Bayesian Algorithm Execution for Tuning Particle Accelerator Emittance with a Virtual Objective

Sara Ayoub Miskovich, Willie Neiswanger, William Colocho, Claudio Emma, Jacqueline Garrahan, Timothy Maxwell, Christopher Mayes, Stefano Ermon, Auralee Edelen, and Daniel Ratner

1SLAC National Laboratory, Menlo Park, California 94025, USA
2Department of Computer Science, Stanford University, Stanford, California 94305, USA

Traditional black-box optimization methods are inefficient when dealing with multi-point queries, i.e. when each query of the objective requires multiple secondary measurements, simulations, or other tasks. Existing approaches, including Bayesian optimization (BO), acquire the full series of measurements at each iteration, making the queries slow and information-poor. We propose applying Bayesian Algorithm Execution (BAX) to instead query and model individual measurements. BAX avoids the slow multi-point query by acquiring points through a virtual objective, i.e. calculating the multi-point objective from the learned model rather than from the experiment. As a result, queries in BAX are faster and retain more information compared to those in BO. In this work, we use BAX to minimize emittance at the Linac Coherent Light Source (LCLS) and the Facility for Advanced Accelerator Experimental Tests II (FACET-II) particle accelerators. Although the emittance is a critical parameter for the performance of high-brightness machines, including X-ray lasers and linear colliders, optimization is often limited by the time required for tuning. In an LCLS simulation environment, we show that BAX is 20× faster while also being more robust to noise compared to existing optimization methods. In live tests, BAX performed the first fully-automated emittance tuning at both LCLS and FACET-II, matching the hand-tuned emittance at FACET-II and achieving an optimal emittance 24% lower than that obtained by hand-tuning at LCLS. We anticipate that our approach can readily be adapted to other types of optimization problems involving multi-point queries commonly found in scientific instruments.

I. INTRODUCTION

Black-box optimization is a general approach for systems with unknown behavior [1, 2]. Examples of black-box optimizers include Nelder-Mead simplex, coordinate descent, genetic algorithms, and many others. Bayesian optimization (BO) [3, 4] is an especially appealing approach for black-box systems in which each function query is expensive (in time or other costs), a common situation for scientific and industrial instruments. BO has been used extensively for optimization of scientific instruments, including particle accelerators [5, 9].

For applications where the optimization objective is not a direct system observable, but rather is computed from a series of secondary measurements, even BO may be too inefficient. For example, a single query might require a scan of another control variable, a series of expensive computations, or an internal optimization process. We refer to this type of query as a multi-point query. Black-box optimization tasks with multi-point queries are encountered in a variety of fields, such as material science [10], nuclear astrophysics [11], aerospace engineering [12], bioinformatics [13], in addition to particle accelerators [14]. When dealing with such tasks, traditional optimization methods, including BO, exhibit two types of inefficiencies. First, each query requires a set number of secondary measurement points to calculate the objective, even if a subset of those points already reveals the system settings to be suboptimal. Second, because the black-box function returns only the objective to the optimization algorithm, the full information acquired within the black-box function evaluation (e.g. the full series of secondary measurements) is not shared across different optimization steps, and this information loss leads to inefficient sampling. The result is that each multi-point query is both expensive and information-poor. When time is a limiting factor, optimizer inefficiency may reduce the number of control variables that can be optimized, curtail the ability to find a global optimum, or make the optimization task too costly to even attempt in practice. Optimizer inefficiency is particularly problematic for scientific instruments that are in high demand.

The motivating example for this paper is beam-emittance minimization in particle accelerators. The emittance of a charged particle beam, defined as the volume of particles in position-momentum phase space [15], is an important parameter for a variety of accelerator applications. For accelerator-based light sources, emittance determines the X-ray beam brightness, limiting the shortest wavelength available at X-ray free-electron lasers (XFELs) [16] and affecting the output X-ray power by orders of magnitude [17]. It is especially important for the next generation of undulator designs (e.g. LCLS-II-HE) [18]. For colliders, small emittance is critical to maximize luminosity [19, 20]. Beams with high emittance ratios are needed for a variety of new light source and collider designs [21], and also for injection into components of compact accelerators, such as dielectric wakefield accelerating cavities [22].

