

Genetic algorithm for chromaticity correction in diffraction limited storage rings

M. P. Ehrlichman*

Paul Scherrer Institut, Villigen, Aargau 5232, Switzerland

(Received 8 March 2016; published 25 April 2016)

A multiobjective genetic algorithm is developed for optimizing nonlinearities in diffraction limited storage rings. This algorithm determines sextupole and octupole strengths for chromaticity correction that deliver optimized dynamic aperture and beam lifetime. The algorithm makes use of dominance constraints to breed desirable properties into the early generations. The momentum aperture is optimized indirectly by constraining the chromatic tune footprint and optimizing the off-energy dynamic aperture. The result is an effective and computationally efficient technique for correcting chromaticity in a storage ring while maintaining optimal dynamic aperture and beam lifetime.

DOI: 10.1103/PhysRevAccelBeams.19.044001

I. INTRODUCTION

Multiobjective genetic algorithms have found much success providing nonintuitive solutions to problems that are not adequately solved by analytic methods. Such algorithms have been successfully applied to various aspects of accelerator design [1–5]. In this paper, a genetic algorithm is developed for optimizing dynamic aperture and beam lifetime using sextupole and octupole strengths in next-generation diffraction limited storage rings (DLSR).

The one-turn map for a DLSR contains strong nonlinearities. Such machines use stronger focusing and lower dispersion to achieve an emittance that is below the diffraction limit for some wavelengths of light. The strong focusing of these machines leads to a large chromaticity, which must be corrected by placing strong sextupoles in dispersive regions. Furthermore, the low dispersion necessitates that the sextupole strengths be even higher. Such strong sextupole moments add strong nonlinearities to the lattice.

These nonlinearities must be carefully designed to maintain adequate injection efficiency and beam lifetime. Nonlinearities induce tune shifts and increase the sensitivity of particles to machine imperfections. This reduces the dynamic aperture, which is the volume in 6D phase space containing stable particle trajectories. Touschek scattering and residual gas scattering excite the phase space coordinates of particles stored in a ring. A particle is lost if it exits the dynamic aperture. This imposes a beam lifetime. A reduced dynamic aperture also complicates injection by reducing the capture efficiency of the machine.

The problem of nonlinear optics in a storage ring is to correct the chromaticity without inducing nonlinearities that degrade the other lattice properties too much.

The established technique for optimizing nonlinearities in a storage ring is resonant driving term minimization and was developed for the original Swiss Light Source (SLS) [6]. It is a Lie algebra expansion of the transfer map in driving terms that are functions of the sextupole strength. These terms drive higher order chromatic tune shifts, amplitude dependent tune shifts, and resonances. A gradient optimizer is used to minimize a weighted vector of the driving terms. The weights are determined by judging from the tune diagram and frequency maps which resonances and tune shifts are most important. This technique is straightforward to wield up to second order in sextupole strength. Beyond second order, the number of driving terms which need to be minimized makes the method less robust. The complexity of the optimization space necessitates much trial and error to locate good local minima, and the procedure becomes a bit of a dark art.

A storage ring is a complicated nonlinear oscillator. Its transfer map is the concatenation of many hundreds of linear and nonlinear maps. Explicit algorithms yield only an incomplete control of particle motion in a storage ring. This lack of analytic clarity makes accelerators good candidates for genetic algorithms, which typically do not depend on knowledge of the system. Genetic algorithms optimize a system by selectively breeding trial solutions according to their fitness.

One example where a genetic algorithm performs better than perturbation techniques is in confining the chromatic and amplitude-dependent tune shifts. Second order perturbation theory is unable to bend the chromatic and amplitude dependent tune shifts beyond second order. The genetic algorithm developed here fits tune footprints into tighter areas using higher orders of correction.

* michael.ehrlichman@psi.ch

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

A well-designed genetic algorithm will encourage behaviors in the evolving population that, further down the road, lead to solutions with optimal objective functions. For example, the algorithm presented here includes a dominance constraint (see Sec. III C) on the amplitude of the nonlinear dispersion. A small-amplitude nonlinear dispersion is not one of the objective functions, but it is a property that a lattice with good objective functions will have. By applying a dominance constraint to the nonlinear dispersion, we are, in a sense, breeding characteristics into the early generations, that in later generations will yield improved objective functions.

Genetic algorithms for the optimization of storage ring nonlinearities have been developed and evaluated elsewhere [3,7–10]. The present application stands out in its use of dominance constraints to more efficiently evolve the population. The algorithm requires only modest computing resources. On a Linux cluster consisting of 64 E5-2670 xeon cores, it delivers solutions to ten variable problems in one or two days, and 20 variable problems in two or three days. The scheme consistently does as well or better than Lie algebra approaches, and repeated attempts with different random seeds on different lattice variants suggest that the solutions it finds are globally optimal. So the optimization scheme presented here allows for a lattice development cycle on the order of a couple days, and does not consume expensive computing resources. This framework was developed for the Swiss Light Source upgrade project [20].

Section II of this paper gives an overview of the system, listing the components out of which this optimization scheme is built. Section III introduces multiobjective genetic algorithms, and Sec. IV describes the optimization problem, including the physics behind the calculations of the objectives and constraints. In Sec. V the optimization scheme is applied to upgrade prototypes of SLS and the proposed Armenian light source CANDLE [11]. Misalignment studies are conducted on the SLS upgrade lattice in Sec. VI.

II. SYSTEM ARCHITECTURE

The optimizer is built within the PISA framework [12], which specifies that the sorting algorithm be separated from the rest of the optimizer. The sorting algorithm is implemented as a stand-alone binary that communicates with the rest of the optimizer through the filesystem. This separation simplifies the coding and makes it trivial to switch between different sorting algorithms, such as SPEA2 [13] or NSGA2 [14].

We use the APISA variant [5] of PISA. APISA modifies the original PISA framework by supporting dominance constraints. APISA was originally developed for the design of the Cornell Energy Recovery Linac injector.

Accelerator physics calculations are handled by calls to the BMAD library [15]. The top-level coding, which

includes population management and breeding, parallelization, and additional physics calculations, was developed at PSI and is coded in FORTRAN90. The parallelization paradigm is master-slave and is implemented using COARRAYS, which in Intel’s Fortran compiler is implemented as a high-level language on top of MPI.

The cluster management software is Sun Grid Engine. The cluster is composed of several 16-core E5-2670 compute nodes running 64-bit Scientific Linux. Typically four of these nodes are used in an optimization run.

III. MULTIOBJECTIVE GENETIC ALGORITHMS

The multiobjective optimization problem is formulated as [16]

$$\begin{aligned} \text{Minimize} \quad & f_m(\mathbf{x}), \quad m = 1, 2, \dots, M; \\ \text{subject to} \quad & g_j(\mathbf{x}) \geq 0, \quad j = 1, 2, \dots, J; \\ & h_k(\mathbf{x}) = 0, \quad k = 1, 2, \dots, K; \\ & x_i^{(L)} \leq x_i \leq x_i^{(U)} \quad i = 1, 2, \dots, n. \end{aligned} \quad (1)$$

f_m are the objectives, which generally are competing. g_j are inequality constraints and h_k are equality constraints. $x_i^{(L)}$ and $x_i^{(U)}$ are upper and lower bounds on variables. A vector of variable strengths $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ is called an individual.

A genetic algorithm manages a population of individuals. Every individual i in the population is represented by a vector of variables $\mathbf{x}^{(i)}$. For our purposes $\mathbf{x}^{(i)}$ is real valued, but in general it could contain integer, logical, and complex variables. Each individual i has an associated vector of objective values $\mathbf{f}^{(i)}$ and constraint values $\mathbf{g}^{(i)}$ and $\mathbf{h}^{(i)}$.

