Machine Learning Models for Optimization and Control of X-ray Free Electron Lasers

Auralee Edelen,[*] **Nicole Neveu, Claudio Emma, Daniel Ratner, Christopher Mayes**
SLAC National Accelerator Laboratory

Abstract

Particle accelerators are used in a wide array of medical, industrial, and scientific applications. To meet the needs of these applications, the controllable settings of the accelerator are often optimized to produce specific particle beam parameters, both during the initial design of the system and on-the-fly during operation (i.e. "tuning"). For example, at the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory, settings are adjusted for each scientific experiment to create custom electron beam shapes and energies. At present, this process relies heavily on manual tuning by human operators and simplified analytic models that do not fully capture the system behavior. There is growing interest in using machine learning to help model and control accelerators like the LCLS. Here, we discuss ongoing work in using machine learning to create fast-executing, accurate surrogate models that can aid this tuning process.

1   Introduction

Tuning of the LCLS requires adjustment of tens to hundreds of variables depending on the specific setup. As such, using machine learning to aid the task of tuning is a growing area of interest [1, 2]. If one had an accurate, fast-executing model of the system, online tuning could likely be automated to a greater degree. Prior to a given experiment, more extensive optimization of the setup could also be conducted. However, particle accelerators like the LCLS are challenging to model due to their high dimensionality, nonlinear behavior, the presence of small, compounding errors in the descriptions of beamline elements (e.g. magnetic fields of focusing magnets can have asymmetries), and hidden variables. Finally, while physics simulations that include most of the relevant effects do exist, these are computationally intensive and often still cannot easily be made to match the observed behavior of the accelerator. Such simulations are also too computationally intensive to run "online" during machine operation, thus precluding their use in online optimization and control. The computational burden also severely limits exploration of the full parameter space during offline optimization studies for new setups or new designs. In practice, often only a few narrow "working points" are chosen for detailed optimization studies.

As a complementary approach, we are using neural networks (NNs) to learn fast-executing surrogate models from a sparse sampling of simulations of particle accelerator systems. To do this, we conduct a wide random sampling of the controllable variables and predict the resultant beam characteristics, including scalar values that summarize statistics of the beam's 6D phase space (x, y, z positions and momenta of each electron in the beam) and 2D binned projections of the phase space. The reason for using simulations is that the parameter space cannot in practice be fully explored on the machine because of the operational expense. Pre-training on a wide range of simulation data and then fine-tuning with with measured data may provide a way of filling out the parameter space more fully (prior investigation into this approach looks promising [3, 4]). In addition to being used directly
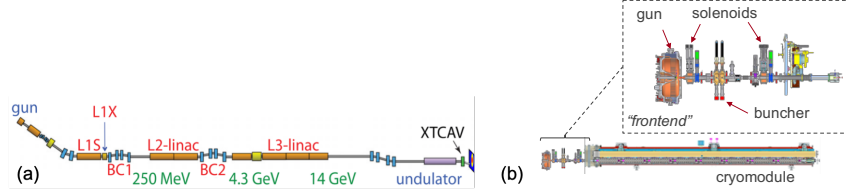
---

[*]`edelen@slac.stanford.edu`

Figure 1: Schematic of the LCLS (a). The electron beam is accelerated using radio frequency (rf) cavities (grouped in linac sections L1, L2, L3), for which the phases and voltages can be adjusted to tune the final electron beam characteristics (e.g. energy, energy spread, shape in longitudinal phase space). The two bunch compressors (BC1 and BC2) are sets of magnets that compress the beam longitudinally. The transverse deflecting cavity (XTCAV) is a diagnostic for the beam's longitudinal phase space (pulse duration vs. energy). In (b), we show a schematic of the LCLS-II injector, with the rf gun, solenoids, and rf bunching cavity used to manipulate the initial electron beam.

in model-based optimization, these surrogate models can be used to help prototype new optimization algorithms with more rapid turnaround than can be achieved with the simulation alone.