The accelerator settings need to be periodically re-
optimized, or tuned, to achieve acceptable emittance values. Emittance depends on the initial beam conditions at the particle source (e.g. the photocathode and associated drive laser for electron accelerators), along with the combined effects of adjustable accelerator settings. The initial conditions can change as a result of both intended beam property adjustments (e.g. changes to the beam charge) and unintended drift over time (e.g. uncontrolled time-varying changes to the drive laser output). Given the importance of the beam emittance to accelerator performance, emittance optimization would ideally be a routine task for accelerator operations.

However, one of the most common methods for determining the transverse emittance involves scanning a focusing magnet’s strength while observing the resulting change in beam size [15], i.e. the emittance is calculated from a multi-point query. Each step of the optimizer thus involves choosing a configuration in the control domain (the settings of the accelerator), then taking a set of measurements over a defined range in the secondary domain (in this case, beam-size measurements at different magnet settings), and finally fitting the measured points in the secondary domain to calculate an emittance value to the optimization algorithm. The inefficiency of measuring emittance severely limits the number of tuning opportunities at most accelerator facilities, despite the importance of achieving small emittance. For example, at the Linac Coherent Light Source (LCLS) [16], injector emittance is typically tuned only after returning from a machine shut-down or a substantial change to the target beam parameters (such as the charge), and it is done almost entirely manually by human experts. There is a clear need for more efficient algorithms for emittance optimization.

We propose adapting a recently-developed method, Bayesian Algorithm Execution (BAX) [23], as a sample-efficient technique for optimization tasks with multi-point queries. In contrast to BO, which directly models the objective function, BAX builds a system surrogate model in the joint control-measurement domain, defined by the Cartesian product of the optimization’s control and secondary domain. BAX then guides the acquisition with a virtual objective, i.e. calculated on the surrogate model rather than obtained directly from the machine. By modeling the joint control-measurement domain, as opposed to modeling only the objective of a multi-point query, BAX maximizes the information gained from individual measurements. In addition, by employing a virtual objective, BAX can query individual points in the joint control-measurement domain, avoiding the need for multiple measurements at each step. We emphasize that querying and modeling direct observables (e.g. beam sizes using BAX), rather than the objective (e.g. emittance using BO) represents a significant conceptual shift in black-box optimization with multi-point queries.

In summary, this paper extends the BAX algorithm to the challenging and outstanding problem of emittance tuning in accelerators, with substantial improvement in sample-efficiency. First, we describe how BAX can improve efficiency for general optimization tasks involving multi-point queries. Second, for the specific case of emittance optimization, we show BAX to be sample-efficient compared to both BO and the Nelder-Mead simplex algorithm [24], using a simulation environment of the LCLS injector. Finally, we show experimental results for BAX in minimization of electron-beam emittance at both LCLS and the Facility for Advanced Accelerator Experimental Tests II (FAET-II) [25] at SLAC National Accelerator Laboratory (SLAC), with the former achieving 24% lower emittance than was found by manual tuning. We anticipate this approach to be widely adaptable to optimization problems involving multi-point queries on real-world instruments in science and engineering.

II. EMITTANCE CALCULATION AND OPTIMIZATION

Despite its importance to the lasing process, emittance optimization is not a part of standard daily tuning. Tuning time is limited by the need to minimize beam interruptions to users, which roughly translates into a limit on the number of queries to the system available for each tuning event. Emittance tuning is particularly slow because the exact response of the emittance to accelerator settings is unknown, and thus emittance is typically tuned as a black-box, iteratively adjusting the injector settings and observing the outcome. As a result, emittance tuning is typically limited to recovery from scheduled down-times and when switching to non-standard beam setups.

