2014

Simultaneous Optimization of the Cavity Heat Load and Trip Rates in Linacs Using a Genetic Algorithm

Balša Terzić
Old Dominion University, bterzic@odu.edu

Alicia S. Hofler

Cody J. Reeves

Sabbir A. Khan

Geoffrey A. Krafft
Old Dominion University

See next page for additional authors

Follow this and additional works at: https://digitalcommons.odu.edu/physics_fac_pubs

Part of the Computer Sciences Commons, and the Physics Commons

Authors

Balša Terzić, Alicia S. Hofler, Cody J. Reeves, Sabbir A. Khan, Geoffrey A. Krafft, Jay Benesch, Arne Freyberger, and Desh Ranjan

Simultaneous optimization of the cavity heat load and trip rates in linacs using a genetic algorithm

Balša Terzić,[1,2,3,*] Alicia S. Hofler,[3] Cody J. Reeves,[4] Sabbir A. Khan,[5] Geoffrey A. Krafft,[2,3] Jay Benesch,[3] Arne Freyberger,[3] and Desh Ranjan[2,5]

[1]*Department of Physics, Old Dominion University, Norfolk, Virginia 23529, USA*
[2]*Center for Accelerator Science, Old Dominion University, Norfolk, Virginia 23529, USA*
[3]*Thomas Jefferson National Accelerator Facility, Newport News, Virginia 23606, USA*
[4]*Department of Physics, University of Florida, Gainesville, Florida 32611, USA*
[5]*Department of Computer Science, Old Dominion University, Norfolk, Virginia 23529 USA*
(Received 21 July 2014; published 15 October 2014)

In this paper, a genetic algorithm-based optimization is used to simultaneously minimize two competing objectives guiding the operation of the Jefferson Lab's Continuous Electron Beam Accelerator Facility linacs: cavity heat load and radio frequency cavity trip rates. The results represent a significant improvement to the standard linac energy management tool and thereby could lead to a more efficient Continuous Electron Beam Accelerator Facility configuration. This study also serves as a proof of principle of how a genetic algorithm can be used for optimizing other linac-based machines.

I. INTRODUCTION

The Continuous Electron Beam Accelerator Facility (CEBAF) is a superconducting facility located at Jefferson Lab. It provides a continuous electron beam of up to 12 GeV for use in nuclear physics experiments in up to four experimental halls simultaneously. CEBAF features two linacs, North and South, each with 200 multicell elliptical superconducting radio frequency (SRF) cavities. Each cavity has a tunable gradient which can be adjusted within its operational range to reproduce the total design energy gain per pass through the linac. Different configurations of the cavity gradients lead to different heat consumption and trip rates (related to cavity design). Cavity trips interrupt the machine operation and reduce the overall time for useful data collection in the experimental halls.

The study is motivated by the desire to find the optimal set of cavity gradients needed to maximize science *and* minimize the cost of operation (electricity bill). Realizing that a large part of Jefferson Lab's multimillion dollar monthly electricity bill is due to cryogenics, even modest improvements in the heat load may translate into millions of dollars in savings over the lifetime of CEBAF operations.

Cavity heat load and radio frequency (rf) trip rates are competing objectives—minimizing one quantity will increase the other—necessitating a multiobjective optimization. Successfully implemented multiobjective optimization

will provide a set of optimal solutions clearly showing the trade-offs between the competing objectives. The ability to know these trade-offs is invaluable in every-day linac operations. For instance, knowing how much money operating the linacs at six trips per hour as opposed to five will lead to judicious experiment and operation planning.

The remainder of the paper is organized as follows. Section II describes the details of the CEBAF linacs at Jefferson Lab. Section III outlines the optimization problem, shows how its simplified, single-objective version can be solved using Lagrange multipliers (Sec. III A) and then how a genetic algorithm (GA) can be used to solve the full multiobjective problem (Sec. III B). Section IV presents the results of the multiobjective optimization. Finally, Sec. V summarizes the findings and discusses their relevance for linac-based machines in general.

II. CEBAF LINACS

CEBAF has two linacs with 25 cryomodules, each containing eight cavities, for a total of 400 cavities. The cavities are grouped in eight-cavity cryomodules called C25, C50 and C100, labeled based on the energy gain expected in MeV per cryomodule. The initial complement, used during the 6 GeV era, consisted of 40 cryomodules, 320 cavities, with 50 cm accelerating length. The recent addition, which ushered the 12 GeV era, consists of ten cryomodules, 80 cavities, of 70 cm active length. Surface to accelerating gradient ratio is 2.56∶1 in the original complement and 2.08∶1 in the new complement.

Using fault data acquired beginning on January 30, 1995, a total of 1563 exponential models [see Eq. (2) in Sec. III] was developed, about five per cavity, due to changing conditions during the following 17 years. Field

---

*Corresponding author.
bterzic@odu.edu

emission during the first half of the period is discussed in [1]. The contemporary set of models was input to the linac energy manager (LEM) optimization tool which initially equally weighted 2 K heating and fault rate [2], but which soon excluded the former because the solutions found were too unstable.

In 2003 a hurricane hit CEBAF. The uncontrolled warm-up and the methods used in the subsequent evacuation of the cavity volumes caused a substantial loss in potential. A refurbishment program was instituted in 2006, and the first refurbished cryomodule was installed in April 2007. Eleven cryomodules have been refurbished to date. At installation, they provided 50 MeV/c gain, but over the course of three years they lost about 20% of capacity. Finally, as part of the 12 GeV upgrade project, ten cryomodules with eight 70 cm cavities were added, designed to provide 100 MeV/c each and designated C100 [2].

The linac energy manager (LEM) program in use since 1999 minimizes fault rate by setting cavities which do not fault at maximum gradient and makes up the remainder of the needed momentum gain by optimizing the C25 cavities using Lagrange multipliers on the set of exponential models. Total heat (refrigerator) constraints are applied manually. The rf gradients were calibrated using the accelerator arcs as spectrometers. Cavity quality factors, Qs, were obtained near the operating gradient by measuring the valve response to resistive heating, turning off individual cavities, and monitoring the valve response over tens of minutes.

For each cavity in the linac the LEM program requires (i) cavity designator; (ii) input Q so required rf power may be calculated (constraint); (iii) maximum gradient determined during commissioning or beam operations; (iv) an allowance for detuning via acoustic noise which alters cavity length/frequency; (v) for C25 cavities, the gradient at which a fault occurs every eight hours; (vi) for C25 cavities, the slope of the exponential fit; (vii) Q_0 of the fundamental accelerating mode at a representative operating gradient, for use in calculating heat generated; (viii) cavity length, 0.5 m for C25 and C50 cavities and 0.7 m for C100. For the newer cavities, C50 and C100, as well as a small fraction of the old C25 cavities, not enough data exist for a reliable exponential model to be developed. From the optimization standpoint, these cavities are treated as if they never trip.

Moore's law now allows a more sophisticated algorithm, one which can provide a stable solution with respect to *both* fault rate and heat load, to be solved in a few minutes when an rf system fails and the allowed momentum gain from a cavity must be substantially reduced or set to zero.

The results presented in this paper are based on the data from the 2009 PVDIS run during which the CEBAF was operated at 6.068 GeV with 500 μA linac beam current. Extending this formalism to data for the new 12 GeV CEBAF configuration when it becomes available is straightforward.

III. THE OPTIMIZATION PROBLEM

Simultaneously optimizing cavity heat load and trip rates is a multiobjective optimization problem. Since there are N_c cavity gradients G_i to adjust ("knobs to turn"), this is also a N_c-dimensional problem. Each cavity gradient G_i, measured in MV/m, is restricted to the domain $[3, D_i]$, where 3 MV/m is the lower limit of the stability of the control system and D_i is the "drive-high"—the maximum gradient at which a cavity can be safely set, which varies from cavity to cavity.