The output of a multiobjective optimizer is of a different nature than that of a single-objective optimizer. A single objective optimizer, or equivalently, an optimizer which reduces $\mathbf{f}^{(i)}$ to a single value by weighting the individual objectives, gives the user one particular \mathbf{x} that is the best solution to the optimization problem it could find. A multiobjective optimizer, on the other hand, returns a population of \mathbf{x} ’s. This returned population is an optimal surface in the objective space called a Pareto front. For any individual on the Pareto front, no improvement to any one of its objectives can be achieved without worsening the others. The user of a multiobjective optimizer typically applies additional criteria when selecting a particular solution from the Pareto front.

A. Ranking the population

The ranking of individuals in a single objective optimization problem is straightforward: the individual with the better objective value is preferred. Ranking in multiobjective optimization problem is more complicated. The core

concept is the *dominance relationship*, which is a way of comparing any two individuals, say $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$. They are compared by asking the question: “Does $\mathbf{x}^{(i)}$ dominate $\mathbf{x}^{(j)}$?” The dominance relationship is defined as follows [16]. *Definition 1.*—An individual $\mathbf{x}^{(i)}$ is said to dominate another $\mathbf{x}^{(j)}$, if both of the following conditions are true: (i) $\mathbf{x}^{(i)}$ is no worse than $\mathbf{x}^{(j)}$ in all objectives; (ii) $\mathbf{x}^{(i)}$ is strictly better than $\mathbf{x}^{(j)}$ in at least one objective.

A sorting algorithm applies the dominance relationship to sort the population from best to worst. There exist many different sorting algorithms. Two algorithms that we have used are NSGA2 and SPEA2. We obtain similar results with both algorithms, but find that the populations resulting from SPEA2 span the objective space more evenly.

For details on SPEA2 see Ref. [13]. In short summary, dominance is determined for every ordered pair of individuals in the population. Each individual in the population is assigned a strength which is the number of individuals it dominates. Then, each individual is assigned a fitness which is the sum of the strengths of all the individuals that dominate it. A lower fitness is better. A “clumping” penalty is added to this fitness based on the shortest distance (in objective space) of the individual to another individual. This encourages the population to span a wider region of the objective space. Incidentally, the clumping penalty makes it unlikely for any two individuals to have the exact same fitness, even if they are both not dominated by any other individual.

Individuals are ordered according to their fitness value. The lowest ranked individuals are deleted from the population. This is typically the worst half or three-quarters of the population each generation. The population is replenished by mating the surviving individuals.

Mating pairs are determined by drawing integers in a procedure called a *tournament*. Two or more random integers are drawn between 1 and N , where N is the number of surviving individuals. 1 corresponds to the most fit, and N to the least fit. The individual corresponding to the smallest of the drawn integers is chosen for mating. Thus, individuals with better fitness are more likely to reproduce. Its mate is chosen through the same process. This is repeated until enough pairs have been selected to replenish the population. Each pair generates two offspring.

A two-step process generates two new children from each mating pair. The first step is simulated binary crossover [17]. Say the two parents are $\mathbf{x}^{(p1)} = \{x_1^{(p1)}, x_2^{(p1)}, \dots, x_n^{(p1)}\}$ and $\mathbf{x}^{(p2)} = \{x_1^{(p2)}, x_2^{(p2)}, \dots, x_n^{(p2)}\}$. They will produce two offspring, $\mathbf{x}^{(c1)}$ and $\mathbf{x}^{(c2)}$.

Start with variable $x_1^{(i)}$ and draw a random real t between 0 and 1. Compare t to P_c , which is a parameter between 0 and 1 that determines how likely it is that a variable is simply copied from parent to child, as opposed to applying

a stochastic function. If $t > P_c$, then variable $x_1^{(p1)}$ is simply copied to the child $x_1^{(c1)}$, and $x_1^{(p2)}$ copied to $x_1^{(c2)}$. If $t \leq P_c$, then draw another random real q between 0 and 1. Then,

$$\beta_q = \begin{cases} (2q)^\kappa & \text{if } q \leq 0.5 \\ \left(\frac{0.5}{1-q}\right)^\kappa & \text{otherwise,} \end{cases} \quad (2)$$

and

$$x_1^{(c1)} = \frac{1}{2}[(1 + \beta_q)x_1^{(p1)} + (1 - \beta_q)x_1^{(p2)}] \quad (3)$$

$$x_1^{(c2)} = \frac{1}{2}[(1 - \beta_q)x_1^{(p1)} + (1 + \beta_q)x_1^{(p2)}], \quad (4)$$

where κ is a parameter that controls the width of the distribution. As depicted in Fig. 1, smaller values of κ cause the offspring to explore a broader parameter space. A typical value for P_c is 0.8. This process is repeated for all n variables.

Next a mutator is applied separately to $\mathbf{x}^{(c1)}$ and $\mathbf{x}^{(c2)}$. For each variable i in $\mathbf{x}^{(c1)}$, draw a random real t between 0 and 1. Compare t to P_m , a parameter that determines how likely it is for a variable to undergo mutation. If $t > P_m$, the variable is not mutated. If $t \leq P_m$, then draw a second random real m . Variable $x_i^{(c1)}$ is adjusted according to

$$\beta_m = \begin{cases} (2m)^\kappa - 1 & \text{if } m \leq 0.5 \\ 1 - [2(1 - m)]^\kappa & \text{otherwise,} \end{cases} \quad (5)$$

and

$$x_i^{(c1)} = x_i^{(c1)} + \sigma_{\text{mut}} \times \beta_m. \quad (6)$$

A typical value for P_m is n^{-1} , so that on average one variable is mutated per child. σ_{mut} is typically set to about

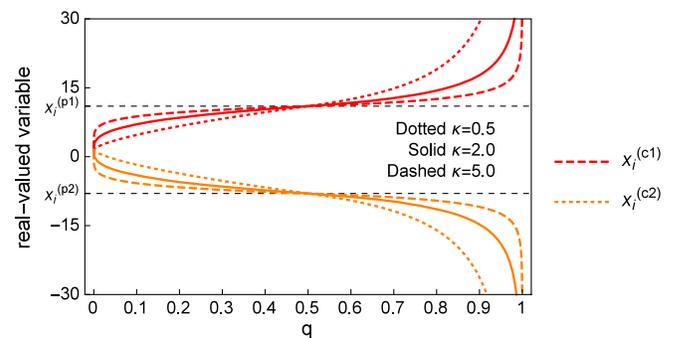


FIG. 1. The crossover operation takes the parent values $x_i^{(p1)}$ and $x_i^{(p2)}$ and a random number q , and returns two offspring values $x_i^{(c1)}$ and $x_i^{(c2)}$. Larger values of the parameter κ result in “near-parent” offspring.

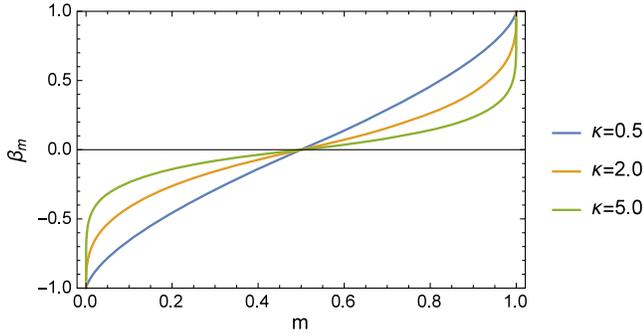


FIG. 2. The mutator operation takes a random real m and adjusts an offspring variable by an amount $M \times \beta_m(m)$. M is a settable parameter which can be customized for each variable type. Larger values of κ make large mutations less likely.

10% of the reasonable variable strength. β_m is depicted in Fig. 2.

After $\mathbf{x}^{(c1)}$, the mutation process is repeated on each variable in $\mathbf{x}^{(c2)}$ and both are added to the population. New individuals are generated until the population is fully replenished.

B. Progression of generations

Once the population is fully replenished, the worker processes calculate the objectives and constraints of the newly generated individuals. The population, consisting of both the parents and children, is resorted and the least fit are deleted. Some of the children will be more fit and displace the older individuals in the surviving population. So is the overall fitness of the population improved. This process of evaluation, sorting, deletion, and replenishment is looped continually. Each loop is referred to as a generation.

