an excellent choice for an ERL-based synchrotron light source to produce very high fluxes of coherent hard x rays, and to provide ultrashort x-ray pulses. The x-ray beam characteristics that will be made available from such a light source are vastly superior to those from the best existing or planned storage ring sources. With the proper photocathode and laser choices, this injector will provide beams of highly polarized electrons for potential applications in a linac-ring version of an electron-ion collider, or, perhaps, for the linear collider.

Quite apart from the optimization of the injector reported here, we have demonstrated the great utility of multiobjective optimization algorithms, implemented in a parallel computing environment, to address complex accelerator physics and engineering challenges. By combining parallelized versions of these algorithms with “front-to-end” simulations of a full accelerator, one can obtain the Pareto-optimal front of the most relevant quantities (e.g., brilliance, luminosity, bunch charge, etc.) [35]. This is a dramatic advantage over a “single-point” design, in which important decision variables (e.g., bunch charge) are kept fixed from the outset. We are constructing the injector described here at Cornell University, and will be able to benchmark the results of these optimizations with real beam measurements in the foreseeable future.

ACKNOWLEDGMENTS

We acknowledge MacCHESS for granting us access to their two computer clusters, which were used extensively in this work. In particular, we thank Sol Gruner, Quan Hao, Ernie Fontes, and Frank Labonte of CHESS for their support and assistance. We thank the LEPP computer group and acknowledge our numerous CLEO and LEPP colleagues who have tolerated our using their theoretically latent desktop computers as a parallel processing resource. Igor Senderovich has written the “quiet start” generator of the initial particle distribution, as well as implemented various metrics to characterize performance of different evolutionary algorithms. This work is supported by Cornell University.

-
- [1] G. R. Neil *et al.*, Phys. Rev. Lett. **84**, 662 (2000).
 - [2] B. Aune *et al.*, Phys. Rev. ST Accel. Beams **3**, 092001 (2000).
 - [3] C. Reece, in *Proceedings of the 10th Workshop on RF Superconductivity, 2001, Tsukuba City, Japan*, <http://conference.kek.jp/SRF2001/>
 - [4] I. Ben-Zvi *et al.*, in *Proceedings of the 2003 Particle Accelerator Conference, Portland, OR* (IEEE, Piscataway, NJ, 2003), pp. 39–41.

- [5] Y. S. Derbenev *et al.*, in *Proceedings of the 2004 European Particle Accelerator Conference, Lucerne* (EPS-AG, Lucerne, 2004), pp. 893–895.
- [6] S. M. Gruner *et al.*, *Rev. Sci. Instrum.* **73**, 1402 (2002).
- [7] S. Benson *et al.*, in *Proceedings of the 26th International FEL Conference, Trieste, 2004* (Comitato Conferenze Elettra, Trieste, 2004), pp. 229–232.
- [8] D. H. Dowell *et al.*, *Appl. Phys. Lett.* **63**, 2035 (1993).
- [9] B. E. Carlsten, *Nucl. Instrum. Methods Phys. Res., Sect. A* **285**, 313 (1989).
- [10] J. Teichert *et al.*, in *Proceedings of the 2004 European Particle Accelerator Conference, Lucerne* (Ref. [5]), pp. 333–335.
- [11] A. Michalke *et al.*, in *Proceedings of the 1992 European Particle Accelerator Conference, Berlin* (World Scientific, Singapore, 1992), pp. 1014–1017.
- [12] D. Janssen *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A* **507**, 314 (2003).
- [13] I. Bazarov and C. Sinclair, in *Proceedings of the 2003 Particle Accelerator Conference, Portland, OR* (Ref. [4]), pp. 2062–2064.
- [14] G. F. Pfister, in *Search of Clusters* (Prentice-Hall, NJ, 1998).
- [15] E. Zitzler *et al.*, Swiss Federal Institute of Technology (ETH) Technical Report No. 103, 2001.
- [16] K. Deb *et al.*, *IEEE Trans. Evolut. Comput.* **6**, 182 (2002).
- [17] K. Deb, *Multi-Objective Optimization using Evolutionary Algorithms* (Wiley, Chichester, 2001).
- [18] C. A. C. Coello, D. A. V. Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems* (Kluwer Academic/Plenum, New York, 2002).
- [19] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, in *Evolutionary Multi-Criterion Optimization (EMO 2003)*, edited by C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, *Lecture Notes in Computer Science* Vol. 2632/2003 (Springer-Verlag, Heidelberg, 2003), pp. 494–508.
- [20] K. Flöttmann, *ASTRA: A Space Charge Tracking Algorithm*, http://www.desy.de/mpyflo/Astra_dokumentation/
- [21] V. Shemelin *et al.*, in *Proceedings of the 2003 Particle Accelerator Conference, Portland, OR* (Ref. [4]), pp. 2059–2061.
- [22] H. Padamsee *et al.*, in *Proceedings of the 2004 European Particle Accelerator Conference, Lucerne* (Ref. [5]), pp. 491–493.
- [23] J. Billen and L. Young, Los Alamos Laboratory Technical Report No. LA-UR-96-1834, 2000.
- [24] D. Myakishev and V. Yakovlev, in *Proceedings of the 1995 Particle Accelerator Conference, Dallas* (IEEE, Piscataway, NJ, 1995), pp. 2348–2350.
- [25] I. Bazarov and G. Hoffstaetter, in *Proceedings of the 2003 Particle Accelerator Conference, Portland, OR* (Ref. [4]), pp. 842–844.
- [26] B. M. Dunham *et al.*, in *Proceedings of the 1995 Particle Accelerator Conference, Dallas* (Ref. [24]), pp. 1030–1032.
- [27] Animation of the beam motion in our injector is available at <http://www.lepp.cornell.edu/~ib38/injector>
- [28] C. Sinclair *et al.*, in *Proceedings of the 2001 Particle Accelerator Conference, Chicago* (IEEE, Piscataway, NJ, 2001), pp. 610–612.
- [29] F. Zhou *et al.*, *Phys. Rev. ST Accel. Beams* **5**, 094203 (2002).
- [30] H. Tomizawa *et al.*, in *Proceedings of the 2002 European Particle Accelerator Conference, Paris* (EPS-IGA and CERN, Geneva, 2002), pp. 1819–1821.
- [31] J. Yang *et al.*, *J. Appl. Phys.* **92**, 1608 (2002).
- [32] D. Orlov *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A* **532**, 418 (2004).
- [33] A. V. Aleksandrov *et al.*, *Phys. Rev. E* **51**, 1449 (1995).
- [34] J. Billen and L. Young, Los Alamos Laboratory Technical Report No. LA-UR-96-1835, 2000.
- [35] I. Bazarov and H. Padamsee, LEPP, Cornell University Technical Report No. CBN 05-3, 2005.