In addition to full system models, we are also constructing modular models of subsystems. Accelerators frequently share designs, and thus these modular components could be swapped between different accelerators. For example, both FACET-II at SLAC and SwissFEL at the Paul Scherrer Institut are, at a high level, similar in design to the LCLS. Here, we focus on models of the LCLS electron beamline and LCLS-II injector, both of which are shown in Fig. 1. The LCLS-II is a major upgrade to the LCLS that is nearing final completion, and the LCLS-II injector will eventually provide beam to a new Linac that is part of this upgrade. In terms of NN-based modeling, the main technical difference to previous work [2 - 11] is the prediction of both scalar beam characteristics and images of the phase space, the use of convolutional layers for the image predictions, the wider range of input parameters used, and the particular systems examined.

2   Modeling for the Linac Coherent Light Source

Forward System Modeling   First, we created a NN model of the LCLS electron beamline (shown in Fig. 1a) using data from the simulation code Bmad [12], which includes all of the major nonlinear beam effects up to the entrance of the undulator. We randomly sampled a wide range of the linac phases and voltages (6 variables in total), which are commonly adjusted to manipulate the beam's shape in longitudinal phase space (LPS), i.e. z-position and z-momentum. The data set covers the majority of the operating range of the LCLS, and it was generated by running appx. 40,000 simulations on the Cori supercomputer at NERSC [13]. The NN was trained to predict 25 scalar beam parameters (the beam's size, phase space area, energy, energy spread, etc.) and the LPS image. We find excellent agreement between simulation results and the predictions from the NN, as shown in Fig. 2. The NN architecture is implemented in Keras [14], and it consists of seven densely-connected layers, followed by two parallel sets of layers for the LPS image and scalar predictions: (1) four sets of alternating convolutional and upsampling layers for the LPS image and (2) two densely-connected layers for the scalars. The convolutional layers each have 16 4x4 filters. The mean squared error in the predictions was used as the cost function, and all activation functions of the hidden units are leaky rectified linear units. An L2 penalty and early stopping were both used to help combat overfitting.

We took the same approach for the LCLS-II injector (shown in Fig. 1b). Injectors are typically more difficult to model because nonlinear effects are more important at low energy. Here we scanned the gun phase, two solenoid strengths, and the buncher amplitude and phase (5 variables in total). We found similarly good agreement between the NN predictions and simulation results, as shown in Fig. 3.

For these two cases, the main technical difference to previous work that involves the prediction of beam images [2 - 6] is the wider range of input parameters, the use of convolutional layers for the image prediction, and the combined prediction with scalar beam parameters. For example, in [5] the LPS is predicted over a narrower range of inputs, and in [2 - 6] a densely-connected network is used to produce the image prediction (the vector output of the NN is then re-shaped into an image).
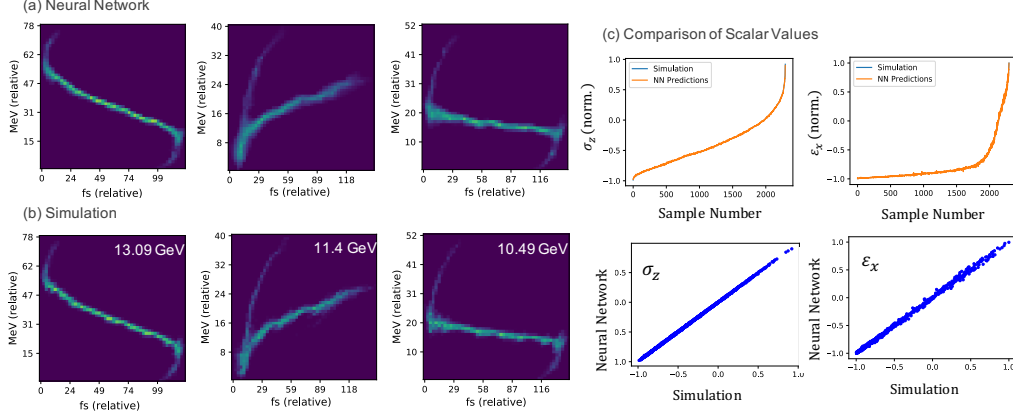
2