As implemented, there is not actually a clear line between one generation and the next. The algorithm is modified to improve computational efficiency. By far, the most time-consuming step is when the objectives and constraints of an individual are calculated. It could often be the case that many nodes in the cluster sit idle while the last few individuals of a given generation are being evaluated. To avoid this situation, the optimizer initially generates extra individuals equal to the number of cores in the cluster. For example, if the population is set to 300 and there are 64 nodes in the cluster, the optimizer will initially generate 364 individuals. Whenever a core finishes evaluating an individual, the individual is added to the population and the core is always immediately given a new unevaluated individual to process. Whenever the population reaches 300, it is resorted, culled, and new individuals are bred. These new individuals replenish the pool of individuals awaiting evaluation. This modification to the algorithm improves cluster efficiency. The loss of a clear demarcation between the generations does not seem to negatively impact the evolution of the population.

C. Dominance constraints

Dominance constraints are a powerful type of constraint that is unique to multiobjective optimization. It is implemented by modifying the dominance relationship. We apply the dominance relationship as implemented in APISA [5].

In addition to calculating objective values $\mathbf{f}(\mathbf{x}^{(i)})$ for each individual i , we also calculate a vector of constraint values $\mathbf{c}(\mathbf{x}^{(i)})$. For example, say c_1 is a dominance constraint for the off-momentum horizontal closed orbit x_{inj} at the injection point, and we wish to constrain $|x_{inj}|$ to be less than x_{max} . Then, $c_1 = x_{max} - |x_{inj}|$. If c_1 is negative, it indicates that the constraint is violated. The magnitude of c_1 represents the degree to which it is violated.

Dominance constraints are implemented by replacing the ordinary definition for dominance in Definition 1 with Definition 2. An individual is called *infeasible* if it violates any of its dominance constraints, else it is called *feasible*. *Definition 2.*— $\mathbf{x}^{(i)}$ is said to constraint dominate an individual $\mathbf{x}^{(j)}$ if any of the following conditions are true: (i) $\mathbf{x}^{(i)}$ is feasible and $\mathbf{x}^{(j)}$ is not; (ii) $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ are both feasible, and $\mathbf{x}^{(i)}$ dominates $\mathbf{x}^{(j)}$ as in Definition 1; (iii) $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ are both infeasible, and both (a) $\mathbf{x}^{(i)}$ is no worse than $\mathbf{x}^{(j)}$ in all constraints and (b) $\mathbf{x}^{(i)}$ is strictly better than $\mathbf{x}^{(j)}$ in at least one constraint.

The behavior of a genetic algorithm implementing dominance constraints flows through three phases: (i) Random population, possibly containing no feasible individuals, spans variable space, sorted according to severity of constraint violations. (ii) Entire population is feasible and spans feasible region of variable space, sorted according to objective values. (iii) Population spans variable space that approaches the Pareto optimal objective space.

Notice that if an individual violates any of its dominance constraints, then its objective values are not taken into account during ranking. Therefore, to save computing time, objective values are calculated only for individuals which do not violate any dominance constraints. Dominance constraints are based on variable bounds, closed orbit amplitudes, and chromatic tune shifts. These quantities are orders of magnitude quicker to evaluate than objective values, which are based on particle tracking.

IV. EVALUATION OF INDIVIDUALS

The design problem is to correct the chromaticity while maintaining acceptable dynamic aperture and beam lifetime. The beam lifetime would be maximized by maximizing the momentum aperture, which is the largest momentum kick that an initially on-axis particle can receive without being lost downstream. The momentum aperture can vary throughout the lattice and is typically calculated element by element or in fixed steps. It is computationally

expensive to calculate, so instead we optimize the off-energy dynamic aperture and apply a dominance constraint to the chromatic tune footprint. The results in Sec. V show that this is an effective proxy for the momentum aperture.

A. Objectives

The dynamic apertures are calculated relative to the linear aperture. The linear aperture is the smallest aperture found by projecting the beam chamber from each point in the lattice to the injection point using the linearization of the map about the particle momentum. The linear aperture, in general, depends on the particle momentum. The on-energy linear aperture does not depend on the sextupole and multipole strengths, but the off-energy linear aperture does. The objective function is formulated relative to the linear aperture such that 1.0 is perfectly bad and 0.0 is perfectly good,

$$f(\mathbf{x}) = \frac{1}{N_{\text{angle}}} \sum_{N_{\text{angle}}} \begin{cases} \left(\frac{L_l(\mathbf{x}) - L_{\text{da}}(\mathbf{x})}{L_l(\mathbf{x})} \right)^2, & \text{if } L_{\text{da}} < L_l \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where N_{angle} is the number of rays along which the aperture is calculated. L_l is the length of the linear aperture ray. L_{da} is the length of the dynamic aperture ray. L_l and L_{da} are depicted in Fig. 3.

The conditional in Eq. (7) reflects our design philosophy that a machine with optimized nonlinearities should behave as if it were linear. The objective function is not rewarded when the dynamic aperture exceeds the linear aperture.

Three objectives are used in our multiobjective optimization problem:

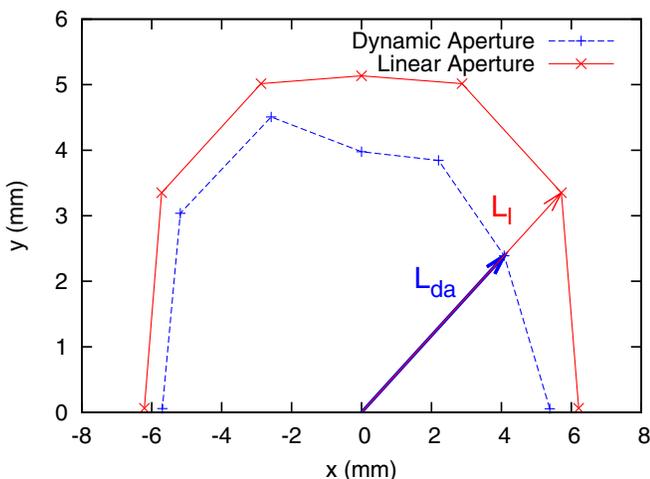


FIG. 3. Dynamic aperture is calculated using an element-by-element tracking code. Particles are tracked for 200 turns. The aperture is found using a binary search for particle loss. In this plot $N_{\text{angle}} = 7$.

$$f_0(\mathbf{x}) = \text{on energy dynamic aperture} \quad (8)$$

$$f_+(\mathbf{x}) = \text{dynamic aperture at } \Delta E_{+\text{DA}} \quad (9)$$

$$f_-(\mathbf{x}) = \text{dynamic aperture at } \Delta E_{-\text{DA}}, \quad (10)$$

where $\Delta E_{+\text{DA}}$ and $\Delta E_{-\text{DA}}$ specify the energy offset where dynamic aperture is evaluated. Typical energy offsets are 2% or 3%.

B. Constraints

Three constraint techniques are used: dominance constraints, modified objective functions, and variable space projection.

Five constraints are implemented as dominance constraints. They are (i) c_{mag} : boundary on sextupole and multipole strength; (ii) c_{+co} : global bound on nonlinear dispersion at $\Delta E_{+\text{DA}}$; (iii) c_{-co} : global bound on nonlinear dispersion at $\Delta E_{-\text{DA}}$; (iv) $c_{+\chi}$: confine chromatic tune footprint between 0 and $\Delta E_{+\chi}$; and (v) $c_{-\chi}$: confine chromatic tune footprint between 0 and $\Delta E_{-\chi}$.

The constraint on sextupole and multipole strength is calculated as

$$c_{\text{mag}} = \sum_i^{\text{magnets}} \begin{cases} \frac{x_i^{(U)} - K_i}{|x_i^{(U)}| + |x_i^{(L)}|}, & K_i > x_i^{(U)} \\ \frac{K_i - x_i^{(L)}}{|x_i^{(U)}| + |x_i^{(L)}|}, & K_i < x_i^{(L)} \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where K_i is the strength of magnet i , and $x_i^{(U)}$ and $x_i^{(L)}$ are the upper and lower bounds on the magnet strength.

