

MULTI-OBJECTIVE BAYESIAN OPTIMIZATION AT SLAC MeV-UED *

Fuhao Ji[†], Auralee Edelen, Joel England[‡], Xiaozhe Shen, Sara Miskovich, Ryan Roussel, Stephen Weathersby, Duan Luo, Mianzhen Mo, Patrick Kramer, Christopher Mayes, Xijie Wang, Alexander Reid, Michael Minitti,
SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Abstract

SLAC MeV-UED, part of the LCLS user facility, is a powerful “electron camera” for the study of ultrafast molecular structural dynamics and the coupling of electronic and atomic motions in a variety of material and chemical systems. The growing demand of scientific applications calls for rapid switching between different beamline configurations for delivering electron beams meeting specific user run requirements, necessitating fast online tuning strategies to reduce set up time. Here, we utilize multi-objective Bayesian optimization(MOBO) for fast searching the parameter space efficiently in a serialized manner, and mapping out the Pareto Front which gives the trade-offs between key beam parameters, i.e., spot size, q-resolution, pulse length, pulse charge, etc. Algorithm, model deployment and first test results are presented.

INTRODUCTION

Machine learning(ML) and artificial intelligence(AI) have revolutionized many computational and real world tasks in recent years, from autonomous driving, protein folding prediction [1] to fusion reactor control [2]. Speeding up and aiding online optimizations of complex particle accelerators is one of the key areas where AI/ML can make substantial contributions [4–10]. At MeV-UED [3], the growing demand of scientific applications calls for highly automated and rapid switch between different machine configurations for delivering electron beams meeting specific user run requirements, necessitating fast online tuning strategies to reduce set up time. At mean time, rapid R&D activities are undergoing for enabling future science applications, which pose additional challenges for beam tuning and optimizations.

We utilize AI/ML based techniques for speeding up online beam tuning at MeV-UED. In particular, the recently introduced Multi-Objective Bayesian Optimization(MOBO) scheme [8] was used for searching the parameter space efficiently in a serialized manner, and mapping out the Pareto Front which gives the trade-offs between key beam parameters, i.e., spot size, q-resolution, pulse length, pulse charge, etc. This method uses a set of Gaussian Process(GP) surrogate models, along with a multi-objective acquisition function, to reduce the number of observations needed to converge by at least an order of magnitude over current methods, i. e., Multi-Object Genetic Optimization (MOGA), and is

thus a critical step toward online multi-objective optimization on real accelerator systems.

MULTI-OBJECTIVE BAYESIAN OPTIMIZATION

The goal of multi-objective optimization is to search the input parameter space and minimize(for simplicity, here we discuss the minimization case) the vector objective function $\mathbf{f}(\mathbf{x}) = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_M(\mathbf{x})\}$. Usually, there is no single solution \mathbf{x}^* that simultaneously minimizes all objectives. Instead, objective vectors are compared using Pareto domination: an objective vector $\mathbf{f}(\mathbf{x})$ dominates another vector $\mathbf{f}(\mathbf{x}')$ if $f_m(\mathbf{x}) \leq f_m(\mathbf{x}')$ for all $m = 1, \dots, M$ and there exists at least one n such that $f_n(\mathbf{x}) < f_n(\mathbf{x}')$. The Pareto Front (PF) \mathcal{P} is the set of non-dominated objective vectors which gives the optimal trade-offs between objectives. Thus the goal of a Pareto optimization algorithm is to identify the PF within a pre-specified budget of function evaluations. Hypervolume(HV) is an often used metric to evaluate the quality of a PF, the HV quantifies the hypervolume(area in the 2 objective case) of the set of points dominated by \mathcal{P} intersected with a region of interest in objective space bounded below by a reference point \mathbf{r} (see Figure 1). MOBO attempt to maximize HV during the optimization process. Given a set of N observations: $D_N = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$, each objective is modeled as an independent GP surrogate model:

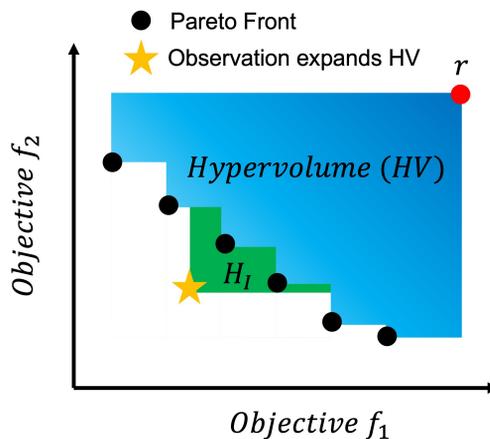


Figure 1: Multi-objective optimization, Pareto Front is formed by non-dominated vectors in the objective space, the optimizer attempts to maximize Hypervolume(HV) during the optimization process.

* Work supported by U.S. Department of Energy Office of Science, Basic Energy Sciences under Contract No. DE-AC02-76SF00515

[†] fuhaoji@slac.stanford.edu

[‡] england@slac.stanford.edu

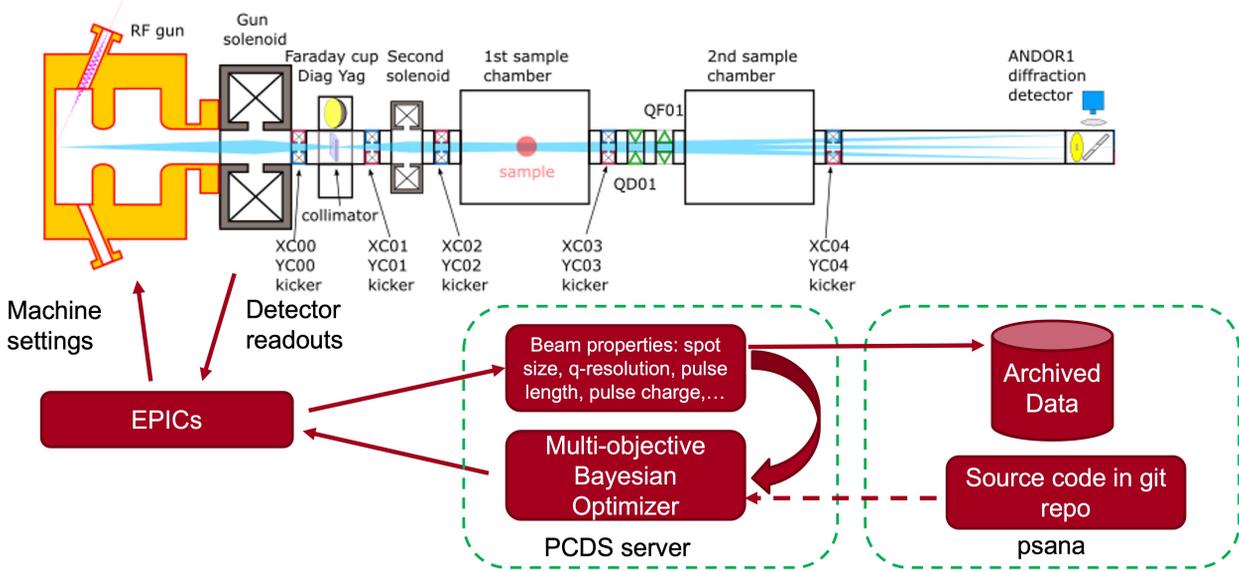


Figure 2: UED beamline layout and MOBO work flow.

$$f_m(\mathbf{x}) \sim \mathcal{GP}_m[\mu_m(\mathbf{x}), k_m(\mathbf{x}, \mathbf{x}')] \quad (1)$$

