

ADAPTIVE MACHINE LEARNING AND AUTOMATIC TUNING OF INTENSE ELECTRON BUNCHES IN PARTICLE ACCELERATORS*

A. Scheinker[†], Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract

Machine learning and in particular neural networks, have been around for a very long time. In recent years, thanks to growth in computing power, neural networks have reshaped many fields of research, including self driving cars, computers playing complex video games, image identification, and even particle accelerators. In this tutorial, I will first present an introduction to machine learning for beginners and will also touch on a few aspects of adaptive control theory. I will then introduce some problems in particle accelerators and present how they have been approached utilizing machine learning techniques as well as adaptive machine learning approaches, for automatically tuning extremely short and high intensity electron bunches in free electron lasers.

Introduction

Precise control of bunch lengths, current profiles, and energy spreads of increasingly shorter electron beams at femtosecond resolution is extremely important for all advanced particle accelerators, including free electron lasers (FEL). FEL X-ray bursts with tunable wavelength are generated by tuning the energies of extremely short electron bunches (~fs). Two of the most advanced FELs are the Linac Coherent Light Source (LCLS) and the European XFEL (EuXFEL). The LCLS provides users with photon energies of 0.27 keV to 12 keV based on electron bunches with energies of 2.5 GeV to 17 GeV with electron bunch charges ranging from 20 pC to 300 pC and the bunch duration from 3 fs to 500 fs [1–3]. The EuXFEL, utilizes electron bunches with energies of up to 17.5 GeV, with charges ranging from 0.02 to 1 nC per bunch, and photon energies of 0.26 keV up to 25 keV [4]. Both the LCLS and the EuXFEL face challenges in quickly tuning between different beam types and achieving precise control for desired current and energy profiles and complex experiments such as two color mode and self seeding [5–8].

Machine Learning

Recently, powerful machine learning (ML) techniques have been studied for various particle accelerator applications. ML-based tools, such as neural networks (NN), can be trained to automatically tune and control large complex systems such as particle accelerators [9–12]. In a preliminary simulation study for a compact THz FEL, a NN control policy was trained to provide suggested machine settings to switch between desired electron beam energies while preserving the match into the undulator and a fast surrogate model was also trained from PARMELA simulation results in order to facilitate the training of the control policy [13].

For mapping inputs to outputs of an analytically unknown, but sampled system, a standard linear regression approach assumes a noise corrupted linear model of the form

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{w}, \quad y = f(\mathbf{x}) + \epsilon, \quad \epsilon \sim N(0, \sigma_n^2), \quad (1)$$

where ϵ is a identically distributed Gaussian distribution with zero mean and variance σ_n^2 . The goal here is to determine an approximation of the weights, \mathbf{w} , in order to learn the mapping $\mathbf{x} \rightarrow y$. Given a collection of measurements, $M = (\mathbf{X}, \mathbf{y})$, where the matrix \mathbf{X} has rows given by m sets of input parameters $\mathbf{x}_i = (x_{i1}, \dots, x_{in})$, $i \in [1, m]$, and the vector $\mathbf{y} = (y_1, \dots, y_m)$ is a collection of outputs, a Bayesian approach gives the following approximation for the weights \mathbf{w} , based on the assumption that they are mean 0 with covariance matrix Σ :

$$\hat{\mathbf{w}} = \sigma_n^{-2} \left(\sigma_n^{-2} \mathbf{X} \mathbf{X}^T + \Sigma^{-1} \right)^{-1} \mathbf{X} \mathbf{y}. \quad (2)$$

This approach works extremely well and is the least squares-based optimal solution for (1) given a set of measurements X , but fails once nonlinearities are introduced in the mapping $f(\mathbf{x})$. The most straight forward way to extend this approach to nonlinear systems is to choose a set of functions, such as polynomials, project an input x into a higher dimensional space, of the form $\mathbf{g}(x) = (1, x, x^2, \dots)$, and then perform a similar approach as above on an assumed model of the form $f(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T \mathbf{w}$. Another approach is to work directly in function space, utilizing Gaussian processes, which are collections of random variables with joint Gaussian distributions, with mean $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ and covariance

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \quad (3)$$

The choice of covariance function determines the shapes of response functions and their smoothness. A typical choice for a smooth covariance function is given by an exponential:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-|\mathbf{x} - \mathbf{x}'|^2 / 2\right), \quad (4)$$

which corresponds to a Bayesian linear regression model with an infinite number of basis functions. A thorough overview of Gaussian processes is available in [14].