The two constraints on global nonlinear dispersion are calculated as

$$c_{\pm co} = \frac{1}{x^{(\text{co})}} \left[x^{(\text{co})} - \max_{i \in \text{all elements}} \text{abs}(x_{\text{co},i} - \eta_i \times \Delta E_{\pm \text{DA}}) \right], \quad (12)$$

where $x^{(\text{co})}$ is the maximum allowed closed orbit, usually set to millimeter or so. $x_{\text{co},i}$ and η_i are the closed orbit and ordinary dispersion at element i .

The positive chromatic footprint between 0 and $\Delta E_{+\chi}$ is constrained to cross neither the half-integer nor the integer resonances. It is calculated by dividing the interval from $\Delta E = 0$ to $\Delta E = \Delta E_{+\chi}$ into N_{χ} equal segments. The horizontal and vertical tunes are calculated by linearizing the optics about each $\Delta E_j \in \{\Delta E_1, \Delta E_2, \dots, \Delta E_{+\chi}\}$. The smallest of these energy offsets that results in an unstable transfer matrix, or a horizontal or vertical tune that crosses an integer or half-integer resonance, is used to calculate the value of the dominance constraint. The value is calculated as

$$c_{+\chi} = -1 + \frac{j-1}{N_\chi}. \quad (13)$$

If all ΔE_{+j} are stable and do not cross the half-integer or integer, $c_{+\chi}$ is set to 0. A similar procedure is used to constrain the negative chromatic tune footprint $c_{-\chi}$.

Two constraints are implemented by modifying the objective functions. The first is a constraint on the minimum size of the off-energy linear aperture and it modifies the off-energy objective functions f_+ and f_- . The constraint prevents a failure condition where the optimizer improves the off-energy objectives by making the linear aperture tiny, rather than by growing the dynamic aperture. Setting this constraint to 2 or 3 mm is usually sufficient to avoid the condition. If the shortest linear aperture ray at ΔE_{+DA} or ΔE_{-DA} is shorter than this constraint, then f_+ or f_- is set to a perfectly bad value of 1.

The second constraint implemented by modifying the objective functions modifies the on-energy objective function f_0 . It confines the on-energy amplitude dependent tune shift (ADTS). Along the two DA rays closest to the horizontal axis, the horizontal and vertical tunes are calculated. Along the vertical ray, the vertical tune is calculated. This is because the horizontal motion for large vertical and small horizontal offsets is dominated by coupling, rather than by the horizontal optics. The tunes are calculated by summing the element-by-element phase advance in normal mode coordinates. If the tunes of the particle cross the half-integer or integer resonances, then the tracking code considers the particle lost. The apertures along the two horizontal rays and one vertical ray define a “clipping” box. When f_0 is calculated, all DA rays are clipped at the box.

The chromaticity correction is applied by projecting the N_{cs} -dimensional space of chromatic sextupole strengths onto the $(N_{cs} - 2)$ -dimensional surface on which the horizontal and vertical chromaticities have the desired values. This is possible because chromaticity depends linearly on the sextupole strength. N_{cs} is the number of chromatic sextupole families in the lattice. A chromaticity response matrix \mathbf{A} is determined numerically,

$$\mathbf{A} = \begin{pmatrix} \frac{d\chi_x}{dK_1} & \frac{d\chi_x}{dK_2} & \dots & \frac{d\chi_x}{dK_{N_{cs}}} \\ \frac{d\chi_y}{dK_1} & \frac{d\chi_y}{dK_2} & \dots & \frac{d\chi_y}{dK_{N_{cs}}} \end{pmatrix}. \quad (14)$$

The Moore-Penrose pseudoinverse \mathbf{A}_p of \mathbf{A} is calculated via singular value decomposition (SVD) [18]. Then the thin-QR decomposition [18] \mathbf{Q}_1 of $\mathbf{I} - \mathbf{A}_p \mathbf{A}$ is taken, where \mathbf{I} is the $N_{cs} \times N_{cs}$ identity matrix. Note that $\mathbf{Q}_1 \in \mathbb{R}^{N_{cs} \times N_{cs}-2}$.

With \mathbf{A}_p and \mathbf{Q}_1 in hand, take any vector $\vec{\omega} \in \mathbb{R}^{N_{cs}-2}$. The chromatic sextupole strengths \vec{K} given by

$$\vec{K} = \mathbf{A}_p \begin{pmatrix} \chi_{x0} \\ \chi_{y0} \end{pmatrix} + \mathbf{Q}_1 \vec{\omega} \quad (15)$$

result in chromaticities of χ_{x0} and χ_{y0} .

The algorithm does not operate directly on the chromatic sextupole strengths. Instead it operates on $\vec{\omega}$, thus constraining the chromaticities to the desired values and reducing the dimension of the variable space by 2.

V. APPLICATION

Here the genetic algorithm is applied to a prototype lattice for the SLS upgrade, and also the proposed Armenian light source CANDLE.

The SLS upgrade is a 2.4 GeV storage ring built of 12 arcs which consist of five longitudinal gradient bends (LGB) plus two half-bend longitudinal gradient dispersion suppressors. There are three types of straight (short, medium, and long) which reduce the periodicity to 3. The lattice uses antibends to focus the dispersion into the LGBs to minimize the radiation integral \mathcal{I}_5 [19]. The lattice parameters are summarized in Table I.

For the SLS upgrade lattice, the objective functions are the on-energy dynamic aperture and the dynamic aperture at -3% and $+3\%$. The chromatic footprint between -5% and $+5\%$ is constrained such that $37.0 < Q_x < 37.5$ and $10.0 < Q_y < 10.5$. The amplitude-dependent tune shift as described in Sec. IV is constrained to this same region.

For CANDLE, the objective functions are the on-energy dynamic aperture and the dynamic aperture at -2% and $+2\%$. The chromatic footprint between -3% and $+3\%$ is constrained such that $24.5 < Q_x < 25.0$ and $14.0 < Q_y < 14.5$. The amplitude-dependent tune shift as described in Sec. IV is constrained to these same regions.

TABLE I. SLS upgrade is a longitudinal gradient bend plus antibend based replacement for the SLS storage ring. CANDLE is a proposed Armenian light source based on combined function magnets.

	SLS upgrade [20] CANDLE [21]	
Circumference (m)	287.25	258
Emittance (pm)	137	1091
Periodicity	3	16
Topology	$12 \times 7BA$	$16 \times 4BA$
Q_x	37.383	24.700
Q_y	10.280	14.368
Natural chromaticity χ_x	-64.9	-40.6
Natural chromaticity χ_y	-34.5	-26.5
Peak dispersion (cm)	4.9	9.3
# chromatic sextupole families	4	8
# harmonic sextupole families	9	0
# octupole families	10	0

TABLE II. Parameters for genetic optimizer.

	SLS upgrade	CANDLE
# nonlinear magnet families	23	8
# variables	21	6
$c_{\text{mag,sext}} (K_{\text{sext}})^a$	500.0	500.0
$c_{\text{mag,oct}} (K_{\text{oct}}L)^a$	500.0	500.0
ΔE_{+DA}	3%	2%
ΔE_{-DA}	-3%	-2%
$\Delta E_{+\chi}$	5%	3%
$\Delta E_{-\chi}$	-5%	-3%
Footprint ^b $Q_{x,\text{min}}, Q_{x,\text{max}}$	37.0, 37.5	24.5, 25.0
Footprint ^b $Q_{y,\text{min}}, Q_{y,\text{max}}$	10.0, 10.5	14.0, 14.5

^aThe sextupole and octupole quantities have been normalized by $n!$.

^bThe footprint constraints apply to both chromatic tune shift and on-energy ADTS.

The optimization parameters for both lattices are summarized in Table II.

A. SLS upgrade lattice

1. Layout and linear lattice considerations

The following were taken into consideration during the design of the layout and linear optics of the SLS upgrade lattice in order to improve the nonlinearities.