The cavity power transfer to the liquid helium for CEBAF SRF cavities is given in [3]

$$P(\vec{G}) = \sum_{i=1}^{N_c} \frac{G_i^2 L_i}{c_i Q(G_i)}, \tag{1}$$

where $\vec{G} = (G_1, G_2, \ldots, G_{N_c})$, the cavity length $L_i = 0.5$ m and shunt impedance $c_i = 960$ Ω/m for C25/C50, and $L_i = 0.7$ m and $c_i = 968$ Ω/m for C100. $Q_i(G_i)$ is the measured (unloaded) cavity quality factor, which tends to decrease as G_i increases. In this implementation, given the data modeled, Q_i is constant for each cavity [$Q_i(G_i) = Q_i$].

The cavity trip rate exponential model (per hour) is reported in [2] as

$$T(\vec{G}) = 3600 \sum_{i=1}^{N_c} \exp\left[A + B_i(G_i - F_i)\right], \tag{2}$$

where $A = -10.26813067$, B_i the model trip slope and F_i the fault gradient. As with any GA simulation, the quality of the results obtained depend on the appropriateness of the underlying model.

To be feasible, any solution produced by the optimization must also satisfy the constraint that the total energy gain in the linac is within 2 MeV of a prescribed energy E_{linac}.

Therefore, the full multiobjective, multidimensional, constrained optimization problem becomes

$$\text{Minimize} \quad P(\vec{G}), T(\vec{G})$$

$$\text{Subject to} \quad \left| E_{\text{linac}} - \sum_{i=1}^{N_c} G_i L_i \right| < 2,$$

$$3 \le G_i \le D_i. \tag{3}$$

A. Single-objective optimization using Lagrange multipliers

It is both pedagogical and illuminating for a moment to study two related single-objective problems to discern the expected asymptotic behavior of the multiobjective problem: (i) constrained minimization of the heat load only and (ii) constrained optimization of the trip rate only.

A single-objective optimization is significantly simpler and can be implemented using a number of standard tools. One such tool is an approach that utilizes Lagrange multipliers.

First the Lagrangian for the constrained single-objective optimization problem is defined as

$$\mathcal{L}(\vec{G}, \lambda) = f(\vec{G}) + \lambda \left(E_{\text{linac}} - \sum_{i=1}^{N_c} G_i L_i \right), \qquad (4)$$

where $f(\vec{G})$ is either $P(\vec{G})$ or $T(\vec{G})$, and λ is a Lagrange multiplier. For simplicity and without loss of generality, the constraint imposed here is an equality constraint that the total energy gain in the linac must exactly equal the prescribed energy E_{linac} (not within 2 MeV as formulated in the full multiobjective version). This optimization does not confine the cavity gradients G_i to their operational ranges $[3, D_i]$. The solution to the constrained problem is obtained upon setting the partial derivatives of the Lagrangian with respect to the coordinates G_i and the Lagrange multiplier λ to zero:

$$\frac{\partial \mathcal{L}(\vec{G}, \lambda)}{\partial (G_i, \lambda)} = 0, \qquad i = 1, N_c. \qquad (5)$$

While taking a partial derivative with respect to the Lagrange multiplier λ only ensures that the constraint is satisfied, the other derivatives lead to the optimal solution:

$$\frac{\partial P(\vec{G})}{\partial G_i} = \frac{\partial}{\partial G_i} \left[\frac{G_i^2}{Q_i(G_i)} \right] \frac{L_i}{c_i} \approx \frac{2 G_i L_i}{c_i Q_i} = \lambda_P L_i = \text{const},$$

$$\frac{\partial T(\vec{G})}{\partial G_i} = 3600 B_i \exp[A + B_i(G_i - F_i)] = \lambda_T L_i = \text{const}. \qquad (6)$$

$\approx$ denotes our treatment of Q_i as a constant, independent of G_i, which we carry forward throughout the remainder of the paper. After solving for G_i and substituting into the constraint,

$$G_i = \frac{\lambda_P}{2} c_i Q_i, \qquad \lambda_P = \frac{2 E_{\text{linac}}}{\sum_{i=1}^{N_c} c_i Q_i L_i},$$

$$G_i = \frac{\ln \frac{\lambda_T L_i}{3600 B_i} - A}{B_i} + F_i,$$

$$\lambda_T = \exp \left[\frac{\delta E - \sum_{i=1}^{N_c} \frac{L_i}{B_i} \ln \frac{L_i}{3600 B i} + A \sum_{i=1}^{N_c} \frac{L_i}{B_i} - \sum_{i=1}^{N_c} F_i L_i}{\sum_{i=1}^{N_c} \frac{L_i}{B_i}} \right], \qquad (7)$$

where $\delta E = E_{\text{linac}} - \sum_{i=1}^{N_c} D_i L_i$, with the summation including only values of i for which $B_i = 0$. Combining cavity parameters on one side of the equation and constants on the other gives the conserved quantities of the two single-objective optimizations:

$$\frac{G_i}{Q_i} = \frac{\lambda_P c_i}{2},$$

$$B_i \exp[A + B_i(G_i - F_i)] = \frac{\lambda_T L_i}{3600}. \qquad (8)$$

These quantities are plotted in Fig. 1. Replacing the equality constraint with the inequality constraint, i.e., allowing for a small scatter in total energy gain within 2 MeV around the nominal linac energy E_{linac} will translate into a small scatter in the conserved quantities, as will be shown in the next section.

The solutions of the single-objective optimization— minimizing the heat load only or the trip rates only— provide the asymptotic behavior of the multiobjective optimization that will be described in the next section. Such trade-off solutions obtained from the multiobjective optimization—those that judiciously minimize both objective functions simultaneously—will be more or less bunched around the conserved quantities.

These analytical solutions obtained by solving the single-objective constrained optimization using Lagrange multipliers are not only pedagogic in that they describe the asymptotic behavior of the full multiple-objective optimization, but also useful in providing theoretical range on the values of the objective functions. In particular, the theoretical minimum heat load is given by the value of the single-objective minimization of the heat load only, while the maximum heat load is computed from the set of cavity gradients corresponding to the solution of the single-objective minimization of the trip rate. Similarly, the theoretical minimum trip rate is given by the value of the single-objective minimization of the trip rate only, while the maximum trip rate is computed from the set of cavity gradients corresponding to the solution of the single-objective minimization of the heat load. For the configuration considered here, the theoretical operational ranges on the heat load are [1015, 1405] W for the North linac and [948, 1437] W for the South linac, while the theoretical operational ranges on the trip rate are $[0.74, 6 \times 10^9]$ per hour for the North linac and $[0.2, 4 \times 10^{14}]$ per hour for the South linac.

When comparing these results to the full multiobjective optimization, one has to remember that this implementation using Lagrange multipliers does not ensure (by design) that the cavity gradients G_i are within operational ranges $[3, D_i]$. Therefore, the Lagrange multiplier approach implemented here, while qualitatively accurate, only provides a rough quantitative estimate on the full optimization problem.

The Appendix outlines an interesting concept of casting often-difficult multiobjective optimization as a set of single-objective optimizations.