At LCLS, the three primary emittance-tuning variables are the solenoid magnet (SOL1) and two corrector quadrupoles (CQ1 and SQ1) [26]. To run an emittance calculation for some injector setting configuration, a “quadrupole scan” [15] changes a quadrupole (denoted Q5) while measuring the beam size on a wire. Emittance can then be calculated from the parabolic behavior of beam size with respect to magnet settings. Figure 1 shows a simplified layout of the LCLS injector and emittance calculation setup. Note that while single-shot emittance measurements for accelerators do exist (e.g. multi-slit and pepper-pot masks [27,30]), they are generally not suitable for all types of beams and beamline layouts, including beams at the LCLS.

With standard black-box emittance optimization, each step involves choosing a new control variable configuration (SOL1, CQ1, and SQ1) [26]. To run an emittance calculation for some injector setting configuration, a “quadrupole scan” [15] changes a quadrupole (denoted Q5) while measuring the beam size on a wire. Emittance can then be calculated from the parabolic behavior of beam size with respect to magnet settings. Figure 1 shows a simplified layout of the LCLS injector and emittance calculation setup. Note that while single-shot emittance measurements for accelerators do exist (e.g. multi-slit and pepper-pot masks [27,30]), they are generally not suitable for all types of beams and beamline layouts, including beams at the LCLS.

With standard black-box emittance optimization, each step involves choosing a new control variable configuration (SOL1, CQ1, and SQ1) [26]. To run an emittance calculation for some injector setting configuration, a “quadrupole scan” [15] changes a quadrupole (denoted Q5) while measuring the beam size on a wire. Emittance can then be calculated from the parabolic behavior of beam size with respect to magnet settings. Figure 1 shows a simplified layout of the LCLS injector and emittance calculation setup. Note that while single-shot emittance measurements for accelerators do exist (e.g. multi-slit and pepper-pot masks [27,30]), they are generally not suitable for all types of beams and beamline layouts, including beams at the LCLS.

With standard black-box emittance optimization, each step involves choosing a new control variable configuration (SOL1, CQ1, and SQ1) [26]. To run an emittance calculation for some injector setting configuration, a “quadrupole scan” [15] changes a quadrupole (denoted Q5) while measuring the beam size on a wire. Emittance can then be calculated from the parabolic behavior of beam size with respect to magnet settings. Figure 1 shows a simplified layout of the LCLS injector and emittance calculation setup. Note that while single-shot emittance measurements for accelerators do exist (e.g. multi-slit and pepper-pot masks [27,30]), they are generally not suitable for all types of beams and beamline layouts, including beams at the LCLS.
III. BAX FOR EFFICIENT EMITTANCE OPTIMIZATION

To improve the efficiency of optimization with multipoint queries, we propose to model the joint control-measurement domain, e.g., to learn a model of beam size as a function of both control and measurement variables. Figure 2 (b) presents an illustration of how joint optimization increases the information gained from each individual beam-size measurement. In joint optimization, we require a method that minimizes a computed function (e.g., emittance) operating on a learned model of a direct observable (e.g., beam size). However, standard model-based optimization methods such as BO are restricted to modeling and optimizing the same quantity. In the case of emittance tuning, standard BO can only model and optimize the computed emittance from a full quadrupole scan at each query point in the control domain.

Instead of standard BO, we employ an information-based BAX method \cite{23, 31, 32}, which lets us model a direct observable and optimize the output of any function computed from that model, rather than the model itself. By using BAX, we convert the joint optimization problem into a type of adaptive sampling (experimental design) \cite{33, 34} problem in a joint control-measurement domain, as shown in Fig. 2. For each query from the control domain, instead of making a full scan in the second domain, we measure a single point. In this manner, each function query is cheap while still performing the same optimization problem, i.e., finding a point in the control domain that optimizes the black-box function. Full details on the BAX algorithm are given in Appendix C.

In the LCLS emittance application, BAX models the beam size with respect to the SOL1, CQ1, and SQ1 control variables and the Q5 quadrupole measurement variable. The full BAX emittance optimization procedure is illustrated in Fig. 4. The procedure is initialized with randomly sampled points from the joint control-measurement domain. Then, at each iteration, the BAX algorithm queries the beam size at a single value of the Q5, SOL1, CQ1 and SQ1 variables, constructing a model of the beam size as a function of both control and measurement variables. This internal model represents the prior distribution on the function describing the beam-size response to the quadrupole strength, and is implemented in this work by a Gaussian process model (GP) with a radial basis function kernel.