1. The ADTS is suppressed if the horizontal tune is close to

$$Q_{\text{opt}} = \frac{2q+1}{2}N, \quad (16)$$

where N is the periodicity of the machine and q is some integer [22] (Chap. 14.3.1). The periodicity of SLS is 3.

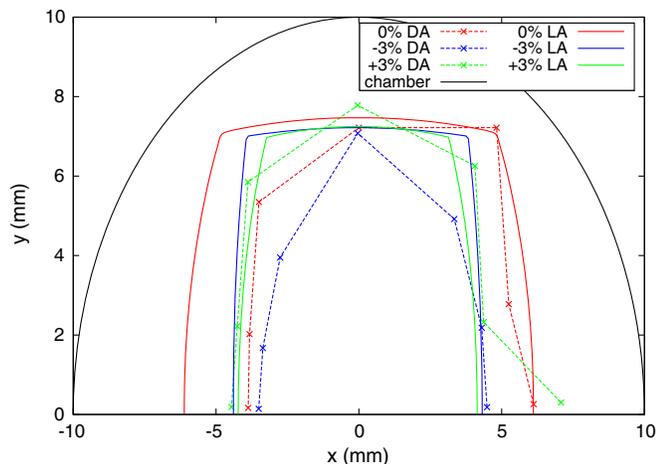


FIG. 4. Dynamic aperture for SLS upgrade at 0%, -3%, and 3% energy offset. Dashed lines are dynamic aperture, solid lines are linear aperture. At the injection point $\beta_x = 3.3$ m and $\beta_y = 6.5$ m. The maxima throughout the machine are $\beta_x = 8.7$ m and $\beta_y = 11.7$ m.

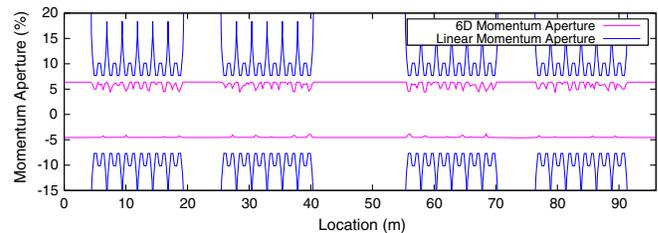


FIG. 5. Momentum aperture of SLS upgrade. Linear aperture is calculated using an element-by-element linearization of the optics. The 6D aperture is calculated with full 6D tracking including radiation damping and synchrotron oscillation for 2.5 periods of the linear synchrotron tune. The asymmetry of the 6D momentum aperture is the influence of nonlinear momentum compaction. The rf voltage is set such that the linear calculation gives a $\pm 5\%$ bucket.

Thus ADTS is reduced by selecting a horizontal tune near one of the following: $\{\dots, 34.5, 37.5, 40.5, \dots\}$.

2. Chromatic sextupoles are placed where dispersion is large, and where either $\beta_x \ll \beta_y$ or $\beta_x \gg \beta_y$. Harmonic sextupoles are placed in dispersion-free regions where $\beta_x \gg \beta_y$, $\beta_x \ll \beta_y$, or $\beta_x \approx \beta_y$.

3. The arcs are constructed from five identical unit cells. The horizontal and vertical phase advances per unit cell are 0.4 and 0.1 radians, respectively. Over the five unit cells, the lowest order chromatic and geometric resonant driving terms are canceled out [6].

2. Optimization

The population size for the SLS upgrade optimization is 300 and begins with a pool of 364 unevaluated, randomly generated, individuals. Each individual is described by 21 variables representing the 23 magnet strengths. The strengths are bounded by $c_{\text{mag,sext}}$ and $c_{\text{mag,oct}}$, as given in Table II. Each seed is farmed via MPI to a CPU which

TABLE III. Touschek lifetime calculation and assumptions. Bunch length is natural without third harmonic cavity. The currents for both SLS upgrade and CANDLE assume a 500 MHz rf system. The horizontal emittance is that given by the radiation integrals calculation. The Touschek lifetime is calculated with 6D tracking including radiation damping and synchrotron oscillations. The reference Touschek lifetime is calculated for the SLS upgrade by assuming a 5% momentum acceptance everywhere, and by assuming 3% for CANDLE.

	SLS upgrade	CANDLE
Horizontal emittance (pm)	137.0	1091.0
Vertical emittance (pm)	10.0	10.0
Current per bunch (mA)	1.0	1.0
Number of particles per bunch (10^9)	6.0	5.4
Bunch length (mm)	0.261	0.494
6D Touschek lifetime (hr)	4.58	3.82
Reference Touschek lifetime (hr)	4.35	3.63

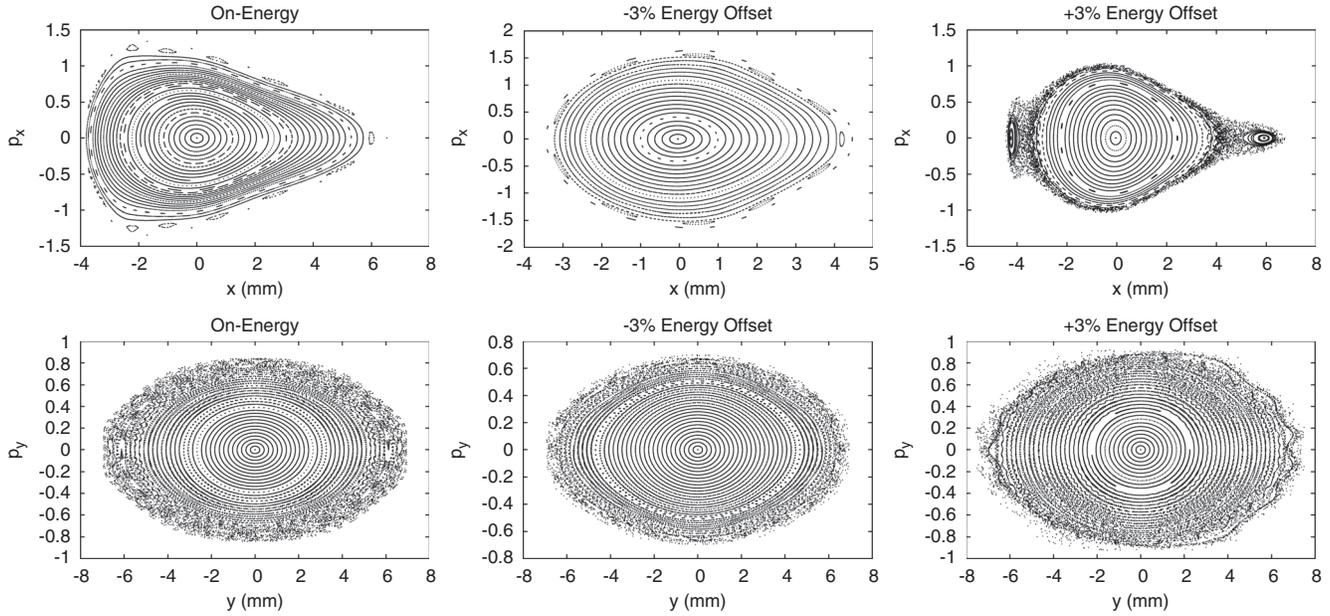


FIG. 6. Horizontal and vertical phase space portraits for SLS upgrade on-energy, and at -3% and $+3\%$ energy defect.

evaluates its constraint and objective values. As soon as 300 individuals have been evaluated, the population is sorted and the 150 least fit are deleted. A four-way tournament is used to select parent pairs from the surviving population. From each pair, simulated binary crossover plus mutation is used to generate two new children. The new children are

added to the pool of unevaluated individuals, and the process repeats.