B. Multiobjective optimization using a genetic algorithm

The solutions obtained from the multiobjective optimization will comprise a *Pareto-optimal front*: a set of
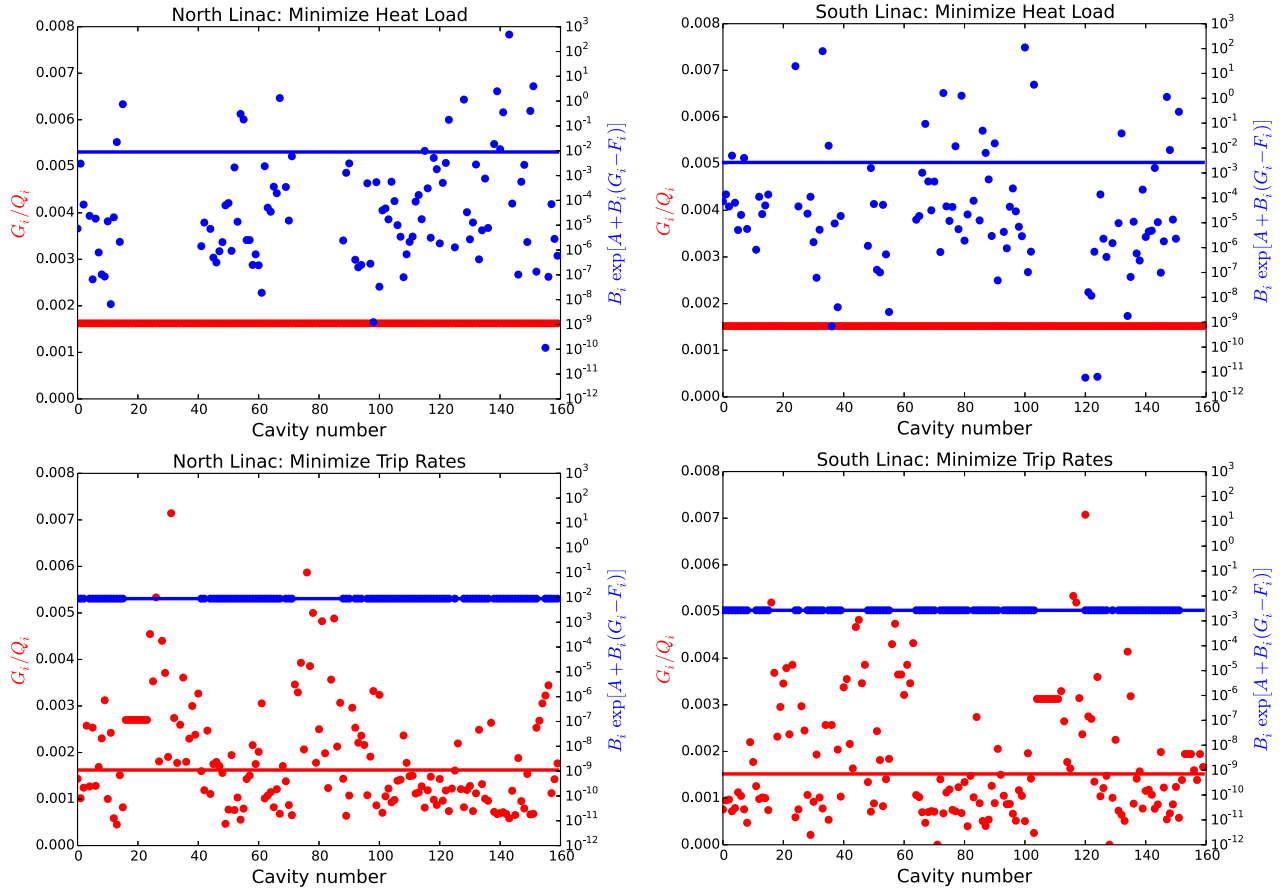
FIG. 1.   Plots of the two conserved quantities given in Eq. (8)—G_i/Q_i shown in red and $B_i \exp\left[A + B_i(G_i - F_i)\right]$ shown in blue—for the heat load only minimization (top row) and the trip rate minimization (bottom row) for the CEBAF's North Linac (left column) and South Linac (right column). Solid lines represent the analytically computed values of the conserved quantities.

nondominated solutions to the optimization problem. A solution is defined as nondominated if there exist no other solutions for which all objective functions are more optimal. The concept of nondominance is invoked to provide a gradation among all feasible solutions for a multiobjective optimization (Fig. 2).

The GA-based multiobjective optimization uses the platform and programming language interface for search algorithms (PISA) developed at ETH Zürich [4,5] and alternate PISA (APISA) from Cornell University [6]. PISA is a modular test bed system for GAs, which separates the GA parent selection process from the optimization problem evaluation and population generation processes into two programs: the *selector* and the *variator*. This design allows for easy application of different GA selection algorithms, selector programs, to several academic bounded-domain optimization problems for performance and convergence comparisons. APISA expands the functionality of PISA to provide support for strict inequality constraints. This paradigm has solved a number of multidimensional, single-objective and multiobjective problems in accelerator physics, as described in [7–9]. A pedagogical, self-contained overview of the GA-based optimization is given in [7].

The GA-based simulation starts with a random population of individuals with different operational gradients G_i, each sampling its allowed range $[3, D_i]$ without a bias. Only individuals which satisfy the energy constraint are kept in the pool of available solutions. The selector strategy is the Strength Pareto Evolutionary Algorithm 2 (SPEA2) [10], mostly encouraged by its successes in earlier work [7,9,11].

An early attempt to use GA-based methods to simultaneously optimize the cavity heat load and trip rates [12] established the proof of concept that the GA-based methods can be used for improving the heat load given the preset limit on the trip rates. The study stopped short of being operationally useful due to two important issues: (i) the two objective functions, cavity heat load and trip rates, were combined into one, thereby resulting in a single-objective optimization and collapsing the Pareto-optimal front to a single point (see the Appendix); (ii) the optimization was carried out in PERL interpreted scripting language, which is extremely slow when compared to compiled languages like C. The important advance of this work over the earlier study is that the GA optimization implemented here is fully multiobjective—it cultivates the entire Pareto-optimal front throughout the simulation, enabling a complete picture of
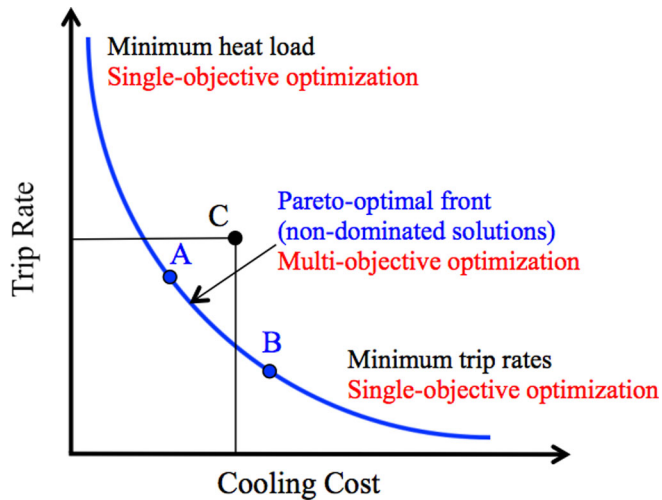
FIG. 2. The Pareto-optimal front for the simultaneous optimization of the linac's cooling cost and trip rate (multiobjective optimization). Asymptotically, the solutions at the extremes of the Pareto-optimal curve represent the single-objective minimization of the cooling cost (top left portion of the curve) and the single-objective minimization of the trip rates (bottom right portion of the curve). Solutions A and B are on the Pareto-optimal front, while solution C is not because it is dominated by solution A (A is within the rectangle defined by C's coordinates).

the trade-off between the two competing objectives. Furthermore, the GA-based code used here is written in C, leading to a substantial speedup over the earlier implementation in PERL and reaching the stringent operational specifications on execution time.