Neural networks are another class of extremely powerful ML tools for learning input-output relationships for complex, many parameter systems. In particular, convolutional neural networks (CNN) are very useful for images, for example to map 2D LPS measurements to accelerator component values, because they take into account spacial relationships. Mathematically, a convolutional layer can be written as

$$h_{(i,j),c}^l = \sum_{m=-s}^s \sum_{n=-s}^s \sum_{c'} w_{(m,n),c,c'} h_{(i-m,j-n),c}^{l-1} + b_{c,c'} \quad (5)$$

* Work supported by Los Alamos National Laboratory
† ascheink@lanl.gov

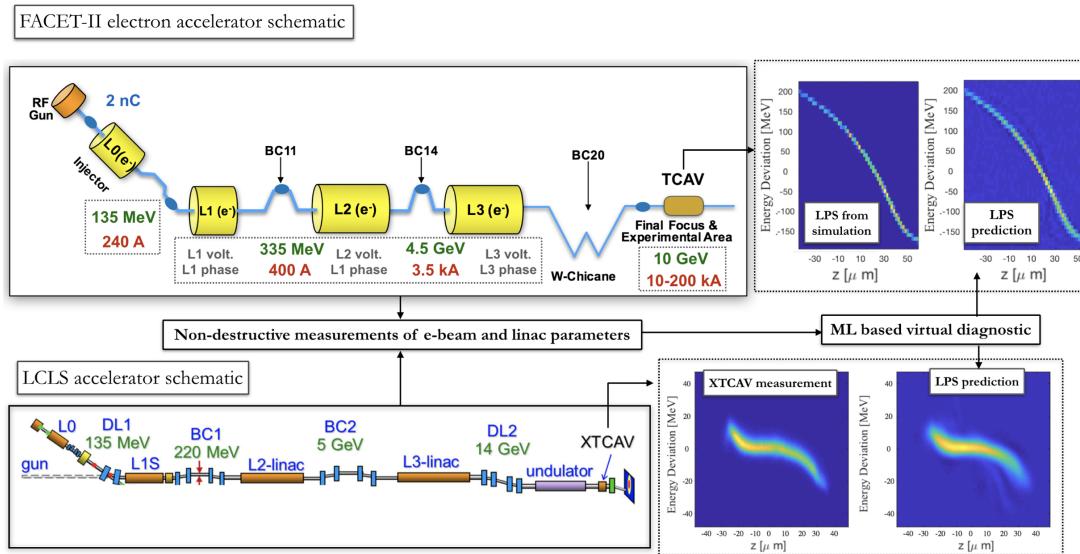


Figure 1: Schematic of the FACET-II and LCLS accelerators and examples of LPS virtual diagnostics compared to LCLS measurements and FACET-II particle tracking simulations. Figure from [16].

