ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOPT070

SURROGATE MODELLING OF THE FLUTE LOW-ENERGY SECTION

C. Xu*, E. Bründermann, A.-S. Müller, J. Schäfer, and A. Santamaria Garcia
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract

Numerical beam dynamics simulations are essential tools in the study and design of particle accelerators, but they can be prohibitively slow for online prediction during operation or for systematic evaluations of new parameter settings. Machine learning-based surrogate models of the accelerator provide much faster predictions of the beam properties and can serve as a virtual diagnostic or to augment data for reinforcement learning training. In this paper, we present the first results on training a surrogate model for the low-energy section at the Ferninfrarot Linac- und Test-Experiment (FLUTE).

INTRODUCTION

Compared to single beam dynamics, collective effects are computationally more expensive to calculate. For example, a detailed space charge particle tracking simulation often takes minutes to run and thus makes parameter optimization or training of machine learning algorithms on simulation data very time consuming or even infeasible. A surrogate model can be used to provide rapid evaluations and replace the time-consuming simulations by approximating the outputs. Common methods to build a surrogate model include Gaussian process regression, random forests, and deep neural networks (NN) [1–3]. In addition, surrogate models could be used as virtual diagnostics, predicting valuable information of the beam in a non-destructive way, e.g. the longitudinal phase space of the electron bunches [4, 5]. In this paper, we present the development of a neural network surrogate model of the low-energy section at the Ferninfrarot Linac- und Test-Experiment (FLUTE). We describe the training process and compare the NN-predicted bunch properties to both simulations and measurement data. Finally, we discuss applications of the surrogate model as an online virtual diagnostic and as a training environment for other algorithms.

TRAINING THE SURROGATE MODEL

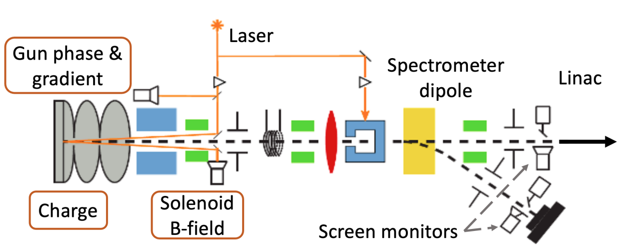

Figure 1: Schematic layout of FLUTE low-energy section with the inputs to the surrogate model marked with an orange box. Figure adapted from [6].

Figure 1 shows the schematic layout of the low-energy section of the the KIT linac-based test facility, FLUTE [7]. The electrons are generated at the RF photoinjector and accelerated up to 7 MeV. The input layer of the surrogate model NN consists of 4 neurons representing the bunch charge, the RF gun phase, the RF gun maximal gradient, and the solenoid magnetic field. The output layer returns 6 scalar values representing the bunch properties, namely the transverse beam size σ_x, bunch length σ_z, mean energy E, relative energy spread σ_E, normalized transverse emittance ϵ_x, and percentage of the remaining particles. These bunch properties can also be measured using the diagnostic devices [8], so that the model can be further retrained and fine-tuned to match the measurement results.

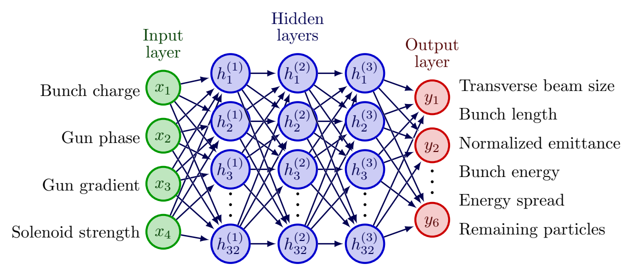

Figure 2: Architecture of the fully connected neural network. The 4 machine parameters are fed into the neural network with 3 hidden layers, each with 32 neurons.

Table 1: Input parameter range used to generate training data

Input Parameters	Range	Unit
Charge	1 to 30	pC
Gun phase	175 to 235	deg
Gun max. gradient	50 to 100	MV/m
Solenoid B-field	0.08 to 0.2	T

We use a fully connected feed-forward NN with 3 hidden layers, as shown in Fig. 2, each layer with 32 neurons and the hyperbolic tangent function (tanh) as the activation function. The size of the network is chosen to sufficiently approximate the transfer map from photocathode to linac, but not too large so that it fully memorizes the dataset. In such a case, the network will not be able to generalize to unseen scenarios. The NN is implemented using the open-source library pyTorch [9]. The training data consists of 10^4 samples randomly selected from the parameter space listed in Table 1. For each parameter setting, the bunch properties are obtained via an ASTRA tracking simulation. The NN training parameters are summarized in Table 2. The input and output parameters are min-max normalized, mapped to [0,1] intervals, to speed up the training process. The output

* chenran.xu@kit.edu

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOPT070

normalization ensures that each predicted bunch property is equally weighted in the loss calculation. We use the Adam optimizer with a batch size of 64, initial learning rate of 10^{-3}, and train for 200 epochs to prevent overfitting. The loss function is defined as the mean squared error (MSE) between the predicted and target bunch properties.