1. Increasing efficiency of the simulation through additional constraints

The constrained multiobjective problem given in Eq. (3) only features a single constraint specifying how the dependent variables (cavity gradients G_i) combine to satisfy the total energy gain requirement. When no explicit

constraints on the value of the objective functions $P(\vec{G})$ and $T(\vec{G})$ are given, the GA will explore the entire (infinite) Pareto-optimal front, thereby spending quite a bit of time in the regions which are of little or no physical importance (where the trip rates or the heat load are too high to be operationally useful). While such simulations are important from a pedagogical standpoint for establishing the traditional single-objective optimization as asymptotes of the multiobjective optimization presented here (see Fig. 2), they are operationally inefficient. Instead, introducing additional constraints on the objective functions to the optimization in Eq. (3) focuses the computations on the operationally interesting ranges:

$$\text{Subject to} \quad P(\vec{G}) \le P_{\max},$$
$$T(\vec{G}) \le T_{\max}, \qquad (9)$$

where $P_{\max}$ and $T_{\max}$ are specified at the beginning of the GA simulation. These constraints were treated in the same manner as the constraint on the total energy of the linac. Imposing these new constraints forgoes information on the extremes in favor of more relevant solutions emerging in the center of the Pareto-optimal front.

To the extent investigated here, the simulation with these extra constraints leads to an estimated Pareto front which is about as accurate as that produced without the additional constraints but using twice as many generations. Therefore, additional constraints of the objective functions reduce the computation time by about a half. This is illustrated in Fig. 3.

2. Reusing solutions from previous GA simulations

Another approach to improving the efficiency of the simulation is starting it with individuals produced by a previous simulation. This is implemented by treating individuals from a previously computed Pareto-optimal
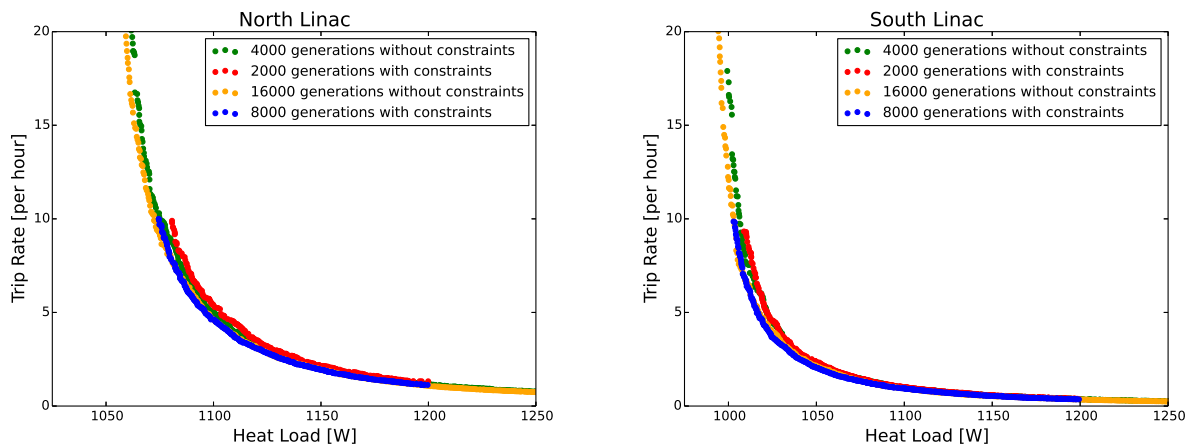

FIG. 3. Effect on the quality of the Pareto-optimal front of constraining the objective functions as in Eq. (9). Both North linac (left panel) and South linac (right panel) are constrained by $T_{\max} = 10$ trips per hour and $P_{\max} = 1200$ W.
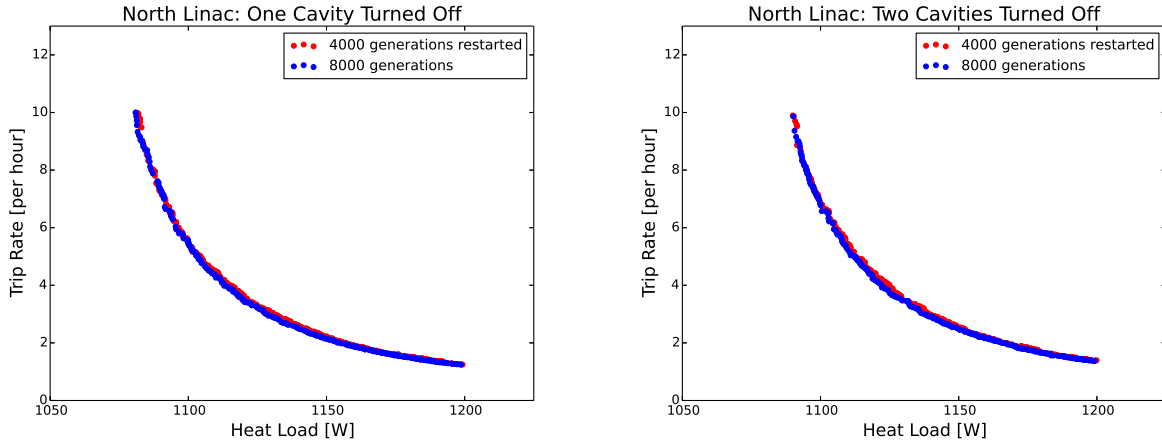
FIG. 4.   Effect on turning off one cavity (left panel) or two cavities (right panel) and restarting the simulation from the previous solution with all cavities on. The blue points denote the Pareto-optimal front produced by 8000 generations of the GA simulation started from a random sampling of the search space, and the red points show the 4000 generations of the GA simulation restarted from the solution to the full cavity problem with 4000 generations.

front as the initial generation of the new simulation, thereby essentially continuing where the last simulation left off.

This restart feature is particularly useful during the operation of the linac when one of the cavities needs to be turned off, usually due to frequent tripping or otherwise problematic behavior. With the problematic cavity turned off, the energy must be redistributed across the remaining cavities, which changes the optimization problem and thus the Pareto-optimal front. Starting such a new simulation from the solution of the full cavity problem (which one would have at the beginning of linac operations) significantly reduces the execution time. Figure 4 shows that the Pareto-optimal front for the reduced problem started from the Pareto-optimal front of the full cavity problem needs about a half as many generations of the GA simulation (and therefore half as long in terms of execution time) than if it was started from a random sampling of the search space. This is true for both one (left panel of Fig. 4) or two cavities turned off (right panel of Fig. 4).

3. Optimization of the PISA code

In PISA implementation, *variator* and *selector* run concurrently as two separate programs. They communicate through text files, thereby significantly slowing down the execution. In order to improve the efficiency of the PISA GA code so that it can be used as a high-level, real-time application for CEBAF operations, communication through shared memory replaces text file communication between the *variator* and *selector*. Shared memory may be simultaneously accessed by multiple programs with an intent to provide efficient means of passing data among them and avoid redundant copies. In a shared memory system all processors share a single view of data and the communication between processors can be as fast as memory accesses to a same location.

Replacing file communication by shared memory, as well as implementing other minor optimizations (streamlining function calls, restructuring the loops, etc.), improved program execution time by a factor of about two and a half. For instance, computing 1000 generations with 512 individuals on a 2.53 GHz Intel® Xeon® CPU with the original implementation takes 245 seconds, while the optimized version executes in 89 seconds.

Further optimizing the code by parallelizing it on clusters of cores comprised of CPUs and graphical processing units did not result in a marked improvement in performance because the intrinsically parallelizable portions of the code (i.e., objective function evaluation etc.) are not the computational bottlenecks for this problem.