Figure 2: Representative example of NN predictions (a) and simulations (b) for the LCLS longitudinal phase space (beam duration vs. energy in z). In (c) we plot NN predictions for the normalized x emittance (ε_x) and z beam size (σ_z) against the simulation results (bottom row). We also show the agreement between sorted values (top row). The scalar predictions are in normalized units $[-1, 1]$, and while only two examples are shown here, the agreement for the other 25 outputs is similar. All examples shown are drawn from the test set.
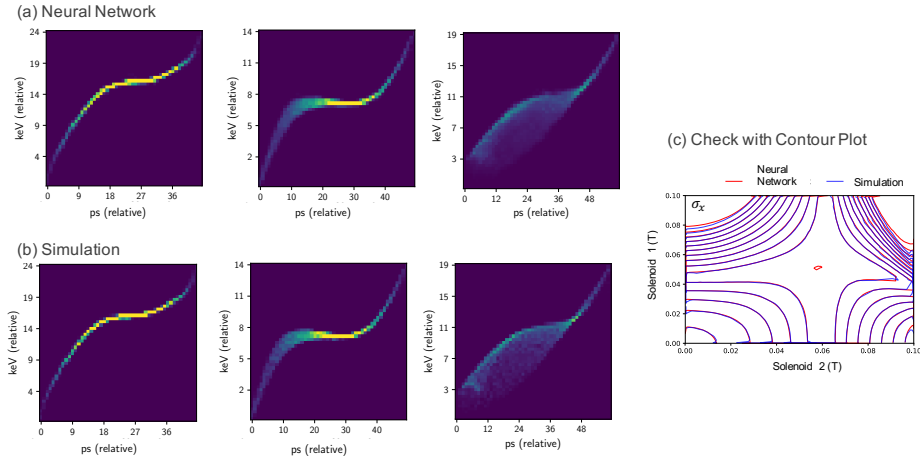

Figure 3: NN predictions (a) and simulations (b) for the LCLS-II injector LPS. All examples are from the test set. In (c) we compare the contour plot of the x beam size (σ_x) that is produced by scanning the two solenoid strengths for the simulation and the NN (i.e. a new scan of the inputs).

Finally, we trained a model of the bunch compressor section BC2. Here, the input is the LPS before entering BC2, and the output is the LPS at the exit of BC2. As shown in Fig. 4, the agreement is good. However, some fine detail is smoothed over in the predictions. This possibly could be remedied by fine-tuning the architecture. In this case, we use an encoder-decoder style network: three convolutional layers (with 10 4×4 filters each) alternated with two 2×2 max pooling layers, followed by four densely-connected layers, followed by three convolutional layers alternated with two upsampling layers.

Inverse System Modeling and Optimization   Ultimately, we would like to use these models in optimization and tuning. Two initial routes for this are: (1) running an optimization algorithm on the surrogate model, and (2) using an inverse model to provide an initial guess of settings to use for particular target beam characteristics. Preliminary work in both areas has shown promise. In [7] accelerator surrogate models are evaluated for use in multi-objective optimization, in [8, 9] an initial
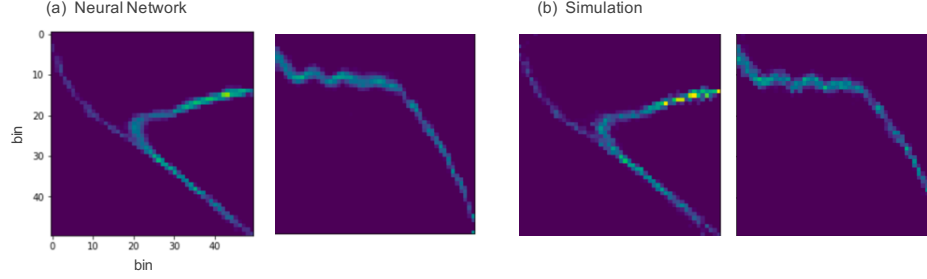
Figure 4: Representative examples of simulations and NN predictions for the bunch compressor BC2, each drawn from the test set. Here we are showing the raw binned values rather than full-scale units (each image consists of 50×50 bins).
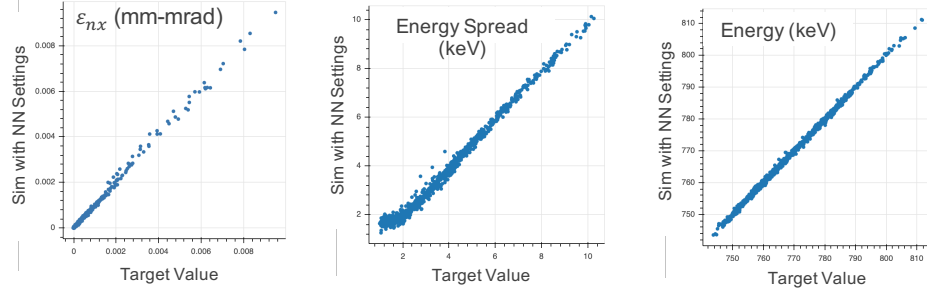