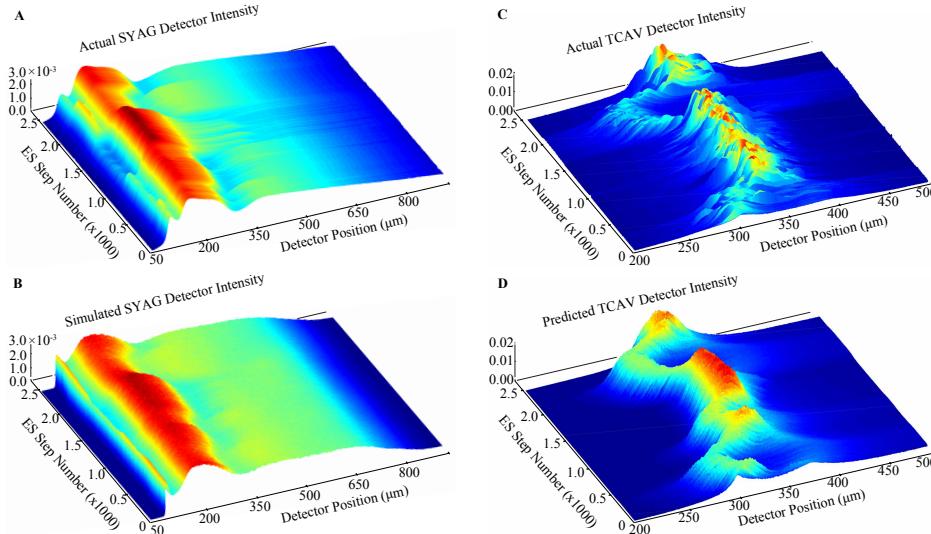


Figure 2: As the accelerator setup was varied, the adaptive scheme continuously adjusted the model to track the measured energy spread spectrum (A) with the simulated energy spread spectrum (B) and thereby was able to predict the measured longitudinal phase space of the beam (C) with simulated XTCAV measurements (D). Images from [25].

which is followed by ab activation function such as

$$ReLU \left(h_{(i,j),c}^l \right) = \max \left\{ 0, h_{(i,j),c}^l \right\}, \quad (6)$$

where $h_{(i,j),c}^l$ is the feature map intensity at (i, j) pixel coordinates, s is the stride, c is the channel index in layer l . All intensities outside the region of the feature map are set to 0. The $w_{(m,n),c',c}$ are the convolutional kernel weights between channel c and channel c' of neighboring layers and $b_{c,c'}$ is a bias term. Powerful CNNs are deep and have very wide layers, resulting in very large numbers of weights and biases that must be tuned by some sort of gradient descent method based on very large collections of training data.

Transverse deflecting cavities provide some of the most important diagnostics that exist, measuring a beam's LPS [15]. Recently, a novel ML approach has been developed to train neural networks to predict a beam's LPS based on accelerator settings [16], as shown in Fig. 1. A novel Bayesian optimization framework that uses sparse online Gaussian processes has been applied for quadrupole magnet tuning in an FEL [17]. Bayesian optimization methods have also been developed for maximizing FEL pulse energy [18]. Various ML tools, including clustering for identifying faulty beam position monitors (BPM) using outlier detection and ML methods for optics corrections has been developed and performed at CERN [19, 20]. For more examples and details the reader is referred to [11, 12] and the references within.

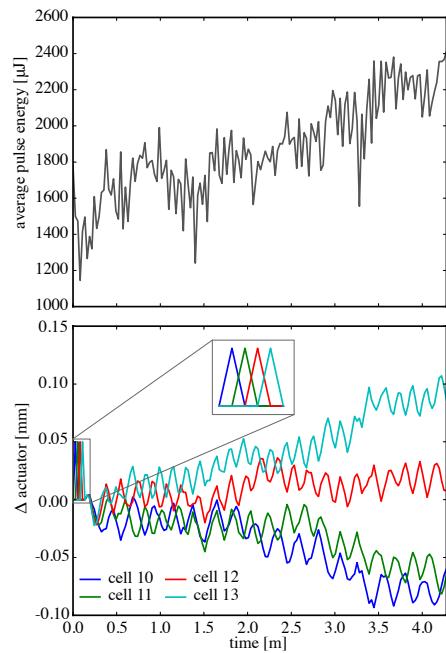


Figure 3: ES scheme applied at the EuXFEL more than doubling average pulse energy over ~4 minutes.