During the internal step (Fig. 4b), BAX executes the full emittance algorithm on a predefined number of quadrupole scans drawn from the model’s posterior distribution, with each virtual scan producing a virtual objective. Finally, the information-based BAX acquisition
One potential hurdle for the implementation of the BAX algorithm is the computational complexity, as each iteration requires repeated calculation of the virtual objective on draws from the model posterior, which in our case can require a large number of emittance calculations. Although this work did not focus on computational efficiency of the algorithm, in cases where the virtual objective is expensive, parallelization of the BAX algorithm execution procedure can significantly reduce wall-clock time. During live optimization at LCLS, a parallelized version using five CPU cores takes 4.7 seconds, on average, to select a beam-size function query at each iteration.

IV. RESULTS OF OPTIMIZATION IN NOISY LCLS SIMULATION SETTING

We study the performance of BAX in emittance optimization using a surrogate model of the LCLS copper injector trained on IMPACT-T simulation data. Details of this model are presented in Appendix D. The BAX procedure queries the beam size from the LCLS surrogate model given a configuration of virtual injector control variables for SOL1, CQ1 and SQ1, as well as a single value for the Q5 strength. For each run, BAX was initialized with 10 points that were uniformly sampled from the joint control-measurement domain defined by the bounds set on each variable domain: SOL1: (0.46, 0.485) (kG·m), CQ1 and SQ1: (-0.02 0.02) (kG). All tests ran for 200 iterations, corresponding to 200 total beam size queries from the simulation. Each algorithm execution on the posterior (i.e. the virtual objective) involves a scan of 10 initial points, followed by 10 final points in both X and Y after adjusting the scan range based on the minimum of the parabola (see Appendix B). To assess BAX’s performance, we repeat the optimization with black-box methods operating directly on the multi-point query. We chose BO and Nelder-Mead simplex as baselines, as both have seen extensive use in accelerator operations. Details of the BAX, BO, and simplex algorithm implementations are given in Appendices C and E.

For each algorithm, we ran 40 tests, each with different initial samples. For all simulated measurements,
Gaussian noise of the form $N(0, \sigma_n)$ was added to the beam size for each function query, with $\sigma_n = 10\%$ as a conservative estimate of typical experimental noise levels at LCLS (see Appendix A). Figure 5 shows the mean and 2-sigma standard error of 40 individual runs with randomized starting conditions are shown for each algorithm. The best emittance seen at each iteration is plotted, and the values are normalized by the minimum of the ground truth emittance. Bottom: A magnified view of BAX optimization results (blue), with the error on the BAX beam-size prediction compared to the true (observed) beam size in the LCLS surrogate model are shown on the right-hand axis (purple).

We can further check that the BAX GP model learns the correct behavior of the beam size during scans over the secondary domain. Figure 6 shows samples of the GP posterior during iterations 40 and 100, with the control variables set to the optimal configuration that BAX estimates at that given iteration. By iteration 100, BAX models the beam-size response to a quadrupole scan with high accuracy.

BAX significantly surpasses BO and simplex in finding an optimal emittance. While each BO and simplex iteration corresponds to an average of 18 beam-size measurements, each BAX iteration is a single measurement. In this noisy environment, BAX finds the optimal emittance after about 160 measurements on average, while the mean of the BO runs takes over 3000. With each beam-size measurement taking 18 seconds on a wire scanner at LCLS, BAX would require 48 minutes of invasive measurements compared to 15 hours with BO.

We hypothesize that BAX should exhibit increased robustness to noise because its GP directly models the beam size as opposed to the calculated emittance. To test this robustness, we repeated the previous experiment with both a noisier ($\sigma_n = 30\%$) and a less noisy ($\sigma_n = 5\%$) beam-size function. The results are shown in Fig. 7. The increased noise leaves BAX’s performance largely unchanged (blue), while degrading BO’s performance significantly (orange).