Figure 5: Simulations and NN predictions for the inverse model of the LCLS-II injector. The target values are sent to the NN inverse model, and here we plot these against the values obtained from the simulation when using the suggested settings from the NN. Each beam characteristic request consists of 12 target beam parameters that are relevant to injector optimization, and five injector variables can be adjusted to achieve them (gun phase, two solenoid strengths, and buncher amplitude and phase).

simulation study in model inversion for a compact FEL is examined, and in [10] an inverse NN is used to provide initial settings for two variables at the LCLS given a target LPS projection.

Extending these studies to higher dimension, we trained an inverse NN of the LCLS-II injector by requesting a set of target beam characteristics, sending the resultant settings from the inverse NN to the forward mode, and then backpropagating the difference between the target and predicted beam characteristics through the combined setup (i.e. freezing the forward model weights and training the inverse model by letting it interact with the forward model). This is similar to the approach taken in [8, 9], which is adopted because the system is not directly invertible (the outputs are not unique with respect to the inputs). Finally, we checked the performance of the inverse NN by sending the settings for various beam requests to the simulation and seeing how close the resultant beam characteristics were to the requested values. We find that we are able to obtain the target beam characteristics for a very broad range of requests (see Fig. 5). Each request consists of 12 target beam parameters that are important for injector optimization, and the NN is able to adjust the five available controllable settings to achieve them (gun phase, two solenoid strengths, and buncher amplitude and phase).

3 Discussion and Future Outlook

We have constructed NN models of the LCLS that are $\mathcal{O}(10^6)$ times faster to execute than the original physics simulations (corresponding to under a ms on a laptop for one execution). This is an important step toward creating a set of models to aid machine optimization. To use these models in practice, we plan to include uncertainty estimates in the predictions (e.g. via model ensembling, Bayesian NNs, or MC dropout). The models can then also be incorporated into Bayesian optimization schemes for live tuning of the accelerator that are currently under development (e.g. see [15] and [16] for successful tests that use Gaussian Process models with Bayesian optimization for LCLS and SPEAR3, another accelerator at SLAC). Expanding the work in [3, 4], we plan to update the models with measured

4

data, and we also plan to predict the full set of phase space projections, rather than just the LPS. One concern is how these models will drift in accuracy over time. We plan to address this by updating the models over time (online retraining) and combining them with local optimization to fine-tune settings (e.g. see [9] for an initial demonstration of this latter approach). Finally, although here we focus on the LCLS, these techniques are being developed with both LCLS and FACET-II [17] in mind (e.g. see [5, 6] for applications to the latter accelerator). The design of FACET-II is similar to LCLS, but it operates with different beam parameters and has different operational challenges (e.g. higher jitter).