IV. RESULTS

With the optimized version of the PISA GA platform, the constrained simultaneous cavity heat load and trip rates minimizations for CEBAF North and South linacs are independently run using 512 individuals evolved over a maximum of 16000 generations. The results are given in Figs. 5 and 6, which depict, for each linac, the Pareto-optimal front (top left panels) and, for the three selected solutions from the front, the conserved quantities from the single-objective minimizations given in Eq. (8)—G_i/Q_i shown in red and $B_i \exp\left[A + B_i(G_i - F_i)\right]$ shown in blue— just as in Fig. 1. The extreme points on the Pareto-optimal front appear quite similar to the single-objective optimization results shown in Fig. 1: for each linac, compare solution A to the top row of Fig. 1 and solution C to the bottom row. This, of course, is expected: even though they are solutions of the multiobjective problem, these extreme points so strongly favor one objective over the other that they can be viewed as asymptotically equivalent to the solution of the single-objective
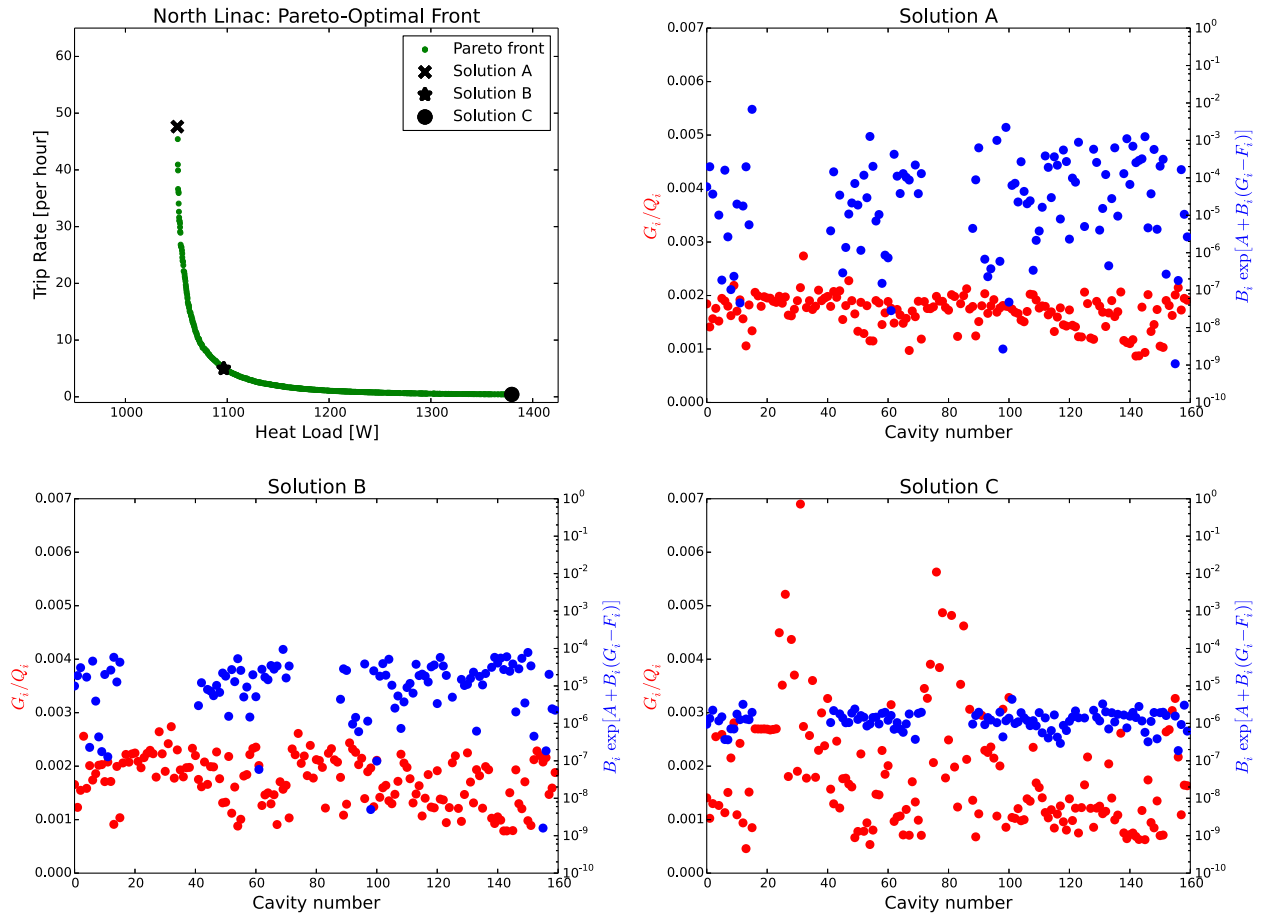
FIG. 5.   Pareto-optimal front for North linac obtained using the GA simulation with 512 individuals evolved for 16000 generations (top left panel). Two of the three solutions marked are closely related to the asymptotic solutions obtained by single-objective minimization of the heat load only (solution A) and trip rates only (solution C). Also shown is a solution from the Pareto-optimal front which has a trip rate of exactly five (solution B). The remaining panels show the single-objective conserved quantities for solutions A, B, and C, as in Fig. 1. Note the red outliers in the bottom right panel.

optimization. It is evident that the solutions which more strongly favor minimization of the heat load will have the quantities G_i/Q_i bunched around a constant value and obversely for minimized of trip rates and $B_i \exp\left[A + B_i(G_i - F_i)\right]$ values. A "goldilocks" solution, one that balances the two objective functions (akin to solution B), will have both quantities bunched around constant values, but with larger scatter.

The extreme solutions A and C provide estimates to the operational lower limit for the heat load (A) and the trip rates (C). For the North linac, the minimum attainable heat load is 1048 W, and the minimum trip rate is 0.41 per hour. For the South linac, these are 988 W and 0.13 trips per hour. It is easy to see then that the judicious solution B is within 4.5% of the minimum heat load for the North linac, and 2.9% for the South linac. It is reassuring that the GA-based optimization allows us to reach the attainable minimum within such a small margin.

Operationally, plots of the conserved quantities can serve as diagnostic tools, since they can quickly identify the cavities which are significantly away from the constant value around which they should be bunched. These outlying cavities are noticeably displaced red dots in both solutions A and B for the South linac (Fig. 6). These cavities have low quality factor (Q_i) values. It is interesting to note that some of these "offending" outlying cavities even when driven at the extreme values of their operational ranges will not fall in line near the constant value.