V. RESULTS OF LIVE EXPERIMENTAL OPTIMIZATION AT LCLS AND FACET-II

We applied BAX to the online optimization of the beam emittance in the LCLS injector with an identical setup to the simulated environment. The injector control variables were the solenoid SOL1, and the two quadrupoles CQ1 and SQ1. Emittance calculations scanned the Q5 quadrupole while measuring beam size with a downstream wire scanner. Details of the experi-
FIG. 7. Robustness to noise during optimization. Comparison of the optimization performance on the LCLS injector surrogate model with varying levels of noise added to the beam size for BAX (blue shades) and BO (orange shades). The best emittance up to that iteration is shown, normalized by the minimum of the ground truth emittance. The lower plot shows a magnified view of BAX optimization.

mental setup are given in Appendix A. The bounds on the SOL1 device domain were identical to the simulation runs, (0.46, 0.485) (kG·m), while CQ1 and SQ1’s domains were made smaller to (-0.015, 0.015) (kG) due to constraints on the region where beam-size measurements were valid and within range of the wire scanner. In future work, such constraints can be incorporated in the optimization to learn to avoid bad or invalid regions without having to place tight restrictions on the control domains.

Figure 8 shows the resulting emittance optimization as a function of the number of beam-size measurements, initialized with 10 random points. To assess the performance of BAX, full experimental emittance scans were performed every 10 iterations at the optimal control variables found by BAX at that iteration (blue). The full emittance scans are slow and are only needed to assess convergence for the study. Because an operator would not have access to such scans in practice, it would not be possible to select an earlier setting with better emittance as would be done with BO. Therefore we show an upper bound on emittance at each point (green), i.e. the worst emittance seen after that iteration (green), the error on the BAX beam-size prediction compared to the true beam size (purple), and the best hand-tuned emittance on that day for reference (red).

We can see that BAX converges to small beam-size errors and a low emittance value after approximately 70 iterations, corresponding to about 20 minutes of beam time. We note that neither the matching quadrupoles nor the match were included in this optimization.

The BAX behavior in this experimental setting is consistent with the simulation study on the injector surrogate model. Due to the inefficiency of traditional optimizers, their performance could not be compared to BAX on the LCLS machine, as a single BO run was expected to take several hours based on previous experience, and beam time at LCLS is in high demand. Nonetheless,
these results are a strong indication that BAX performance surpasses standard black-box optimization methods in noisy experimental settings, just as we observed in the simulation environment.

In addition to the experimental test on LCLS, we applied the BAX procedure to the optimization of the electron-beam emittance on FACET-II. The FACET-II injector line has an identical setup to the LCLS beam line, and we again optimized SOL1, CQ1 and SQ1. However, FACET-II has a charge of 2 nC, compared to 250 pC in the LCLS run. The control domain bounds were (0.37, 0.41) (kG-m) for SOL1, and (-0.01 0.01) (kG) for CQ1 and SQ1. Here, optical transition radiation (OTR) screens were used instead of wire scanners for measuring the beam size downstream of Q5 (see Appendix A).

BAX was initialized with 10 random beam-size measurements, and full emittance evaluations were performed every 10 iterations using the current optimal control variables. Figure 11 shows the results of the FACET-II experimental run. Following the 10 initial random samples, BAX converges to an optimal configuration after 80 queries, recovering a similar emittance (4.49 ± 0.11 mm-mrad, with a match of 1.07 in X-plane and 1.06 in Y) to the best hand-tuned value found during normal operations that day (4.48 ± 0.09 mm-mrad, with a match of 1.05 in X and 1.02 in Y). Similarly to previous figures, we also show the error on the BAX beam-size prediction compared to the true beam size measured on the machine at every iteration.

VI. CONCLUSIONS AND OUTLOOK

In summary, we have proposed a new approach for highly-efficient optimization of systems involving multi-point queries, and have implemented the method experimentally for the high-impact case of emittance optimization in accelerators. Rather than optimizing on slow multi-point queries, BAX instead learns a model of the system on-the-fly from single measurements in the joint control-measurement domain. BAX then guides the optimization by calculating a virtual objective on the fast-executing surrogate. BAX avoids the need for the full, slow multi-point query while also maximizing the information gain of each measurement, increasing the overall efficiency of the optimization.