Table 2: Neural network training parameters

Hyperparameters	Value
Loss function	MSE
Batch size	64
Optimizer	Adam
Learning rate	10^{-3}
Epoch	200

The trained model is then evaluated on test data, consisting of 10^3 random parameter settings to which the NN is not trained on. The prediction error is shown in Fig. 3, where almost all the predicted values show good agreement to the target outputs. The prediction of the mean energy E is slightly lower than the target values, but the overall discrepancy is still well under 0.1 MeV.

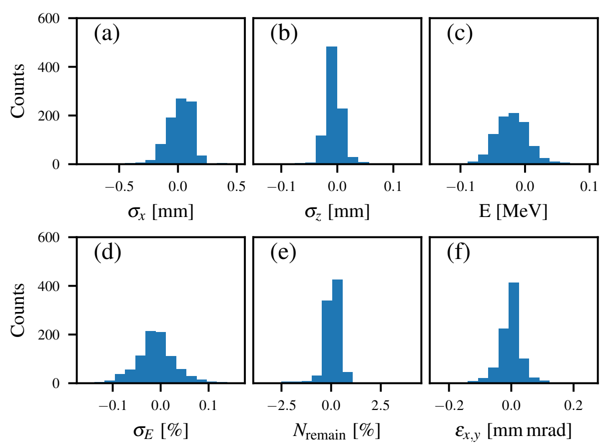

Figure 3: Prediction error on the test data for each output variable: (a) transverse beam size, (b) bunch length, (c) mean energy, (d) relative energy spread, (e) remaining particles, and (f) transverse emittance.

Once the surrogate model is trained, it can predict the beam properties very fast. As an example, Fig. 4 shows a 2D subspace of the 4D input parameter space of the transverse beam size σ_x depending on the solenoid B-field and bunch charge. The left plot is generated by ASTRA simulations with a 10×10 grid, which takes ~5 h for serial execution. The right plot is a 50×50 grid predicted by the trained NN surrogate, which only takes milliseconds in total. As shown in the figure, the NN-predicted beam sizes have a similar structure as the ASTRA simulation results, e.g. the solenoid field required to minimize the beam size at the screen is slightly increasing with the bunch charge due to space charge effects. Deviations from the ideal field strength lead to larger σ_z due to under- or over-focusing of the bunch. The minimum beam sizes, marked as white stars, are very comparable, with $\sigma_x = 0.166$ mm for the ASTRA simulation and $\sigma_x = 0.164$ mm for the NN prediction.

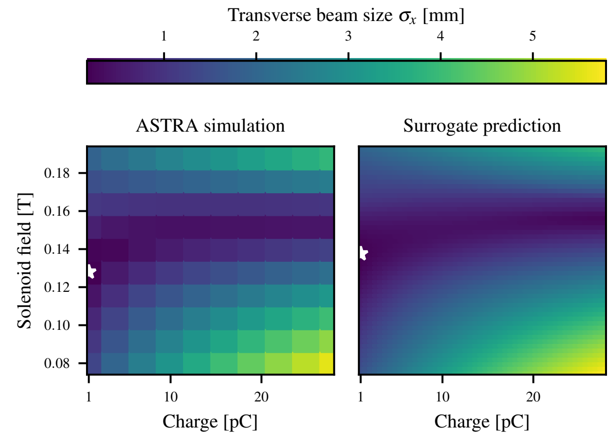

Figure 4: Example of 2D subspace with fixed RF gun phase at 200 deg and gradient 75 MV/m by ASTRA 10×10-grid scan (left) and NN surrogate 50×50-grid prediction (right): transverse beam size σ_x depending on solenoid B-field and bunch charge, where the minima are marked as white stars.

SURROGATE MODEL APPLICATIONS

Virtual Diagnostics

One goal of this study is to use the surrogate model as an online virtual diagnostic for accelerator operation. This could provide shot-to-shot predictions of the beam parameters in a non-destructive manner. For example at FLUTE the energy is measured destructively with a spectrometer dipole magnet, also shown in Fig. 1. The bending angle of the dipole magnet is proportional to the bunch energy for a fixed magnetic field. By varying the magnetic field, the bunch can be deflected and observed on an yttrium aluminium garnet (YAG) screen in the spectrometer arm, where the energy can be determined from the bunch position. Even with an automated procedure, an energy measurement could take up to several minutes, because the dipole needs to be cycled, so that the remnant field does not further affect the normal operation. In order to validate the surrogate model, we compared its results to energy measurements [10] as shown in Fig. 5. For each data point, an ASTRA tracking simulation and surrogate model prediction is performed with the corresponding accelerator parameters. The computation time of the surrogate model is within a millisecond, which is again negligible compared to ~1 h of required ASTRA simulation time and several hours for actual beam time. Both, the surrogate predictions and the ASTRA simulation results, correspond very well to the measurement data with a deviation of under 0.1 MeV. It is interesting to note that the NN was only trained with data within the intervals shown in Table 1, particularly with a minimum gun gradient of 50 MV/m. This corresponds to a gun power of ~3.8 MW, so