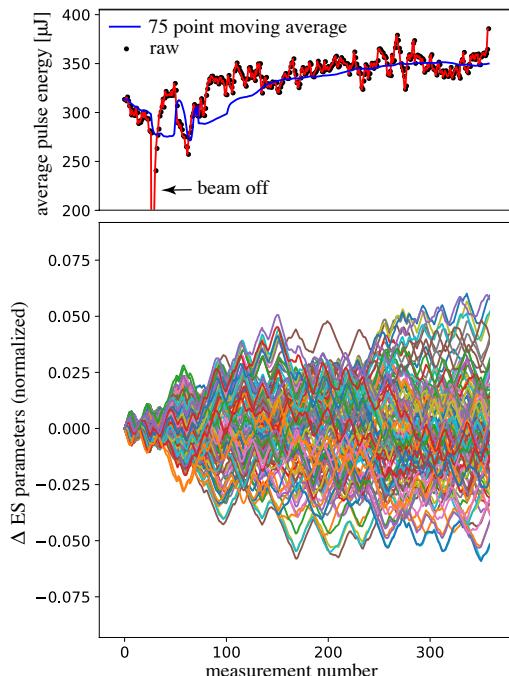


Figure 4: Tuning 105 parameters to maximize average bunch energy based on raw function measurements. The 75 point moving average is plotted to help visualize energy evolution.

Extremum Seeking

The tuning algorithm that we utilized is based on a model-independent adaptive extremum seeking (ES) feedback approach developed for the stabilization of unknown, nonlinear, unstable dynamic systems. The main strengths of the method are that it works based on noisy measurements, can handle

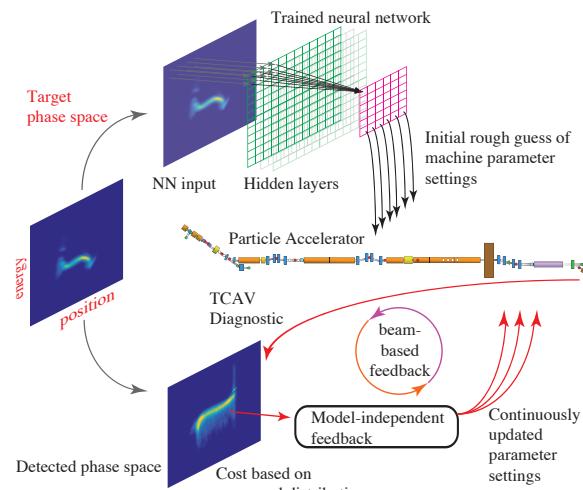


Figure 5: Adaptive ML setup for ILPS tuning at LCLS.

nonlinear, time-varying systems, and can tune many parameters simultaneously. Analytic proofs of convergence for a wide range of systems can be found in the literature [21–25].

For iterative accelerator tuning applications, we consider some analytically unknown cost function that we would like to minimize or maximize base only on noisy measurements, $C(\mathbf{p}, t)$. For this work, $C(\mathbf{p}, t)$ is the pulse energy of the light generated by an FEL and we would like to automatically maximize this cost function. This cost is a function of accelerator parameters $\mathbf{p} = (p_1, \dots, p_m)$, such as magnet power supply settings which control magnetic field strengths or RF system phase and amplitude settings, which control the acceleration of the charged particle beams. Furthermore, all of these components, the beam itself, and the diagnostics drift with time due to external influences such as temperature variation, and therefore there is a time dependence. Also, we are usually only able to sample a noise-corrupted version of such a cost, of the form $\hat{C}(\mathbf{p}, t) = C(\mathbf{p}, t) + n(t)$. Although the interaction of charged particles with external sources of electromagnetic fields, including RF cavities, magnets, and other particles in the bunch, is analytically described via Maxwell's equations and special relativity, when considering a realistic electron bunch and its travel down the length of a particle accelerator, there is no analytic formula relating all component settings to the light pulse energy.

Tuning of parameters \mathbf{p} is based on the dynamics:

$$\frac{dp_i}{dt} = \sqrt{\alpha\omega_i} \cos(\omega_i t + k\hat{C}(\mathbf{p}, t)), \quad (7)$$

where all of the frequencies are distinct, $\omega_i = \omega_{ri} \neq \omega_{rj} = \omega_j$, a good way to choose the dithering frequencies ω_j is to evenly space them in the range $[\omega, 1.75\omega]$, for large ω , so that no two dithering frequencies are integer multiples of each other. α is related to the dithering amplitude of each parameter, upon reaching equilibrium, each parameter oscillates with an amplitude of $\sqrt{\alpha/\omega_j}$ about a steady state value, and k is a gain. Based on [21–25], one can prove that