We applied BAX to the specific task of electron-beam emittance optimization in the LCLS injector in both noisy simulation and experimental settings, and in the FACET-II injector in an experimental setting. In a simulation with typical LCLS noise levels, we saw a 20× increase in efficiency in reaching the optimum when using BAX compared to BO. In experiments on the live machine, BAX was able to reach an emittance that was 24% lower than that achieved by hand-tuning at LCLS, and recovered a similar emittance to the best hand-tuned emittance in FACET-II.

Future work will include expanding the BAX method presented here to higher dimensions to incorporate more quadrupoles and control variables along the accelerator, and targeting more complex objectives (e.g. the beam matching parameter along with the emittance). We expect performance can improve further with stronger priors, such as information on the physical correlations between the control and measurement variables 7. A variety of other accelerator tasks require multi-point queries and could employ BAX. In addition to direct optimization, BAX could be used in tandem with more comprehensive machine models, and highlights a new path toward replacing expensive indirect beam measurements with computation on easy-to-acquire samples from surrogate models.

The BAX approach solves a challenging and long-standing tuning problem in accelerators, and we anticipate it to be broadly applicable to complex optimization problems involving multi-point queries.
commonly found in science and engineering. While we focus on multi-point queries, similar advantages exist for single-point queries that produce rich outputs, e.g. running a simulation. BAX is particularly advantageous over traditional methods in high noise environments and with high dimensional inputs, providing a much needed gain in sample efficiency when optimizing on real-world scientific instruments and devices.

ACKNOWLEDGMENTS

We would like to thank Nicole Neveu for her help in setting up the LCLS emittance calculation, Finn O’Shea for his feedback on the emittance calculation procedures, and Feng Zhou and John Sheppard for their guidance and support with the LCLS injector machine development time.

Appendix A: Experimental Setup

**LCLS Injector Overview.** The LCLS injector begins with a photocathode RF gun followed by two S-band (2.856 GHz) linac sections (L0a and L0b) producing one or more electron bunches at a 120 Hz repetition rate with an energy of 135 MeV (see Fig. 1). In this work, the electron bunch charge was 250 pC. The algorithm and analysis software used to calculate beam profiles was created using experimental devices using PyEpics [39], a Python interface to the EPICS Channel Access (CA) library for the EPICS control system.

**LCLS RF Gun and Measurement Quadrupole.** The focusing gun solenoid (SOL1) is located directly downstream of the RF gun and upstream of L0a, and has two small quadrupole coils incorporated into its magnet. These coils, one normal (CQ1) and one skew (SQ1) quadrupole, are single wires spanning the length of the solenoid, and are intended to cancel a small quadrupole field error at the end of SOL1. The measurement quadrupole (Q5) is 11.7 m from the gun solenoid, and is separated from the wire scanner by a 2.1-m long drift. Figure 1 shows an illustration of this layout.

**LCLS Beam Diagnostics.** The transverse beam profiles at LCLS were measured with minimally-invasive wire scanners for both $X$- and $Y$-planes. They require multiple shots for each measurement, and provide a multipulse averaged integrated beam profile from photomultiplier tube measurements where a Gaussian distribution was used to extract the RMS beam size from the integrated profiles. The statistical fitting error of 3% on average was propagated to the emittance result. The systematic error from the choice of beam size distribution, not taken into account in the experimental emittance results, was estimated to be around 4%/7% in $X$-/$Y$-plane. Each beam-size measurement took at least 18 seconds to execute a wire scanner measurement. While optical transition radiation (OTR) screens would be faster than wire scanners, only the slow wire measurements were available at LCLS at the time of this work.