The initial random population contains no individuals which satisfy the dominance constraints (i.e. all individuals are infeasible). Therefore the population members initially compete for who has the least-bad constraint violations. For the particular optimizer run shown here, the first feasible individual appears at generation 21, and at generation 41 the population consists entirely of individuals which satisfy all of the dominance constraints. From generation 1 to 20 requires 2 minutes, and from generation 20 to 41 requires 14 minutes. This first stage of the optimization proceeds quickly because infeasible individuals are evaluated only

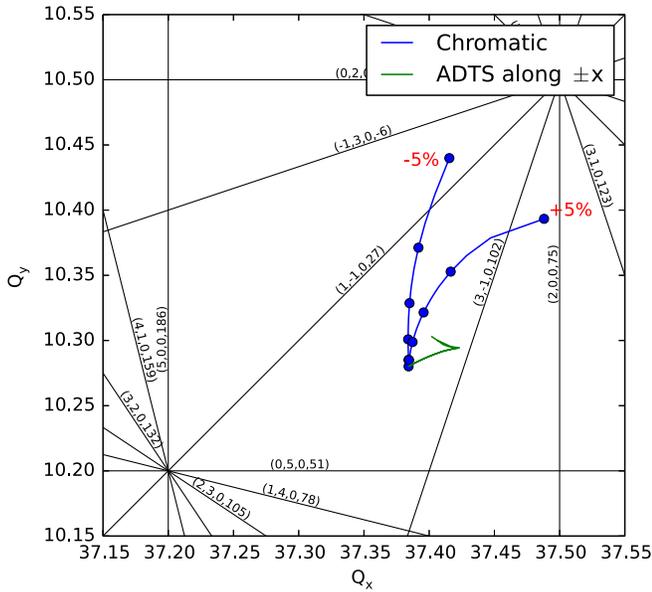


FIG. 7. SLS upgrade: Chromatic footprint from -5% to $+5\%$ and ADTS along x from the $-x$ DA to the $+x$ DA. Blue points are 1% increments. Resonance lines are labeled (p, q, r, n) where $pQ_x + qQ_y + rQ_s = n$. Low order resonance lines and higher order lines are plotted. Higher order resonance lines excluded by periodicity 3 are not shown.

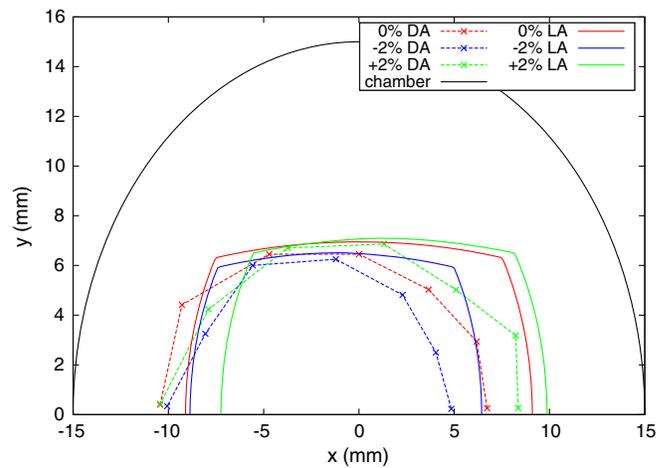


FIG. 8. Dynamic aperture (DA) and linear aperture (LA) for CANDLE at 0%, -2% , and 2% energy offset. At the injection point $\beta_x = 3.3$ m and $\beta_y = 1.8$ m. The maxima throughout the machine are $\beta_x = 9.1$ m and $\beta_y = 8.4$ m.

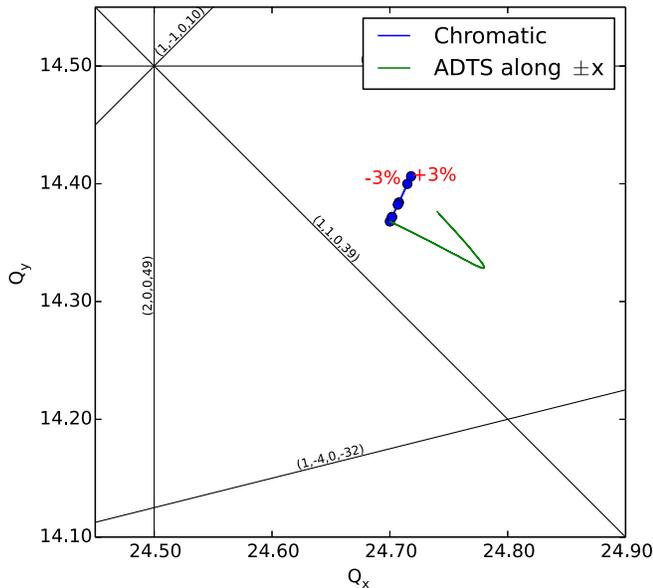


FIG. 9. CANDLE: Chromatic footprint from -3% to $+3\%$ and ADTS along x from the $-x$ DA to the $+x$ DA. ADTS is calculated with a small vertical offset to allow for accurate calculation of vertical tune. Higher order resonance lines excluded by periodicity 16 are not shown.

for off-momentum closed orbit amplitudes and tune footprints. During the remainder of the optimization run, individuals are competing based on dynamic aperture. The entire optimization completes after 45 hours to converge at generation 616. Computing resources are 64 E5-2670 xeon cores. These compute time requirements are generally indicative of the resources required by the algorithm on the SLS upgrade.

Following the optimization run, the seeds with the most promising objective function values are selected by hand for further evaluation. Further evaluation includes on- and off-energy rastered survival plots, higher resolution chromatic and ADTS tune footprints, momentum aperture, and Touschek lifetime evaluation. Using these additional evaluations, the lattice designer selects a best individual.

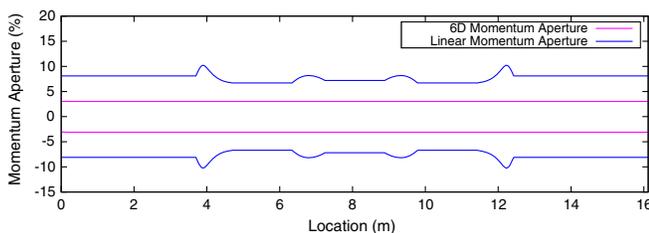


FIG. 10. Momentum aperture and linear momentum aperture for one period of the CANDLE lattice. Linear momentum aperture is calculated from Twiss optics linearized about the on-energy particle. The rf voltage is set such that the linear calculation of the rf bucket depth is $\pm 3\%$.

The dynamic aperture of this individual at 0% , -3% , and $+3\%$ is shown in Fig. 4. Figure 5 compares the momentum aperture and linear momentum aperture. The linear momentum aperture is calculated by linearizing the one-turn map at each element. The reference Touschek lifetime is calculated from the linear rf bucket height, and is taken as the benchmark against which to judge the effectiveness of the algorithm.

The assumptions used for the Touschek lifetime calculation are shown in Table III. The lifetime exceeds the reference lifetime because the nonlinearity of the longitudinal phase space causes it to exceed the dimensions of the linear rf bucket, and the momentum aperture is not otherwise limited by the transverse nonlinearities.

Recall that the genetic algorithm does not directly optimize the Touschek lifetime nor momentum aperture. Rather, it constraints the chromatic and amplitude-dependent tune footprints and maximizes the dynamic aperture area at 0% and $\pm 3\%$. The element-by-element variation in the momentum aperture is small. This indicates that the Touschek lifetime is limited by the longitudinal dynamics, and not by nonlinearities in the transverse optics. Judging by this result, an off-momentum dynamic aperture optimization plus tune footprint constraint is a valid proxy for optimizing the Touschek lifetime and momentum aperture.

Shown in Fig. 6 are $x - p_x$ and $y - p_y$ phase space portraits for the optimized SLS upgrade lattice. The portraits are calculated using 4D tracking for 100 turns. The lack of large resonance islands and lack of thick chaotic layers inside the stable region is a positive result that should contribute to the robustness of the solution when misalignments are added.

Figure 7 shows the chromatic tune footprint and ADTS along $\pm x$. The chromatic tunes are calculated by linearizing the off-energy optics. The ADTS is calculated by summing element-by-element phase advances in normal mode coordinates. The ADTSs along $+x$ and $-x$ mostly overlap.