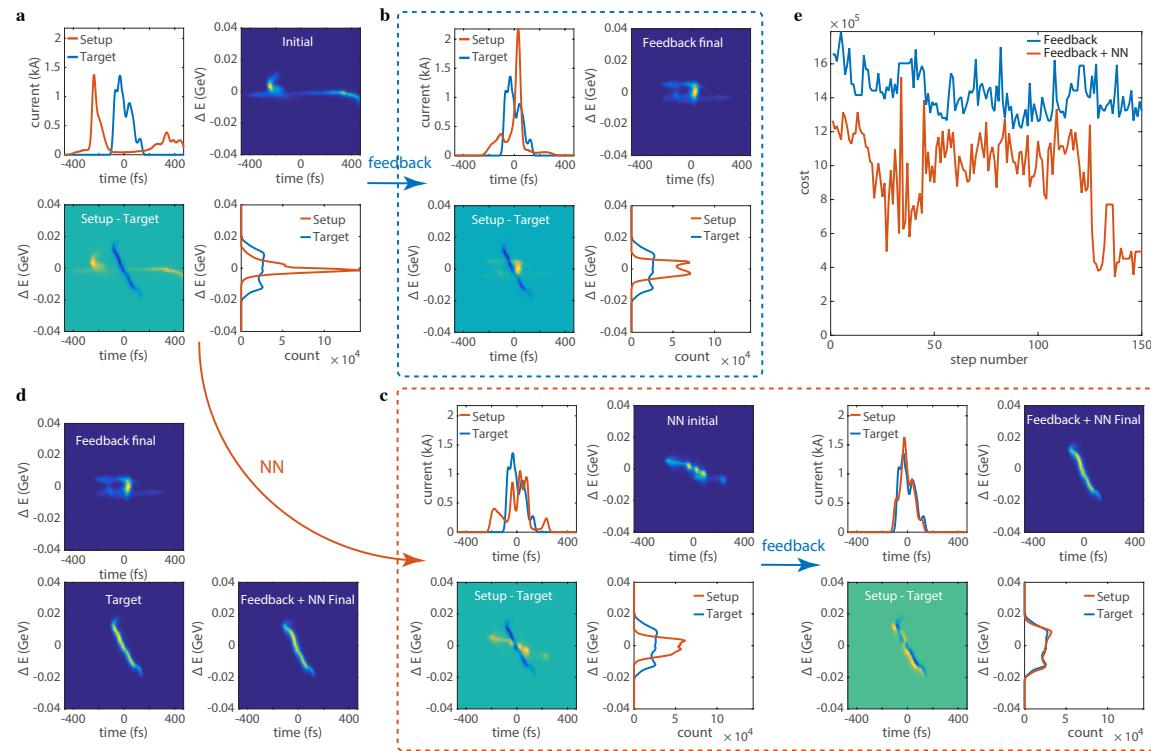


Figure 6: **a:** Longitudinal phase space of initial accelerator setup and target phase space (arbitrary color scales). **b:** The parameters started very far away from their optimal values, feedback alone did not converge within 150 steps, likely stuck in a local minimum. **c:** Utilizing the trained NN to give a closer initial guess, the feedback algorithm was able to converge to the desired phase space within 150 steps. **d:** Final phase space distributions. **e:** Cost function evolution for both case. [27]

on average, for large ω_i , the dynamics of (7) are

$$\frac{dp_i}{dt} = -\frac{k\alpha}{2} \frac{\partial C(\mathbf{p}, t)}{\partial p_i}, \quad (8)$$

a gradient descent of the analytically unknown function C , despite only seeing its noisy measurement \hat{C} .

For digital iterative parameter updates, a finite difference approximation of the derivative in (7) is utilized:

$$p_i(n+1) = p_i(n) + \Delta_t \sqrt{\alpha \omega_i} \cos \left(\omega_i n \Delta_t + k \hat{C}(n) \right), \quad (9)$$

where Δ_t is chosen such that $\Delta_t < \frac{2\pi}{5 \max \omega_i} \ll 1$, so that the finite difference approximation of the derivative holds.