**FACET-II Setup.** The FACET-II injector produces a single electron bunch at a 120 Hz repetition rate with an energy of 125 MeV. In this setup, the measurement quadrupole is 2.7 m from the measurement screen, separated only by a drift. The beam profiles were measured with an OTR screen. For each measurement at each quadrupole setting, after a 3 second wait-time for the magnet to settle, four images were acquired, background subtracted, then averaged. The $X$- and $Y$-plane beam profile projections of the final image were fitted with a Gaussian distribution whose width was taken as the beam size. The statistical error from the fit (around 3% on average) was propagated to the emittance result.

Appendix B: Beam Emittance Calculations

In this work, we only consider the transverse emittance in the $X$- and $Y$-dimensions, which represents the area of the beam in the transverse planes defined by the positions and momenta, $x$ and $p_x$ in the horizontal plane, and $y$ and $p_y$ in the vertical plane (the $X$- and $Y$-axes are defined as orthogonal and the $Z$-axis as parallel to the path of the beam). All mentions of emittance in this article refer to the geometric mean of the normalized transverse emittance, unless otherwise noted.

Common methods to measure electron-beam size in the transverse planes involve either scanning a wire through the beam or passing the beam through a yttrium aluminum garnet (YAG) or OTR screen [40]. An emittance scan is done using either multiple wires/screens at different positions in the accelerator, or, as we used in this work, a single wire/screen while changing an upstream quadrupole magnet [15, 41]. Here, each full emittance scan was performed using the Q5 quadrupole with adaptive scanning methods as outlined in [35]. For each full emittance scan executed on the machine, four initial points are measured as a rough scan along the full quadrupole range to find the approximate waist location, which can change as the upstream injector configuration is varied. Once the waist is found, a finer scan of seven points is measured around the waist, with redundant points skipped during acquisition. During analysis, any data outside of the convex region are removed from the set based on the inflection points, and points are added as needed to recover a symmetric parabola. The process repeats in each of the $X$ and $Y$ dimensions, for a total of 18 beam-size measurements on average.

When accelerator operators tune the LCLS photoinjector by hand, emittance calculations typically involve a scan of at least four measurements over the domain of the scanning quadrupole upstream of the wire. However, in some configurations of the injector, the waist of the beam—a necessary feature to capture—in the transverse $X$- and $Y$-plane does not occur at the same quadrupole strength, thus requiring two quadrupole scans over two different ranges (or one large range). Moreover, to incor-
porate the calculation into automated tuning, the calculation must be robust, requiring additional measurements. To achieve robust and accurate measurements, the quadrupole scan range needs to be large enough to fully capture the waist of the beam along both axes, while still being narrow enough to accurately quantify the minimum of the parabolic curve. This range can vary for different machine settings and is typically adjusted by hand until a good scan is achieved.

Here, the number of measurements per scan executed on the machine (18 on average) was empirically chosen. Using fewer beam-size measurements per scan was insufficiently robust during optimization, leading to erroneous emittance calculations. Using additional measurements per scan resulted in a slightly smaller final emittance, but at the cost of even slower convergence. BAX uses 30 measurements to calculate the virtual objective on the internal model since acquisition time is many orders of magnitude faster than querying the machine.

Appendix C: BAX Algorithm

We can describe joint measurement and optimization problems as the task of inferring a computable property $O_A$ of a potentially noisy black-box function $f$ given an algorithm $A$. BAX presents a general framework for inferring computable properties $O_A$ within $T$ function evaluations, given $A$ and a prior distribution on $f$, $p(f)$, that captures the initial uncertainty about the true function. In this work, $p(f)$ is defined by a Gaussian process (GP) with a radial basis function (RBF) kernel, and we denote $p(f|D_t)$ as the posterior distribution of $f$ given a dataset $D$ of $t-1$ observations. We then use $p(O_A|D_t)$ to denote the induced posterior distribution over the algorithm output $O_A$.