A. Quantifying the improvement over earlier implementations

By design, the multiobjective optimization presented here is expected to be superior to the earlier single-objective optimizations. Simply put, a multiobjective optimization subsumes any single-objective optimization as its special case, an asymptotic solution. The LEM tool currently in operation at CEBAF does not take in consideration heat load at all, and bases all of its optimization decisions on trip rates of the cavities for which the empirical models exist.
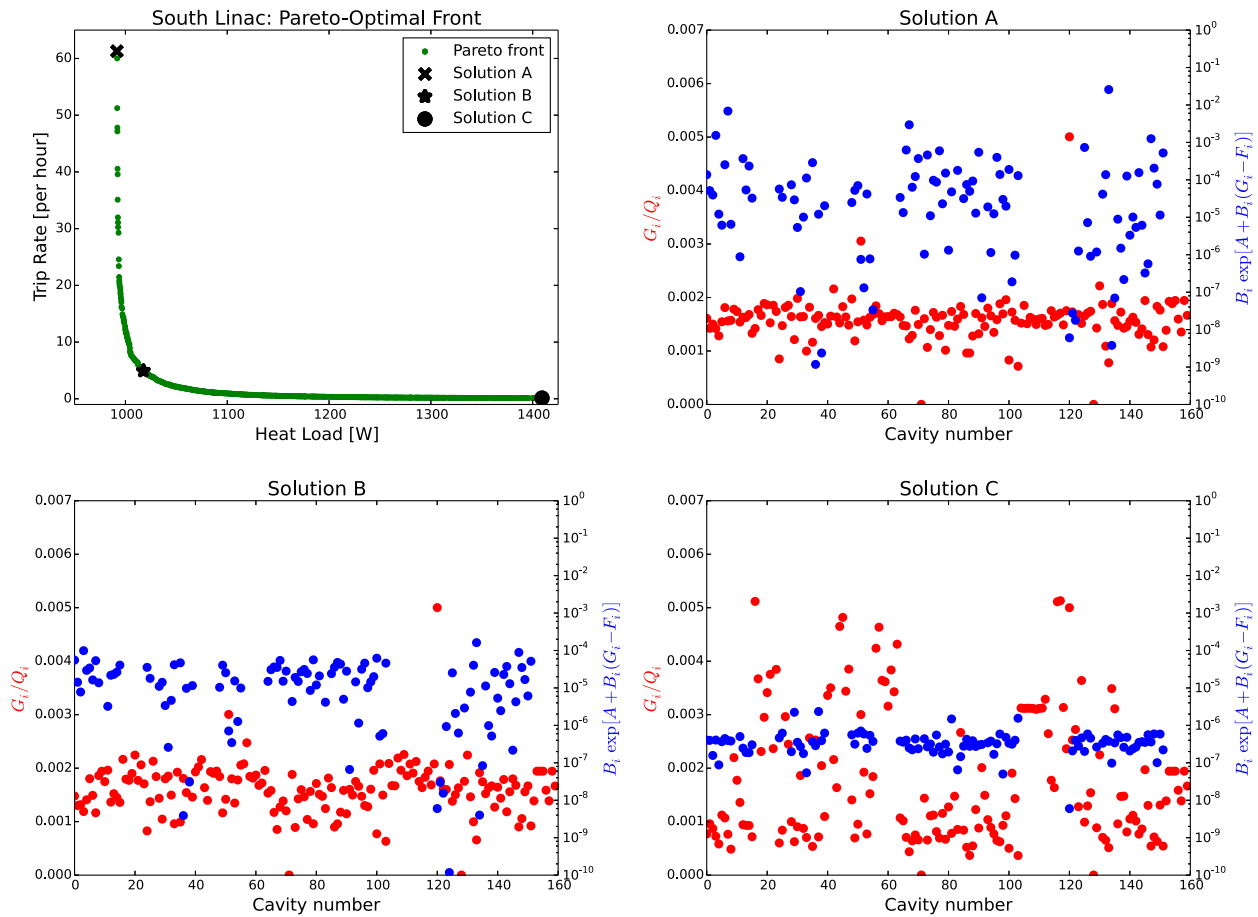
FIG. 6. Same as Fig. 5, except for the South linac. Note the red outliers in solutions A and B.

A prototype of an improvement over the LEM code presented in [12] used a GA-based method for a single-objective optimization, in which the heat load was minimized given a fixed trip rate enforced as a constraint. This approach resulted in an appreciable improvement over the existing LEM tool: it reduced the overall dynamic heat load of the two linacs by about 10%, from 2689 W to 2435 W.

Figure 7 compares the Pareto-optimal front computed using this GA-based multiobjective optimization with the best results of the GA-based single-objective optimization from [12], which had a number of trips per hour fixed at five. A cursory point-to-point comparison reveals that new GA implementation with 16000 generations and 512 individuals yields the minimum heat load of 1097 W (1018 W) for the
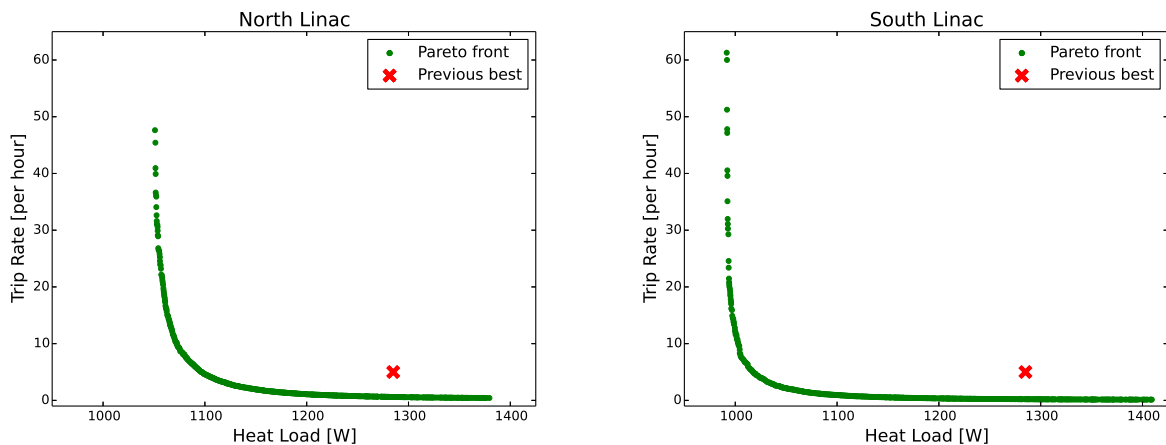

FIG. 7. Pareto-optimal front computed using a GA-based multiobjective optimization with 16000 generations and 512 individuals versus the single-objective GA-based optimization reported in [12].

North linac (South linac), which is 15% (12%) better than the best solution found in [12] (1285 W and 1150 W, respectively). Furthermore, the multiobjective optimization here not only provides a superior solution for the same trip rate as the single solution found previously, but also yields an entire set of feasible solutions for a full range of trip rates. This is particularly helpful in operations when the Pareto-optimal front can be used to estimate the marginal change in operational cost with the change in trip rates.

Combining the solutions for the North and South linacs found using the GA-based multiobjective approach, the total heat load at five trips per hour is 2115 W, which is 13% (325 W) lower than the prototype GA-based solution from [12], and 21% (579 W) lower than the operational point computed by the LEM tool at which the run was executed.

B. Convergence of the Pareto-optimal front

The estimate of the Pareto-optimal front for each generation is a collection of the best nondominated solutions from all the previous generations. Naturally, as the GA-based optimization is evolved, newer generations will produce new solutions which will dominate and push out some older solutions from the Pareto-optimal front. This, in turn, will lead to an evolution in the Pareto-optimal front itself. At some point, after evolving the GA-based optimization through enough generations, the Pareto-optimal front is expected to converge to the optimal solution of the problem, an approximation to the *true* Pareto-optimal front. This true Pareto-optimal front, of course, is not known—the best that one can hope to attain is a good approximation by ascertaining that the Pareto-optimal fronts produced by the GA-based optimization have converged. This is shown in Fig. 8. It is difficult to resolve the differences between the Pareto-optimal fronts after 8000 and 16000 generations, which strongly indicates its convergence.

From the operational standpoint, the difference in optimal solutions obtained after 1000 (2000) generations and 16000 generations is about 0.5% (1%) of the total heat

load. Given that the amount of extra work needed to "squeeze out" this final 0.5% (1%) of performance is eightfold (sixteenfold)—since the code execution time scales linearly with the number of generations—it is clear that for operational purposes the accuracy of the simulation with 1000 generations will suffice. The execution time of our optimized version for 1000 generation is on the order of 90 seconds, close to the specifications of the machine operations tool.

The computational load is further halved when constraints on the objective functions are implemented as discussed in Sec. III B 1, leading to a solution within 1% of the reported optimum within about 90 seconds of execution time on a 2.53 GHz Intel® Xeon® CPU.

C. Sensitivity to measurement error

From the operational standpoint, it is desirable to estimate how sensitive the optimal machine configuration is to the errors in measurement of the cavity gradients G_i or the cavity quality factor Q_i. It is expected that the measurement errors are on the order of 10% for the cavity quality factor and about 5% for the gradient errors [3].