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOPT070

the model extrapolated to an unseen parameter range for the first few points in Fig. 5(a), showing good agreement with the measurement results. To further improve the prediction results, the NN model can be partially retrained on measurement data. This could not only reduce the discrepancy between simulation and measurement, but also mitigate the long-term drifts of the accelerator components.

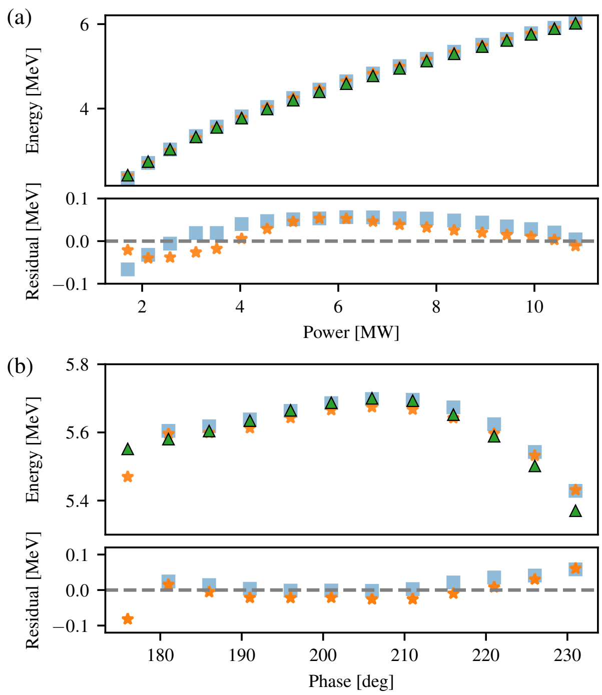

Figure 5: Energy measurements (green triangles) for (a) a power scan and (b) a phase scan of the gun, compared to ASTRA simulations (blue squares) and surrogate model predictions (orange stars). The residuals are the differences to the measurement values. Each measurement is averaged from 20 shots with a standard error of ca. 0.1%.

Fast Evaluation for Other Algorithms

Recently, reinforcement learning (RL) algorithms have been used in the accelerator physics field for the automation of control tasks [11, 12], such as autonomous beam focusing at ARES [13]. Autonomous bunch control with RL is also being considered at FLUTE. RL can also be used for more ambitious goals such as beyond human level control tasks [14–16]. However, the training process of the RL agents is notoriously time consuming, often taking millions of steps. Therefore, it is unpractical to train RL directly on computationally intensive physics simulations or accelerators with a low repetition rate. Thanks to the fast computation time of the surrogate model, it can also be used as a training environment for reinforcement learning. Although the surrogate model has certain discrepancy compared to the actual accelerator, training on it will allow the RL agent to learn roughly the policy and minimize the beam time needed to retrain on the accelerator.

Training a single surrogate model for the whole accelerator including every magnet is unpractical, as the minimum number of training samples required to reasonably represent the accelerator increases exponentially with the number of parameters. For example, generating 10^4 training samples used in this study took about 10 hours, whereas 10^8 simulations are needed to generate samples with same density for 8 input parameters. This corresponds to about 14 years of computation time on the same system. One way to mitigate this is by reducing the parameter space via training surrogate models around some fixed working points. Otherwise, one can also train different models for different stages at the accelerator and connect them for a start-to-end prediction. Despite these limitations, the low-energy section surrogate model can also be used to speed up the optimization of other, more sample efficient algorithms of the full accelerator. For instance, we also developed a Bayesian optimization (BO) algorithm to optimize for intense THz radiation at FLUTE [17]. Due to its exploratory behaviour, the algorithm sometimes samples at undesired operational settings. The surrogate model presented in this study can optimally constrain the parameter space to explore by excluding the settings where the electron bunch has, for example, a large beam size or a large energy deviation when entering the linac. In this way the optimization can be more efficient compared to the presented BO attempt.