The physical parameter update period, T_w , and the digital algorithm's numerical time step, Δ_t , are two completely independent quantities. In the digital algorithm, Δ_t is chosen to be arbitrarily small, based on dithering frequency choices, as described above. The update time, T_w , is the physical time between parameter updates and is chosen based on how fast accelerator parameters can be adjusted.

The iterative scheme is applied as follows: 1). Initial parameter settings, $\mathbf{p}(1)$, are set. 2). A wait time T_w is allowed to pass for accelerator components to settle to their set points, $\mathbf{p}(1)$. This may be ~ 1 second for slow mechanical systems such as phase shifters and ~ 0.1 seconds for digital RF amplitude or phase set-points, then record the cost function, $\hat{C}(1)$.

3). Calculate new parameter settings, $\mathbf{p}(2)$, based on $\mathbf{p}(1)$ and $\hat{C}(1)$, according to (9) and continue iteratively.

The ES scheme has been applied at FACET to create a non-invasive longitudinal phase space diagnostic, by adaptively tuning a model to match a non-destructive energy spread spectrum. Once this match was accomplished, the model's accurately predicted and tracked the longitudinal phase space (LPS) of the electron beam, as shown in Fig. 2 [25]. Further work in this direction is ongoing for even more accurate LPS predictions at FACET-II. We utilized the ES scheme for automatically maximizing the average pulse energy of both the LCLS and the EuXFEL FELs [26]. Figure 3 shows the technique being applied at the EuXFEL more than doubling average pulse energy over ~ 4 minutes and Fig. 4 shows the results of applying the same technique at the EuXFEL with 105 parameters (84 air coils and 21 phase shifters) and a noisy cost function without averaging. This was during initial machine setup in which various parts are incrementally tuned to establish SASE.

Adaptive Machine Learning

Whereas a model-independent method, such as ES, can handle time-varying systems, it is a local approach and can possibly get stuck in local minima. Trained NNs can tune globally, but only for the data sets they were trained on, and therefore cannot handle time-varying systems. Therefore, we created an adaptive ML framework in which a trained

NN takes a first global guess and then adaptive feedback is turned on and zooms in on and track time-varying optimal parameters, as shown in Fig. 5. The approach was to train an NN based on a parameter scan, where for each parameter setting of the LCLS, we recorded a TCAV image of the LPS, to learn how to map phase spaces to parameters [27]. In Fig. 6 we demonstrate the ability of the adaptive machine learning approach. To achieve a desired phase space, a first guess for machine parameters via a train NN takes place (a), ES is then applied ES based on real time TCAV measurements where the cost is the difference between the desired and current 2D phase space images (b), resulting in convergence (c).

Conclusions

Advanced adaptive feedback, machine learning, and adaptive machine learning tools are being developed for automatic accelerator tuning, optimization, and for the development of non-invasive diagnostics based on combinations of real-time measurements and fast online models.