To simulate these errors 10000 different machine configurations are generated with 10% (5%) random fluctuations in the measured cavity quality factor Q_i (measured cavity gradients G_i). The solutions in the Pareto-optimal front are recalculated using these new noisy cavity Q_i and G_i and compared to the unadulterated GA-derived Pareto-optimal front solutions. The resulting plot, shown in Fig. 9, illustrates the effects of these measurements errors, considered one at a time, on the Pareto-optimal front. Red points denote the original Pareto-optimal front, blue points the solutions that do not satisfy the energy constraint, and green points the solutions that do satisfy the energy constraint. It is evident that the errors in the cavity quality factor Q_i only affect the heat load, since the trip rate model is independent of this quantity, causing the horizontal smear symmetric around the original, unperturbed Pareto-optimal
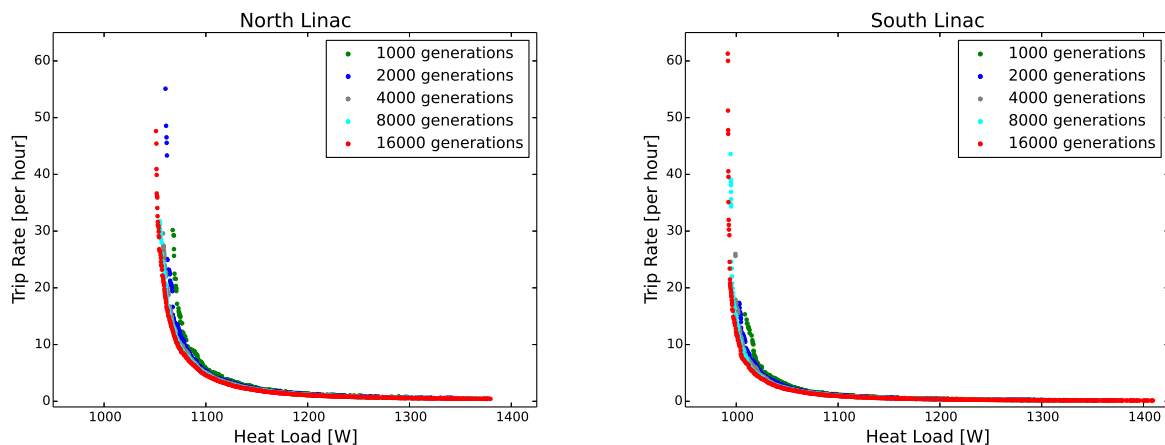

FIG. 8.   Convergence of the Pareto-optimal fronts for the North (left) and South (right) linacs, as a function of the number of generations in a GA simulation.
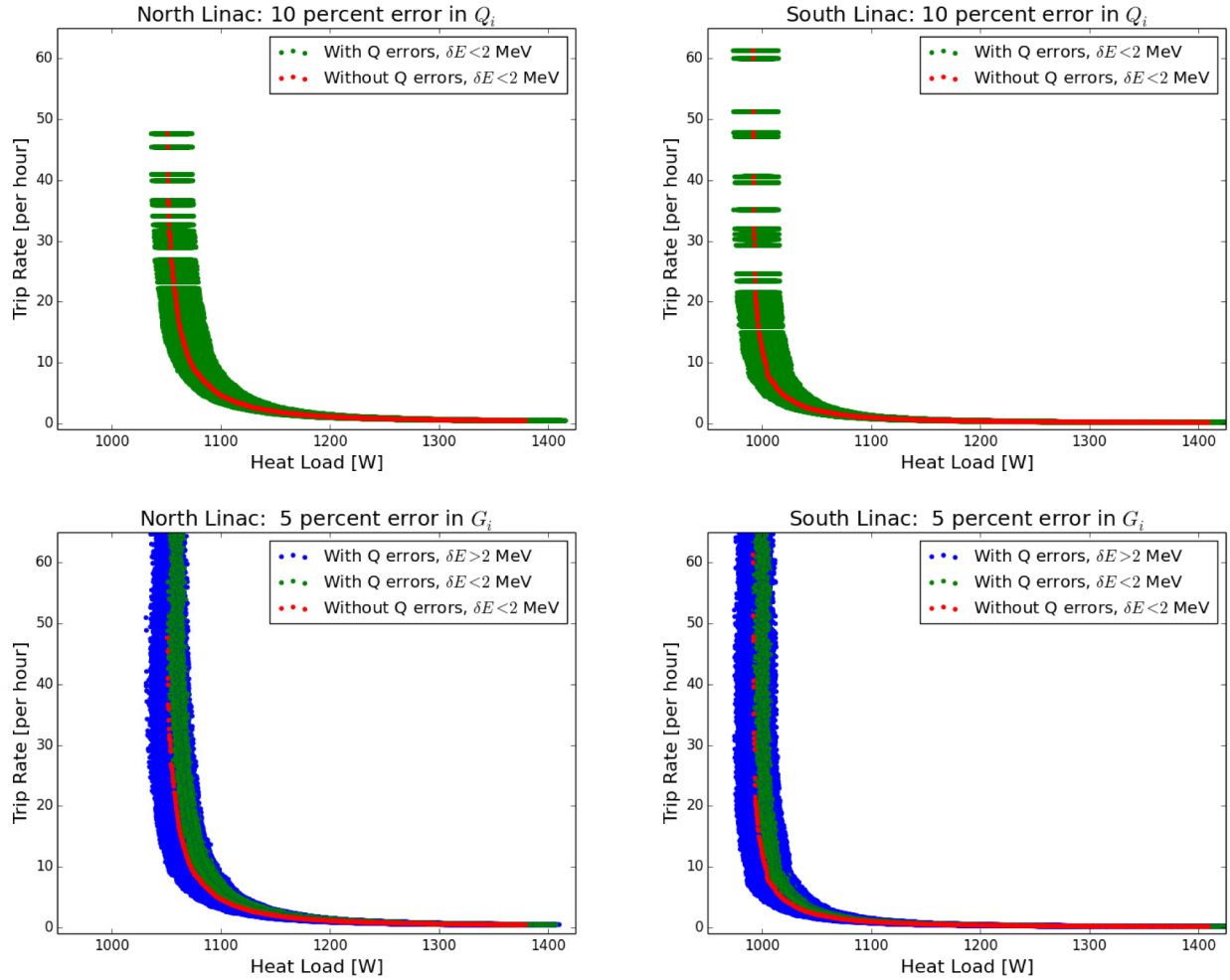
FIG. 9.   Sensitivity on the Pareto-optimal front to a 10% measurement error in the cavity quality factor Q_i for the North (top left) and South linacs (top right) and a 5% measurement error in gradients G_i for the North (bottom left) and South linacs (bottom right). The blue (energy constraint not satisfied) and green (energy constraint satisfied) represent 10000 different machine configurations with measurement errors artificially introduced. The (unperturbed) Pareto-optimal fronts (red) are the same as those in Figs. 5 and 6, generated with 512 individuals evolved for 16000 generations.

front. The energy constraint is also independent of Q_i, as reflected in Eq. (3) and Fig. 9. For the errors in the measured gradient G_i, the smearing around the unperturbed Pareto-optimal is dominated with the solutions that violate the energy constraint; the solutions which still obey the energy condition all have slightly larger heat load consumption.

V. DISCUSSION AND SUMMARY

A GA-based optimization has been developed to simultaneously minimize the dynamic cavity heat load and trip rates of the Jefferson Lab's CEBAF linacs. A numerical simulation using this approach is demonstrated to be superior to the one-dimensional algorithm based on Lagrange multipliers. At the same trip rate and overall energy gain for the linac, the simulated dynamic heat load is reduced by over 20%. While very important in its own right, this study also serves as a proof of principle that GA-based multiobjective, multidimensional optimization can result in

a substantial improvement in efficiency of other linac-based machines, leading to significant savings in operations cost.