CONCLUSION AND OUTLOOK

We trained a NN surrogate model for the low-energy section at FLUTE, which is able to predict important bunch properties like beam size and energy fast and accurately. The prediction results are compared to both ASTRA simulations and measurement results, showing very good agreement. We plan to integrate the surrogate model prediction into the EPICS control system used at FLUTE, so that this could be a daily support tool used in accelerator operation. In the future, more input parameters will be added to the model by reducing the parameter ranges. The outputs will be extended from scalar variables to complete phase space images of the electron bunch. Finally, we plan to build another surrogate model for the second part of FLUTE, which takes the predicted phase space information of the low-energy section model as input and further predicts the bunch properties at the end of the accelerator. With those two surrogate models connected, we can perform online optimization and tailor the bunch properties to meet different experimental requirements.

ACKNOWLEDGEMENTS

This work is in part funded by the Initiative and Networking Fund by the Helmholtz Association (Autonomous Accelerator, ZT-I-PF-5-6). C. Xu and J. Schäfer acknowledge the support by the DFG- funded Doctoral School "Karlsruhe School of Elementary and Astroparticle Physics: Science and Technology".

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOPT070

REFERENCES

[1] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[2] L. Bajer, Z. Pitra, and M. Holeňa, "Benchmarking gaussian processes and random forests surrogate models on the BBOB noiseless testbed," in Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation, 2015. doi:10.1145/2739482.2768468

[3] R. K. Tripathy and I. Bilionis, "Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification," Journal of Computational Physics, vol. 375, pp. 565–588, 2018. doi:10.1016/j.jcp.2018.08.036

[4] C. Emma, A. Edelen, M. J. Hogan, B. O'Shea, G. White, and V. Yakimenko, "Machine learning-based longitudinal phase space prediction of particle accelerators," *Phys. Rev. Accel. Beams*, vol. 21, p. 112 802, Nov. 2018. doi:10.1103/PhysRevAccelBeams.21.112802

[5] J. Zhu, Y. Chen, F. Brinker, W. Decking, S. Tomin, and H. Schlarb, "High-fidelity prediction of megapixel longitudinal phase-space images of electron beams using encoder-decoder neural networks," *Phys. Rev. Applied*, vol. 16, p. 024 005, 2 2021. doi:10.1103/PhysRevApplied.16.024005

[6] M. J. Nasse *et al.*, "First electron beam at the linear accelerator FLUTE at KIT," in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 882–884, doi:10.18429/JACoW-IPAC2019-MOPTS018

[7] M. J. Nasse *et al.*, "FLUTE: A versatile linac-based THz source," *Rev. Sci. Instrum*, vol. 84, no. 2, p. 022705, 2013. doi:10.1063/1.4790431

[8] T. Schmelzer *et al.*, "Diagnostics and first beam measurements at FLUTE," in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 2484–2486. doi:10.18429/JACoW-IPAC2019-WEPGW010

[9] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, 2019, pp. 8024–8035, http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[10] T. Schmelzer et al., "Systematic beam parameter studies at the injector section of FLUTE," in Proc. IPAC'21, Campinas, Brazil, May 2021, pp. 2837–2839. doi:10.18429/JACoW-IPAC2021-WEPAB103

[11] F. H. O'Shea, N. Bruchon, and G. Gaio, "Policy gradient methods for free-electron laser and terahertz source opti-mization and stabilization at the FERMI free-electron laser at Elettra," Phys. Rev. Accel. Beams, vol. 23, p. 122 802, Dec 2020. doi:101103/PhysRevAccelBeams.23.122802

[12] V. Kain *et al.*, "Sample-efficient reinforcement learning for cern accelerator control," *Phys. Rev. Accel. Beams*, vol. 23, p. 124 801, Dec. 2020. doi:10.1103/PhysRevAccelBeams.23.124801

[13] A. Eichler *et al.*, "First steps toward an autonomous accelerator, a common project between DESY and KIT," in *Proc. IPAC'21*, Campinas, Brazil, May 2021, pp. 2182–2185. doi:10.18429/JACoW-IPAC2021-TUPAB298

[14] J. St. John *et al.*, "Real-time artificial intelligence for accelerator control: a study at the Fermilab booster," *Phys. Rev. Accel. Beams*, vol. 24, p. 104 601, Oct. 2021. doi:10.1103/PhysRevAccelBeams.24.104601

[15] W. Wang *et al.*, "Accelerated deep reinforcement learning for fast feedback of beam dynamics at KARA," *IEEE Transactions on Nuclear Science*, vol. 68, no. 8, pp. 1794–1800, 2021. doi:10.1109/TNS.2021.3084515

[16] T. Boltz, "Micro-bunching control at electron storage rings with reinforcement learning," Ph.D. dissertation, 2021.

[17] C. Xu *et al.*, "Optimization studies of simulated THz radiation at FLUTE," presented at IPAC'22, Bangkok, Thailand, Jun. 2022, paper WEPOMS023, this conference.