B. CANDLE

CANDLE is a proposed 3 GeV, 216 m Armenian light source project [21] providing 8.54 nm horizontal beam emittance. According to the recent developments in storage ring lattice design and magnet technologies a new upgrade prototype has been designed [11], which is constructed of sixteen 4BA cells and provides 1.1 nm horizontal beam emittance. The study shown here is on this new 1.1 nm prototype. Each cell is composed of combined function bends with both quadrupole and sextupole moments. Some properties are shown in Table I. Two features which contribute to CANDLE's nonlinearities are: (1) The sextupole moments are spread out across a broad phase advance. (2) All sextupole moments are in dispersive regions.

The dynamic aperture on-energy and at $\pm 2\%$ are shown in Fig. 8. The horizontal ADTS and chromatic footprint out to $\pm 3\%$ are shown in Fig. 9.

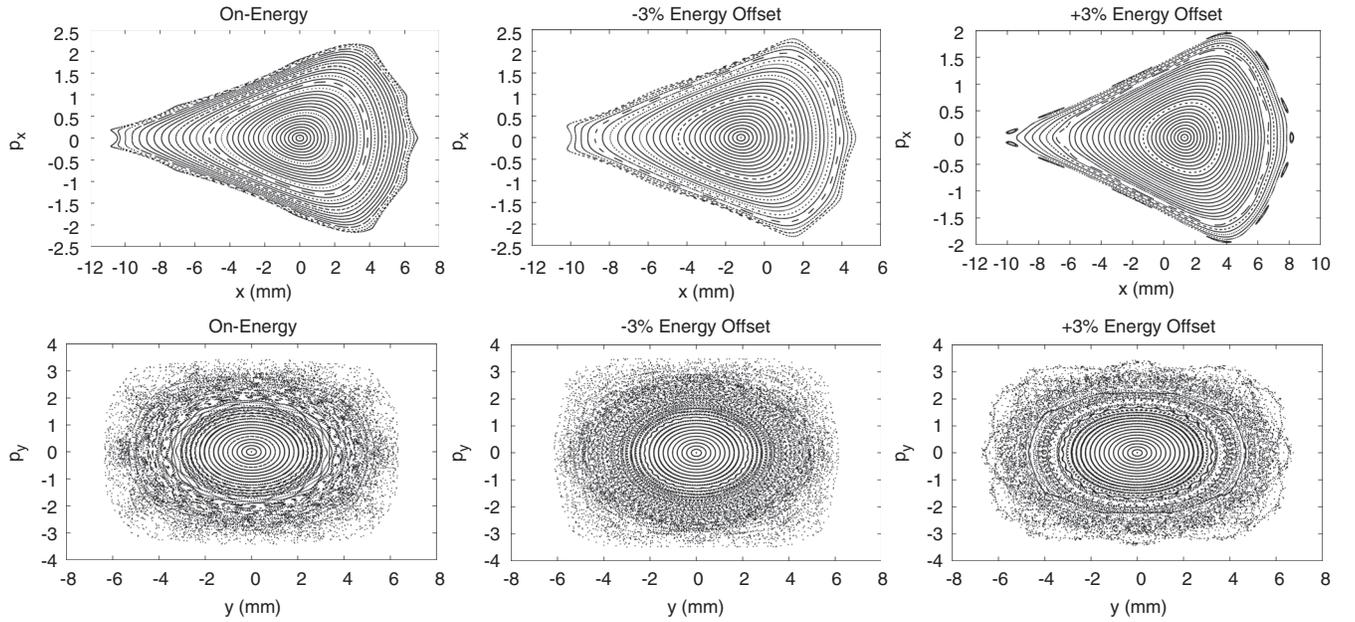


FIG. 11. Horizontal and vertical phase space portraits for CANDLE on-energy, and at -2% and $+2\%$ energy defect.

Touschek lifetime results and assumptions are shown in Table III. The momentum aperture calculated from 6D tracking is shown in Fig. 10. The aperture is determined entirely by the rf bucket and not limited by the transverse optics. The phase space portraits are shown in Fig. 11. Optimizing the $\pm 2\%$ dynamic aperture and constraining the chromatic tune footprint to $\pm 3\%$ has successfully optimized the global momentum aperture to at least $\pm 3\%$.

VI. TOLERANCE TO MACHINE MISALIGNMENTS

The tolerance of the optimized SLS upgrade lattice to machine misalignments is tested. The genetic algorithm is

TABLE IV. Misalignments are drawn from a random Gaussian distribution, subject to a $2\text{-}\sigma$ cutoff. Misalignments which exceed the cutoff are redrawn.

Property	Relative to	Distribution to σ
Quadrupole and sextupole tilt	Girder	$50 \mu\text{rad}$
Quadrupole and sextupole horizontal offset	Girder	$20 \mu\text{m}$
Quadrupole and sextupole vertical offset	Girder	$20 \mu\text{m}$
Bend and antibend tilt	Girder	$50 \mu\text{rad}$
Bend and antibend horizontal offset	Girder	$20 \mu\text{m}$
Bend and antibend vertical offset	Girder	$20 \mu\text{m}$
Girder tilt	Lab	$50 \mu\text{rad}$
Girder horizontal offset	Lab	$50 \mu\text{m}$
Girder vertical offset	Lab	$50 \mu\text{m}$
LGB tilt	Lab	$50 \mu\text{rad}$
LGB horizontal offset	Lab	$20 \mu\text{m}$
LGB vertical offset	Lab	$20 \mu\text{m}$

applied to the ideal lattice and a single solution is selected by the lattice designer. The lattice is then misaligned according to Table IV and corrected as described below. The beam lifetime and on-energy dynamic aperture of the resulting misaligned and corrected lattice is calculated. This procedure is repeated for many misalignment seeds.

The correction procedure begins by flattening the horizontal and vertical orbits using an orbit response matrix and SVD. Then a simultaneous horizontal phase, vertical phase, and horizontal dispersion correction is applied using a combined phase and dispersion response matrix. The residual coupling after these corrections ranges from 0.4

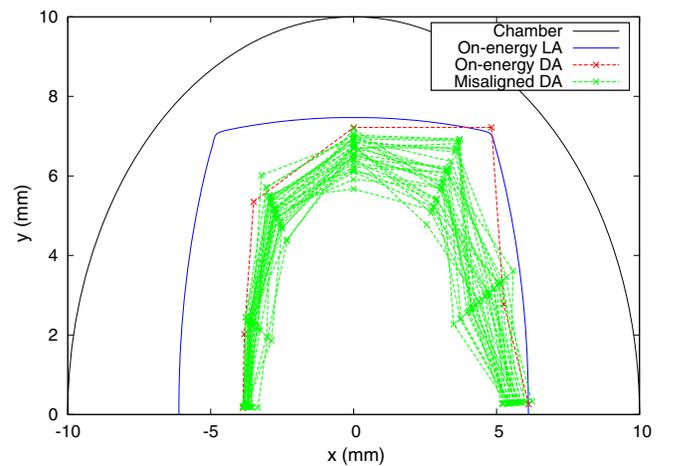


FIG. 12. On-energy dynamic aperture for ideal lattice and 30 misaligned and corrected lattices. At the injection point $\beta_x = 3.3 \text{ m}$ and $\beta_y = 6.5 \text{ m}$. The maxima throughout the machine are $\beta_x = 8.7 \text{ m}$ and $\beta_y = 11.7 \text{ m}$.

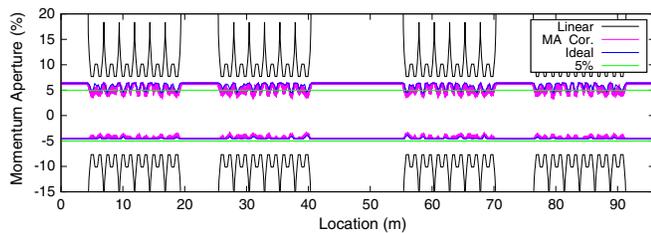


FIG. 13. Momentum aperture from 6D tracking for ideal lattice and 30 misaligned and corrected lattices.

to 3.2%. A dedicated coupling correction is not included in this study.