REFERENCES

- [1] Y. Ding *et al.*, "Measurements and simulations of ultralow emittance and ultrashort electron beams in the linac coherent light source," *Phys. Rev. Lett.*, vol. 102, p. 254801, 2009.
- [2] P. Emma *et al.*, "First lasing and operation of an angstrom-wavelength free-electron laser," *Nature Photonics*, vol. 4, pp. 641–647, 2010.
- [3] D. Ratner *et al.*, "Experimental demonstration of a soft x-ray self-seeded free-electron laser," *Physical Review Letters*, vol. 114, p. 054801, 2015.
- [4] H. Weise and W. Decking, "Commissioning and First Lasing of the European XFEL", in *Proc. FEL'17*, Santa Fe, NM, USA, Aug. 2017, pp. 9–13. doi:10.18429/JACoW-FEL2017-MOC03
- [5] J. Rzepiela *et al.*, "Tuning of the LCLS Linac for user operation," SLAC National Accelerator Lab, Menlo Park, USA, Rep. SLAC-PUB-16643, Jul. 2016.
- [6] T. O. Raubenheimer, "Technical challenges of the LCLS-II CW X-Ray FEL," in *Proc. IPAC'15*, Richmond, VA, USA, 2015. doi:10.18429/JACoW-IPAC2015-WEYC1
- [7] A. A. Lutman, *et al.*, "Fresh-slice multicolour X-ray free-electron lasers," *Nature Photonics*, vol. 10, no. 11, pp. 745–750, 2016.
- [8] A. A. Lutman *et al.*, "High-power femtosecond soft x rays from fresh-slice multistage free-electron lasers," *Phys. Rev. Lett.*, vol. 120 no. 26, p. 264801, 2018.
- [9] M. Buchanan, "Depths of learning," *Nature Physics*, vol. 11, no. 10, pp. 798–798, 2015.
- [10] Y. B. Kong *et al.*, "Predictive ion source control using artificial neural network for RFT-30 cyclotron," *Nucl. Instrum. Methods A*, vol. 806, pp. 55–60, 2016.
- [11] A. L. Edelen *et al.*, "Neural networks for modeling and control of particle accelerators," *IEEE Transactions on Nuclear Science* vol. 63, no. 2, pp. 878–897, 2016.
- [12] A. L. Edelen *et al.*, "Machine Learning to Enable Orders of Magnitude Speedup in Mult-Objective Optimization of Particle Accelerator Systems," *arXiv:1903.07759*, 2019.
- [13] A. L. Edelen *et al.*, "Using a Neural Network Control Policy for Rapid Switching Between Beam Parameters in an FEL," Los Alamos National Lab, Los Alamos, USA, Rep. LA-UR-17-28069, Aug. 2017.
- [14] C. E. Rasmussen & C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [15] C. Behrens *et al.*, "Few-femtosecond time-resolved measurements of X-ray free-electron lasers," *Nat. Commun.*, vol. 5, p. 3762, 2014.
- [16] C. Emma *et al.*, "Machine learning-based longitudinal phase space prediction of particle accelerators," *Phys. Rev. Accel. Beams*, vol. 21, no. 11, p. 112802, 2018.
- [17] M. McIntire, D. Ratner, and S. Ermon, "Sparse Gaussian Processes for Bayesian Optimization," in *Proc. of the 32nd Conference on Uncertainty in Artificial Intelligence*, Jersey City, NJ, USA, Jun. 2016, pp. 517–526.
- [18] J. P. Duris *et al.*, "Bayesian Optimization for Online FEL Tuning at LCLS," presented at HB'18, Daejeon, Korea, Jun. 2018, paper THA2WE01, unpublished.
- [19] E. Fol, J. M. Coello de Portugal, and R. Tomas, "Unsupervised Machine Learning for Detection of Faulty BPMs," in *Proc. IPAC'19*, Melbourne, Australia, 2019. doi:10.18429/JACoW-IPAC2019-WEPGW081
- [20] E. Fol, J. M. Coello de Portugal, and R. Tomas, "Optics Corrections Using Machine Learning in the LHC," in *Proc. IPAC'19*, Melbourne, Australia, 2019. doi:10.18429/JACoW-IPAC2019-THPRB077
- [21] A. Scheinker, "Model Independent Beam Tuning", in *Proc. IPAC'13*, Shanghai, China, May 2013, paper TUPWA068, pp. 1862–1864.
- [22] A. Scheinker and D. Scheinker, "Bounded extremum seeking with discontinuous dithers," *Automatica*, vol. 69, pp. 250–257, 2016.
- [23] A. Scheinker and D. Scheinker, "Constrained extremum seeking stabilization of systems not affine in control," *Int. J. Robust Nonlinear Control*, vol. 28, pp. 568–581, 2018.
- [24] A. Scheinker, "Bounded extremum seeking for angular velocity actuated control of nonholonomic unicycle," *Optimal Control Applications and Methods*, vol. 38, pp. 575–585, 2017.
- [25] A. Scheinker and S. Gessner, "Adaptive method for electron bunch profile prediction," *Phys. Rev. Spec. Top. Accel Beams*, vol. 18, no. 10, p. 102801, 2015.
- [26] A. Scheinker *et al.*, "Model-independent tuning for maximizing free electron laser pulse energy," *Phys. Rev. Spec. Top. Accel Beams*, vol. 22, no. 8, p. 082802, 2019.
- [27] A. Scheinker *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, no. 4, p. 044801, 2018.