Where $\mu_m(\mathbf{x})$ is the predicted mean of objective function $f_m(\mathbf{x})$ and $k_m(\mathbf{x}, \mathbf{x}')$ is the covariance function(kernel) based on the objective function behavior(A common kernel is the radial-basis function(RBF) [11]). To proceed with optimization, a scalar acquisition function need to be constructed to find points which are likely to maximally increase the Pareto Frontier HV. Different acquisition functions such as Expected Hypervolume Improvement(EHVI) [12], Parallel Expcted Hypervolume Improvement(qEHVI) [13] and Upper Confidence Boundary Hypervolume Improvement(UCB-HVI) [12] were proposed for application under different conditions. For the application of online optimization at MeV-UED, we choose UCB-HVI for it's simplicity and low computation cost, especially in high-dimensional objective spaces. The UCB-HVI acquisition function describes an optimistic view of the HV improvement:

$$\alpha_{UCB-HVI}(\mu, \sigma, \mathcal{P}, \mathbf{r}, \beta) := H_I(\mathcal{P}, \mu - \sqrt{\beta}\sigma, \mathbf{r}) \quad (2)$$

Where the predicted mean μ and uncertainty σ in objective space is weighted by a hyperparameter β . When $\beta \ll 1$ the acquisition function prioritizes exploitation and when $\beta \gg 1$ it prioritizes exploration. Combining the GP surrogate model and acquisition function, MOBO optimization can be performed. For each optimization step, the algorithm updates the GP surrogate model and decide which point to measure next based on output of the acquisition function, by this way the PF can be mapped out in a serialized manner.

IMPLEMENTATION AT SLAC MEV-UED

The MOBO optimizer was deployed on SLAC PCDS server(Figure 2) capable of interfacing UED EPICs system

for beamline control, fast data acquisition, as well as performing online data processing/surrogate modeling and archiving data on the LCLS data system(psana). Beam spot size at sample plane was measured using a YAG screen coupled to a Questar Microscope with spatial resolution better than 10 μm , q-resolution(equivalent to beam spot size at diffraction detector) was measured using a P43 phosphor screen coupled to Andor EMCCD. Pulse charge was obtained using integrated beam intensity with calibration based on Faraday cup measurement. Electron pulse length was measured using a high field enhancement THz streaking structure [14, 15] developed at SLAC.

So far, most studies on AI/ML for accelerators have focused on proof-of-principle tests of methods under narrow operating conditions, i. e., sampling in small range, low dimensional parameter space or optimizing a single object [7, 10]. This is mainly due to the difficultness of taking sufficient amount of data in multi-dimensional parameter/objective space(usually $10^2 - 10^3$ data points needed and for each data point, multiple beam property measurements,

Table 1: Beamline Parameters Varied During MOBO Optimization

Parameter	Range	Unit
Initial pulse charge	10 - 100	fC
Gun amplitude	70 - 90	MV/m
Gun phase	40 - 60	degree
1 st collimator diameter	100 or 200	μm
Gun solenoid strength	0.1 - 0.25	T
2 nd solenoid strength	0.1 - 0.3	T
Steering magnets	varies with E_k and Sol settings	-

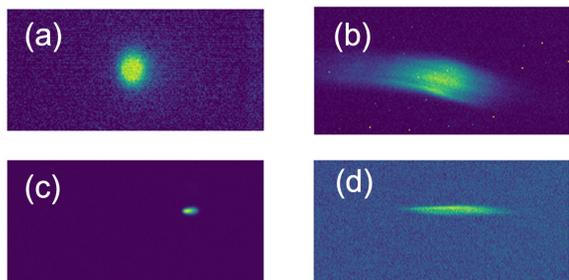


Figure 3: Serialized beam property measurements. (a) beam profile at sample plane (b) virtual cathode (VCC) image (c) beam profile on diffraction detector (d) beam streaked in horizontal using a high field enhancement THz structure, horizontal width corresponds to pulse length.

such as pulse length, spot size, emittance measurements need to be performed). Other practical challenges further complicate the process, for example, beam steering always need to be re-optimized when changing RF cavity amplitudes/phase or tuning focusing elements, also it is often time consuming to change certain variables such as gun amplitude(must do it adaptively to follow cavity temperature change), UV spot size/ pulse length, etc.