So 30 misalignment seeds are generated and corrected. One of these seeds fails to have a closed orbit and is discarded. The on-energy dynamic aperture and momentum aperture for the remaining 29 are calculated and the results are shown in Figs. 12 and 13. Fifty percent of the misaligned and corrected lattices have a lifetime longer than 3.8 hr, and 95% have a lifetime longer than 3.6 hr. The reference lifetime is 4.4 hr.

The misalignment and correction procedure applied here is pessimistic. The fully developed SLS upgrade misalignment model will take into account that the LGBs will be aligned relative to the girders, and the girders relative to one another. Furthermore, coupling correction and vertical dispersion correction will be applied in the actual machine. Despite the pessimistic scenario, the calculated dynamic aperture and lifetimes of the misaligned and corrected lattices are acceptable. From this we conclude that the sextupole and octupole scheme generated by the genetic algorithm is sufficiently robust against imperfections.

The sensitivity of the chromaticity correction scheme to beta beating is tested by applying gradient errors to the

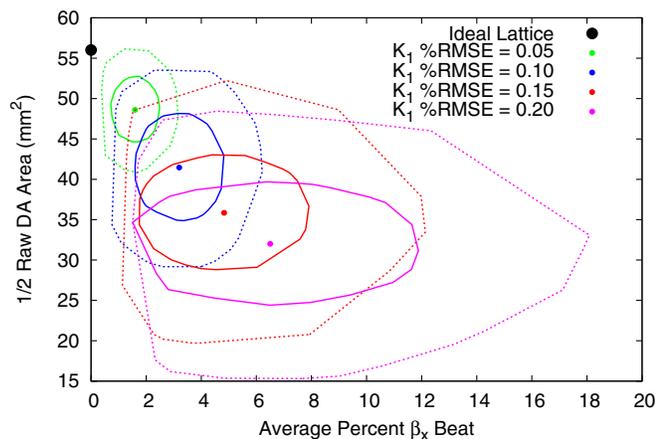


FIG. 14. Area of the $+y$ dynamic aperture versus percent beta beating induced by quadrupole gradient errors. The solid enclosure is the convex hull containing 50% of the seeds nearest the mean, the dashed enclosure contains 90% of the seeds.

quadrupole moments in quadrupoles and gradient bends. Errors with rms values of 0.05%, 0.10%, 0.15%, and 0.20%, subject to a $2\text{-}\sigma$ cutoff, are tested. No corrections are applied. One thousand seeds are generated for each of the four cases. For each seed, the on-energy dynamic aperture and mean percent beta beat is calculated. Plotted in Fig. 14 is the mean for each case and the convex hulls that contain 50% and 90% of the seeds closest to the mean. In the original SLS, beta beating is measured to be 2% [23]. We therefore anticipate a reduction in the on-energy dynamic aperture area in the SLS upgrade due to beta beating of less than 20%.

VII. CONCLUSION

The genetic algorithm presented here offers a robust and computationally affordable technique for generating globally optimal chromaticity correction schemes for diffraction limited light sources. The resulting correction schemes have good on-energy dynamic aperture which should help injection efficiency and give a wide momentum aperture for long beam lifetime. The schemes are sufficiently robust against misalignments.

One feature of this algorithm is the use of dominance constraints to encourage individuals in the early population to take on properties that will later on contribute to healthy objective values.

A second feature is the use of off-energy dynamic aperture along with tune footprint constraints as a proxy for the computationally expensive direct momentum aperture calculation.

Based on development efforts at SLS and results shared by the CANDLE collaboration [11], this genetic algorithm delivers results that are as good or better than those obtained by applying second order resonant driving term minimization. The genetic algorithm converges in a couple days on commonly available computing resources. This turn-around time is comparable to that required for a lattice designer to develop a scheme using resonant driving term minimization. Thus the genetic algorithm presented here is a practical solution for optimizing sextupole and octupole strengths in a diffraction limited light source project.

ACKNOWLEDGMENTS

I would like to thank David Sagan of Cornell University for supporting the development of this genetic algorithm through his work on the BMAD [15] library. Many aspects of this work benefited greatly from discussions with Masamitsu Aiba and Michael Böge. I would like to thank Andreas Streun for his mentorship in storage ring nonlinearities, and for the well-designed SLS upgrade lattice. I would like to thank Artsrun Sargsyan for the opportunity to contribute to the development of CANDLE.

- [1] 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).
- [2] Y. Ineichen, Toward massively parallel multiobjective optimization with application to particle accelerators, Ph.D. thesis, ETH Zürich, 2013.
- [3] M. Borland, V. Sajaev, L. Emery, and A. Xiao, in *Proceedings of the 23rd Particle Accelerator Conference, Vancouver, Canada, 2009* (IEEE, Piscataway, NJ, 2009), pp. 3850–3852.
- [4] Y. Ineichen, A. Adelman, A. Kolano, C. Bekas, A. Curiono, and P. Arbenz, A parallel general purpose multi-objective optimization framework, with application to beam dynamics, [arXiv:1302.2889v1](https://arxiv.org/abs/1302.2889v1).
- [5] I. V. Bazarov and C. K. Sinclair, *Phys. Rev. ST Accel. Beams* **8**, 034202 (2005).
- [6] J. Bengtsson, The sextupole scheme for the Swiss Light Source (SLS): An analytic approach, PSI Technical Report No. SLS Note 9/97, 1997.
- [7] C. Sun, D. Robin, H. Nishimura, C. Steier, and W. Wan, in *Proceedings of the 2nd International Particle Accelerator Conference, San Sebastián, Spain* (EPS-AG, Spain, 2011).
- [8] L. Yang, Y. Li, W. Guo, and S. Krinsky, *Phys. Rev. ST Accel. Beams* **14**, 054001 (2011).
- [9] W. Gao, L. Wang, and W. Li, *Phys. Rev. ST Accel. Beams* **14**, 094001 (2011).
- [10] X. Huang and J. Safranek, *Nucl. Instrum. Methods Phys. Res., Sect. A* **757**, 48 (2014).
- [11] A. Sargsyan, Towards the low emittance design for the CANDLE storage ring, *XXIII European Synchrotron Light Source Workshop*, 2015.
- [12] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, in *Evolutionary Multicriterion Optimization (EMO 2003)*, Lecture Notes in Computer Science, edited by C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele (Springer, Berlin, 2003), pp. 494–508.
- [13] E. Zitzler, M. Laumanns, and L. Thiele, in *Evolutionary Methods for Design, Optimisation, and Control* (CIMNE, Barcelona, Spain, 2002), pp. 95–100.
- [14] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, in *Parallel Problem Solving from Nature—PPSN VI*, edited by M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel (Springer, Berlin, 2000), pp. 849–858.
- [15] D. Sagan, *Nucl. Instrum. Methods Phys. Res., Sect. A* **558**, 356 (2006).
- [16] K. Deb, Multiobjective optimization using evolutionary algorithms: An introduction, KanGAL Technical Report No. 2011003, Indian Institute of Technology Kanpur, 2011.
- [17] K. Deb and R. B. Agrawal, *Complex Syst.* **9**, 115 (1995).
- [18] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore, MD, 1996).
- [19] A. Streun and A. Wrulich, *Nucl. Instrum. Methods Phys. Res., Sect. A* **770**, 98 (2015).
- [20] A. Streun, M. Aiba, M. Böge, M. Ehrlichman, and Á. Saá Hernández, Design studies for an upgrade of the SLS storage ring, in *Proceedings of the International Particle Accelerator Conference, Virginia, 2015*, p. TUPJE047.
- [21] A. Sargsyan (Candle Collaboration), CANDLE: Design study of 3 GeV Synchrotron Light Source, Technical Report No. ASLS-CANDLE R-001-02, Yerevan, Armenia, 2002.
- [22] H. Wiedemann, *Particle Accelerator Physics*, 3rd ed. (Springer, Berlin, 2007).
- [23] M. Aiba, M. Böge, J. Chrin, N. Milas, T. Schilcher, and A. Streun, *Phys. Rev. ST Accel. Beams* **16**, 012802 (2013).