Acknowledgments

References

[1] A.L. Edelen, C. Mayes, D. Bowring, D. Ratner, A. Adelmann, R. Ischebeck, J. Snuverink, I. Agapov, R. Kammering, J. Edelen, I. Bazarov, G. Valentino, J. Wenninger. "Opportunities in Machine Learning for Particle Accelerators'. November 2018, `https://arxiv.org/abs/1811.03172`

[2] A.L. Edelen, et al. "Machine Learning Demonstrations on Particle Accelerators." *North American Particle Accelerator Conference*, 1-6 September 2019, Lansing, Michigan, THXBA1.

[3] A.L. Edelen, J.P. Edelen, et al. "Neural Network Virtual Diagnostic for the FAST Low Energy Beamline." *9th International Particle Accelerator Conference*, 29 April - 4 May 2018, Vancouver, Canada, WEPAF040.

[4] A.L. Edelen, S. Biedron, D.L. Bowring, B.E. Chase, D.R. Edstrom, J. Steimel, J.P. Edelen, P.J.M. van der Slot. "Neural Network-Based Approaches to the Modeling and Control of Particle Accelerators." *9th International Particle Accelerator Conference*, 29 April - 4 May 2018, Vancouver, Canada, THYGBE2.

[5] C. Emma and A.L. Edelen, et al. "Machine learning based longitudinal phase space prediction for accelerators." *Physical Review Accelerators and Beams*, vol. 21, is. 112802, November 2018.

[6] C. Emma, A.L. Edelen, M. Alverson, A. Hanuka, et al. "Machine Learning-Based Longitudinal Phase Space Prediction of Two-Bunch Operation at FACET-II." IBIC, 8 - 12 September, 2019, Malmo, Sweden. THBO01.

[7] A.L. Edelen, N. Neveu, Y. Huber, M. Frey, C. Mayes, A. Adelmann. "Machine learning to enable orders of magnitude speedup in multi-objective optimization of Particle Accelerator Systems." *Under review.* Preprint at `https://arxiv.org/abs/1903.07759`

[8] A.L. Edelen, J. Edelen, et al. "Using Neural Network Control Policies for Rapid Switching Between Beam Parameters in a Free Electron Laser." *NIPS DLPS Workshop*, 6 September 2017, Long Beach, California.

[9] A.L. Edelen, S.G. Biedron, J.P. Edelen, S.V. Milton, and P.J.M. van der Slot. "Using a neural network control policy for rapid switching between beam parameters in an FEL." *38th International Free Electron Laser Conference*, 20 - 25 August 2017, Santa Fe, New Mexico. `https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-17-28069`

[10] A. Schienker, A.L. Edelen, D. Bohler, C. Emma, A. Lutman. "Demonstration of model-independent control of the longitudinal phase space of electron beams in the Linac-coherent light source with femtosecond resolution." *Physical Review Letters*, vol. 121, is. 044801, July 2018.

[11] A.L. Edelen, S. Biedron, J. Edelen, and S. Milton. "First Steps Toward Incorporating Image Based Diagnostics into Particle Accelerator Control Systems Using Convolutional Neural Networks." *2nd North American Particle Accelerator Conference (NAPAC2016)*. 9 - 16 October, 2016, Chicago, IL. TUPOA51.

[12] D. Sagan. "The Bmad Reference Manual." `https://www.classe.cornell.edu/bmad/`

[13] National Energy Research Scientific Computing Center (NERSC). `https://www.nersc.gov/`

[14] F. Chollet, et al. Keras. GitHub, 2015. `https://github.com/fchollet/keras`

[15] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, A. Egger, T. Cope, and D. Ratner. "Bayesian optimization of a free-electron laser." *Under review.* Preprint at `https://arxiv.org/pdf/1909.05963.pdf`

[16] A. Hanuka, J. Duris, J. Shtalenkova, D. Kennedy, A. Edelen, D. Ratner, X. Huang. "Online tuning and light source control using a physics-informed Gaussian process." *These proceedings.*

[17] V. Yakimenko, et al. "FACET-II facility for advanced accelerator experimental tests." *Phys. Rev. Accel. Beams*, vol. 22, is. 101301, October 2019.