To resolve the above issues, a specific data acquisition(DAQ) program was designed and integrated with MOBO. The DAQ automatically switches between different beam detectors and the THz timing structure so that measurements of spot sizes and pulse length at different beamline locations can be quickly performed in a serialized manner(Fig. 3). System parameters were archived simultaneously for surrogate modeling. In addition, manual aid to steering/timing was included. When it needed, the operator can manually tune the steering coils to kick the beam through the collimator and THz structure, as well as tune the THz delay stage to match time of arrival of e-beam and Thz pulse and place the beam at zero-crossing to perform temporal measurements. At mean time, the recorded steering coil and delay stage settings were used to train the surrogate model for predicting correct settings in the following measurements. By using these methods, $10^2 - 10^3$ high quality data points can be obtained during a several hours run.

RESULTS

MOBO optimization was performed at SLAC MeV-UED beamline. Beamline parameters including gun amplitude/phase, solenoid strengths, initial pulse charge, collimator diameter and steering magnets, etc., were varied to produce electron beam with various pulse length, spot size, q-resolution and pulse charges for the use in different ultrafast electron diffraction experiments. Table 1 illustrates the beamline parameters varied during MOBO optimizations. For each run, at least 2 objectives, e. g., spot size vs q-resolution, or pulse length vs q-resolution were assigned, the optimizer first acquire 10 randomly sampled points in the parameter space and built the GP surrogate model, then run

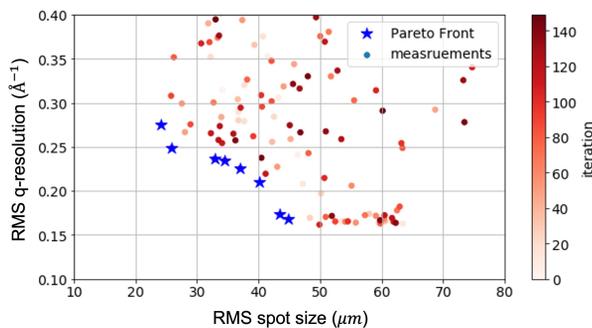


Figure 4: q-resolution vs spot size optimizations, Pareto Front highlighted.

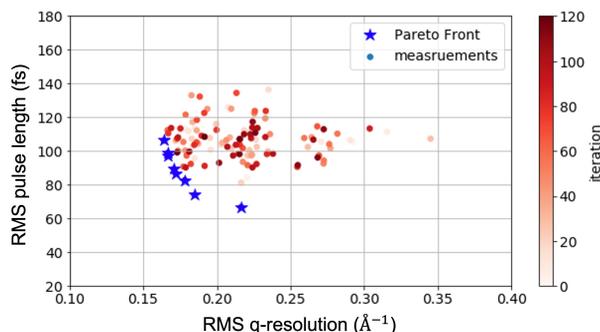


Figure 5: Pulse length vs q-resolution optimizations, Pareto Front highlighted.

MOBO to explore the parameter space and obtain the hypervolume and Pareto Front in objective space. For spot size vs q-resolution optimizations, in total > 2100 data points were taken with various initial pulse charge, beam energy and collimator settings. For pulse length vs q-resolution/spot size optimizations, > 160 data points in total were taken.

Figure 4 illustrates the spot size vs q-resolution optimization result with initial pulse charge 10 fC and 200 μm collimator inserted, complete hypervolume(red) and Pareto Front(blue) was obtained within 140 measurements. Pareto Front shows clear trade-off between spot size at sample plane and rms q-resolution on the diffraction detector, the smallest spot size was down to 23 μm rms with rms q-resolution up to 0.275 \AA^{-1} (left most data point on the Pareto Front). q-resolution gets better with larger spot size and levels off while spot size > 45 μm rms, the corresponding smallest rms q-resolution was 0.165 \AA^{-1} .

Figure 5 illustrates the rms pulse length vs q-resolution optimization result with initial pulse charge 10 fC and 200 μm collimator inserted, complete hypervolume(red) and Pareto Front(blue) was obtained within 120 measurements. Similarly, Pareto Front shows clear trade-off between temporal and q-resolution, in the case of best q-resolution (0.165 \AA^{-1}), rms pulse length was up to 109 fs. Pulse length gets smaller with larger q-resolution, and was found to be down to 66 fs rms with q-resolution = 0.22 \AA^{-1} .