To reduce the number of function evaluations, we use an information-based BAX method, InfoBAX [23, 31, 32], to make targeted queries that maximize the mutual-information between $O_A$ and the next observation $y_t$. The InfoBAX procedure is a sequential algorithm that seeks to maximize the acquisition function, defined here as the expected information gain (EIG) about $O_A$ upon observing $y_t$. The next sampling point is then chosen to be the $x_t$ that maximizes the estimated EIG. We can write EIG as

$$EIG_t(x) = H(O_A|D_t) - \mathbb{E}_{p(y_t|D_t)}[H(O_A|D_t \cup \{(x, y_t)\})],$$

where $H(O_A|D_t)$ is the entropy of $p(O_A|D_t)$, and $p(y_t|D_t)$ is the posterior predictive distribution at $x$ given $D_t$. For details on how EIG($x$) is estimated, we refer the reader to [23].

To compute EIG efficiently, the InfoBAX procedure only executes the algorithm $A$ on function samples $\hat{f}$ drawn from the posterior distribution $p(f|D_t)$, similar to other posterior sampling-based methods for experimental design [12, 43]. In this manner, the true algorithm output is inferred with minimal evaluations of the true function $f$ (e.g. emittance as a function of accelerator tuning settings). Once the next sampling point is selected at each iteration, a single evaluation of the true function is made at that selected point.

Appendix D: LCLS Injector Surrogate Model

The LCLS injector surrogate model used in this work is a neural network (NN) based surrogate model that provides fast, non-invasive predictions of electron-beam properties. The NN was trained on IMPACT-T simulation [44] data using 16 input parameters including the pulse length, laser radius, gun solenoid and quadrupole settings, L0 linac phase, and all matching quadrupoles settings. The NN architecture consists of 17 layers, and the model outputs scalar predictions of the $X$- and $Y$-plane beam sizes. To mimic experimental emittance calculation procedures in a simulation setting, the $X$, $Y$-plane beam sizes were used when performing a beam size measurement, and Gaussian noise $\mathcal{N}(0, \sigma_n)$ was added to each beam size before passing the measurements to the emittance calculation.

Appendix E: Details of Algorithm Comparison in Simulation Setting

Bayesian optimization was implemented using the Bayesian Optimization package [15]. We defined the BO’s surrogate model as a GP constructed with an RBF kernel, and the upper confidence bound (UCB) as the acquisition function with an exploration weight of 2.0 which was maximized to select the next function query. For each run, BO was initialized with three random scans that were uniformly and randomly sampled from the control domain. The simplex algorithm was implemented using the optimization routine available through SciPy [46]. Each simplex optimization was run with an initial guess of the injector configuration that was randomly sampled from the control domain as well.

In some cases when using the NN and querying points far from optimal, the emittance calculation fails to return a number. Such cases were handled by assigning a large number to simplex queries, and the BAX and BO algorithms were set to pass any failed measurements and move to the next optimal point in the acquisition function optimization. Additionally, any emittance calculations with an uncertainty larger than 70% were ignored by the BO and BAX optimizers.
A. Pétetut, S. Hartman, Rokvintar, R. Clarken, B. Martins, D. Allan, T. Birke, P. R. Jemian, N. Claesson, J. Adelman, J. Dwyer, H. Slepicka, G. Brandl, E. Greenberg, D. Vine, and André, “pyepics/pyepics 3.4.0,” (2019).

[40] H. Loos, R. Akre, A. Brachmann, R. Coffee, F. J. Decker, Y. Ding, D. Dowell, S. Edstrom, P. Emma, A. Fisher, J. Frisch, S. Gilevich, G. Hays, P. Hering, Z. Huang, R. Iverson, M. Messerschmidt, A. Miahnahr, S. Moeller, H. D. Nuhn, D. Ratner, and L. /SLAC /LLNL, (2010).

[41] F. Loehl, Measurements of the transverse emittance at the VUV-FEL, Tech. Rep. 1435-8085 (Germany, 2005).

[42] J. M. Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani, Advances in neural information processing systems 27 (2014).

[43] K. Kandasamy, W. Neiswanger, R. Zhang, A. Krishnamurthy, J. Schneider, and B. Poczos, in International Conference on Machine Learning (PMLR, 2019) pp. 3222–3232.

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

[45] F. Nogueira, “Bayesian Optimization: Open source constrained global optimization tool for Python (version 1.1),” (2014).

[46] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, Nature Methods 17, 261 (2020).