The study presented here is based on the numerical simulation of the data of one of the old CEBAF runs, 2009 PVDIS experiment from its 6 GeV era. Its ultimate goal is to pave the way toward a real-time operations tool for the new 12 GeV CEBAF era experiments to replace the existing LEM program. It represents a lion's share of the new operations tool which is currently being developed.

The main reason why our new, superior optimization algorithm has not yet been incorporated into the regular CEBAF operations is unavailability of the data for the new, recently upgraded 12 GeV machine. As the data become available, a numerical simulation with the new data and its implementation on the CEBAF machine will be reported in a separate publication.

From the implementation standpoint, extending the formalism presented in this paper for the old 6 GeV era

CEBAF run with 160 cavities in each linac to the 12 GeV era CEBAF configuration with 200 cavities in each linac—or to any other linac for that matter—is straightforward. The dimensionality of the problem would change accommodating the varying number of cavities in a linac, requiring only a minor modification of the simulation input file. Integrating additional constraints, providing alternate models for the heat load or the trip rates, introducing dependence of the cavity quality factor Q_i on cavity gradient G, or other model-related enhancements would require only minor modification of the source code.

APPENDIX: CASTING A MULTIOBJECTIVE OPTIMIZATON AS A SET OF SINGLE-OBJECTIVE OPTIMIZATIONS

In practice, it is often the case that the two competing objectives are linearly combined into one, thereby reducing the problem from a multiobjective to a more easily tractable single objective. Such a simplification collapses the set of solutions of the multiobjective problem to a single point—a tangential intersection of the coordinate system $c_1 f_1 + c_2 f_2 = $ const and the Pareto-optimal front of the full multidimensional problem (see the left panel of Fig. 10). Since solving a single-objective optimization problem is considerably simpler and much better supported in terms of solving techniques and corresponding literature, it seems like a natural step forward to replace a multiobjective problem with a set of single-objective optimization problems. Since each vector of parameters (c_1, c_2) leads to a single point on the Pareto-optimal front, N single-objective simulations, each based on a different, mutually linearly independent vector, approximates the front with N points (as in the right panel of Fig. 10). Exact normalization of (c_1, c_2) vectors is not necessary, as long as they are linearly independent. Without loss of generality, consider the proposed scheme $c_1 + c_2 = 1$ and $0 \leq c_1, c_2 \leq 1$ that conveniently makes contact with the single-objective optimizations of f_1 alone ($c_1 = 1$, $c_2 = 0$) and f_2 alone ($c_1 = 0$, $c_2 = 1$). The results for the single-objective simulation equivalents of one single multiobjective CEBAF linac optimization are shown in the right panel of Fig. 10. For the same number of generations, a single-objective optimization of $c_1 P + c_2 T$ for individual values of (c_1, c_2) provides a better estimate to the Pareto-optimal front because it does not expend computational effort on issues particular to having multiple objectives (expanding the limits of the Pareto-optimal front). In this example, the estimates of the Pareto-optimal front using a set of single-objective optimizations with 1000 generations are about as accurate as the multiobjective optimization with 4000 generations in estimating Pareto-optimal front.
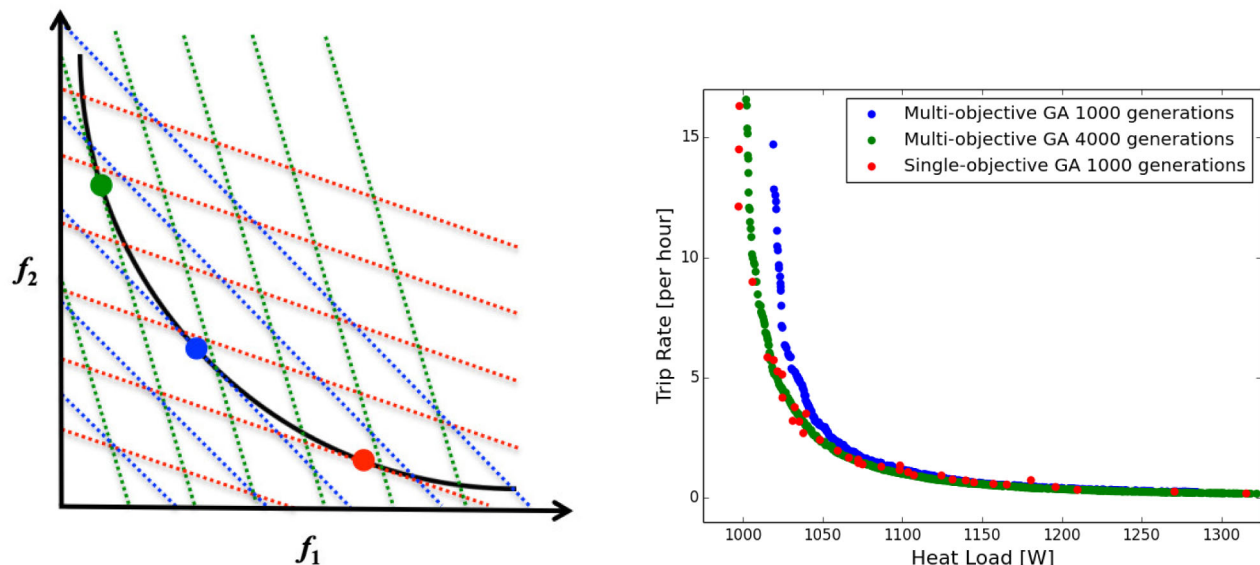

FIG. 10. Left: Relationship between the Pareto-optimal front for the multiobjective optimization of two objective functions f_1 and f_2, and the single-objective optimization of their linear combination $c_1 f_1 + c_2 f_2$. Three different combinations of (c_1, c_2) parameters are shown in green, blue and red. Right: Comparison of the Pareto-optimal fronts from the multiobjective GA simulations (blue and green points) with the results of the corresponding set of single-objective optimizations of $c_1 P + c_2 T$ with varying values of (c_1, c_2) (red points). The simulation is for the CEBAF's South linac.

[1] J. Benesch, http://arxiv.org/abs/physics/0606141.

[2] J. Delayen, L. Doolittle, and C. Reece, in *Proceedings of the 18th Particle Accelerator Conference, New York, 1999* (IEEE, New York, 1999).

[3] J. Benesch, M. Bickley, J. Creel, M. Drury, A. Freyberger, M. Wright, C. Reece, and D. Turner, Jefferson Lab Technical Note No. JLAB-TN-09-041, 2009.

[4] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, Institut fur Technische Informatik und Kommunikationsnetze Technical Report No. 154, 2002.

[5] S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler, in *Evolutionary Multi-Criterion Optimization (EMO 2003)*, edited by C. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Lecture Notes in Computer Science (Springer, New York, 2003), pp. 494–508.

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

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

[8] B. Bartolini, M. Appolonio, and I. P. S. Martin, Phys. Rev. ST Accel. Beams **15**, 030701 (2012).

[9] B. Terzić, K. Dietrick, A. Hofler, and G. Krafft, Phys. Rev. Lett. **112**, 074801 (2014).

[10] E. Zitzler, M. Laumanns, and L. Thiele, in *Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2001)*, edited by K. Giannakoglou, D. Tsahalis, J. Periaux, K. Papailiou, and T. Fogarty (International Center for Numerical Methods in Engineering (CIMNE), Barcelona, Spain, 2002), pp. 95–100.

[11] A. Hofler, Ph.D. thesis, Old Dominion University, 2012.

[12] A. Freyberger, Jefferson Lab Technical Report No. JLAB-TN-12-057, 2012.