CONCLUSION

AI/ML based beam optimization was implemented at SLAC MeV-UED. Particularly, a MOBO optimizer was deployed on SLAC PCDS server, in combination with a specifically designed DAQ, the optimizer was capable of performing serialized beam property measurements and obtain the Pareto Front which gives trade-offs between key beam parameters, i. e., rms pulse length vs q-resolution, q-resolution vs spot size, etc. Further developments include adding beam parameter constraints [8] using Xopt [16], physics informed optimizations [9], as well as adding emittance vs pulse length optimizations which will be particularly beneficial for improving the UED machine performance.

ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. M.M. acknowledges the support from the U.S. DOE Office of Science, Laboratory Directed Research and Development program at SLAC under contract DE-AC02-76SF00515, and Fusion Energy Sciences under FWP 100182

REFERENCES

- [1] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold”, *Nature*, vol. 596, pp. 583-589, 2021.
- [2] J. Degraeve *et al.*, “Magnetic control of tokamak plasmas through deep reinforcement learning”, *Nature*, vol. 602, pp. 414-419, 2022.
- [3] SLAC Megaelectronvolt Ultrafast Electron Diffraction Instrument <https://lcls.slac.stanford.edu/instruments/mev-ued>
- [4] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelman, “Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems,” *Phys. Rev. Accel. Beams*, vol. 23, p. 044601, 2020.
- [5] A. Sheinker *et al.*, “Demonstration of Model-Independent Control of the Longitudinal Phase Space of Electron Beams in the Linac-Coherent Light Source with Femtosecond Resolution”, *Phys. Rev. Lett.*, vol. 121, p. 044801, 2018.
- [6] S. C. Leemann *et al.*, “Demonstration of Machine Learning-Based Model-Independent Stabilization of Source Properties in Synchrotron Light Sources,” *Phys. Rev. Lett.*, vol. 123, p. 194801, 2019.
- [7] J. Duris *et al.*, “Bayesian Optimization of a Free-Electron Laser,” *Phys. Rev. Lett.*, vol. 124, p. 124801, 2020.
- [8] R. Roussel *et al.*, “Multiobjective Bayesian optimization for online accelerator tuning,” *Phys. Rev. Accel. Beams*, vol. 24, p. 062801. 2021.
- [9] A. Hanuka *et al.*, “Physics model-informed Gaussian process for online optimization of particle accelerators,” *Phys. Rev. Accel. Beams*, vol. 24, p. 072802, 2021.
- [10] R. Roussel *et al.*, “Turn-key constrained parameter space exploration for particle accelerators using Bayesian active learning,” *Nature Communications*, vol. 12, p. 5621, 2021.
- [11] C. E. Rasmussen and K. I. Williams, “Gaussian Processes for Machine Learning”, *Adaptive Computation and Machine Learning*, MIT Press, Cambridge, MA, 2006.
- [12] M. Emmerich, K. Yang, A. Deutz, H. Wang and C. M. Fonseca, “A multi-criteria generalization of Bayesian global optimization”, *Advances in Stochastic and Deterministic Global Optimization, Springer Optimization and Its Applications*, pp. 229-242.
- [13] S. Daulton, M. Balandat, and E. Bakshy, “Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization”, *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [14] E. C. Snively, M. A. K Othman *et al.*, “Femtosecond compression dynamics and timing jitter suppression in a THz-driven electron bunch compressor”, *Phys. Rev. Lett.*, vol. 124, p. 054801, 2020.
- [15] R. Li *et al.*, “Terahertz-based subfemtosecond metrology of relativistic electron beams”, *Phys. Rev. Accel. Beams*, vol. 22, p. 012803, 2019.
- [16] C. E. Mayes *et al.*, “Lightsource Unified Modeling Environment (LUME), a Start-to-End Simulation Ecosystem”, in *Proc. IPAC’21*, Campinas, Brazil, May 2021, pp. 4212–4215. doi:10.18429/JACoW-IPAC2021-THPAB217