
Learning Electron Bunch Distribution along a FEL Beamline by Normalising Flows

Anna Willmann

Helmholtz Zentrum Dresden-Rossendorf
Institute of Radiation Physics
a.willmann@hzdr.de

Jurjen Couperus Cabadağ

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf

Yen-Yu Chang

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf

Richard Pausch

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf

Amin Ghaith

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf
Synchrotron SOLEIL

Alexander Debus

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf

Arie Irman

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf

Michael Bussmann

Center for Advanced Systems Understanding
Görlitz

Ulrich Schramm

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf
Technische Universität Dresden

Nico Hoffmann

Institute of Radiation Physics
Helmholtz Zentrum Dresden-Rossendorf

Abstract

Understanding and control of Laser-driven Free Electron Lasers remain to be difficult problems that require highly intensive experimental and theoretical research. The gap between simulated and experimentally collected data might complicate studies and interpretation of obtained results. In this work we developed a deep learning based surrogate that could help to fill in this gap. We introduce a surrogate model based on normalising flows for conditional phase-space representation of electron clouds in a FEL beamline. Achieved results let us discuss further benefits and limitations in exploitability of the models to gain deeper understanding of fundamental processes within a beamline.

1 Introduction and Motivation

Plasma acceleration processes are comprehensively studied in the recent years as high-energy X-rays find many applications such as medical diagnostics and treatment, chip manufacturing or hardening of material surfaces. X-rays are used in research as well as they can provide atomic scale imaging.

One source of high-brilliance coherent X-rays is free-electron lasers (FELs), where a bunch of relativistic electrons is transported through a sequence of optical elements and passed through an undulator - a magnetic structure with varying polarity, where electrons emit photons coherently.

Application of conventional particle accelerators to generate relativistic electrons might end up in kilometers in size, what makes them expensive to keep and maintain. Laser-plasma accelerators(injectors) are compact and can reduce FEL's costs, however successful application of them requires a precise understanding and control over the process.

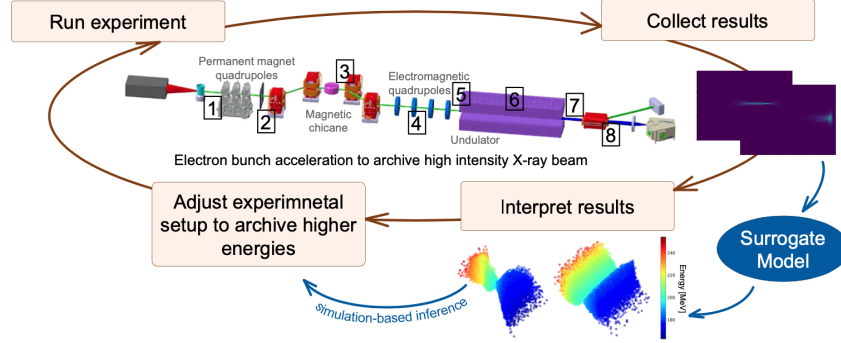


Figure 1: Surrogate model application in laser plasma acceleration research

To achieve clear understanding of the process there are performed many experiments supported by theoretical research. Numerical modeling of a free electron laser consists of multiple parts: simulations of electron acceleration, transport of an electron bunch and radiation in an undulator. In-situ analysis of a FEL by simulation is limited due to complex workflow and potentially long simulation time. Detailed description of used equipment of the beamline together with achieved experimental results is given in Labat et al. [2022].

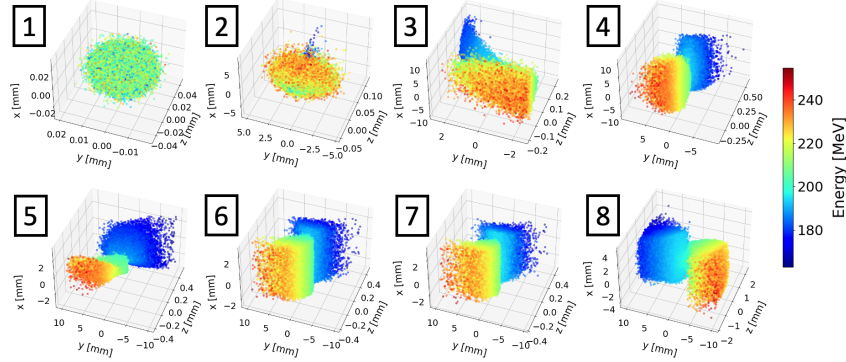


Figure 2: Electron bunch 1 is passed through a sequence of optical elements (fig. 1), each of them is changing the shape of a distribution in order to achieve high coherence of emission in the undulator.

Figure 1 represents a beamline from Andre et al. [2018] for a free electron laser with a sketch of the research pipeline. Electrons are accelerated inside plasma and create a bunch (figure 1, diagnostic screen 1) that is propagated through the manipulation line (figure 1, 2-8) and being transformed as it is shown at figure 2.

The operation of the beamline can be monitored by different diagnostics and quantities, among others, spectrometer measurements at positions of diagnostic screens (e.g. diagnostic screen 2 at figure 1) and intensity of radiation. We can now tune certain performance criteria of the beamline, e.g. brilliance, stability, higher intensity, by adjusting certain parameters of the injector as well as the facility. Alternatively, simulation-based inference (Cranmer et al. [2020]) or data-driven methods for one-step inversion (Hoffmann et al. [2016]) can be used to recover the matching phase-space representation to provide explanation of covered internals of the beamline to advance fundamental research. Simulation-based inference, however, relies on a fast surrogate model that can be used for Bayesian inversion by e.g. rejection-sampling (Approximate Bayesian Computation) or Markov Chain Monte Carlo. Advanced techniques in machine learning based surrogate modeling also promises simulation-based design of accelerators which is currently limited due computational challenges for simulation of the governing system (Lehe et al. [2020]).

Main contribution of this work is therefore a surrogate model that allows for reconstruction of the phase space distribution at different locations of the beamline conditioned on parameters of an accelerated electron bunch by an injector.

2 Method

The non-linear transformation of electrons along a beamline makes it hard to model the corresponding electron distribution deterministically, i.e. by tracking each particle. We therefore propose to learn the distribution of electrons in data-driven fashion by a generative model. There also is an increasing interest in the application of physics-informed neural networks (PINN) for training of surrogate models in physics as demonstrated by (Stiller et al. [2020]). However, we focus on a data-driven surrogate model since PINN increase the computational complexity of the training due to heavy usage of automatic differentiation. There are certain orthogonal state-of-the-art generative models such as GANs(Goodfellow et al. [2014]), variational autoencoders(Kingma et al. [2019]) and normalizing flows(Dinh et al. [2014]) while we concentrate on normalizing flows. Compared to variational autoencoders, it has simpler architecture(only one network instead of encoder and decoder) and does not have a loss of information due to compression. Negative log-likelihood loss function is easier to balance and optimize compared to the adversarial GANs' loss function.

Data. We are working on simulated electron bunches propagated through the beamline and observed at 8 diagnostic screens(figure 1), computed by ELEGANT software package(Borland [2000]).

An electron bunch is represented in phase space, where each particle $\mathbf{p} \in \mathbb{R}^6$ is described by 6 scalars: 3 spatial coordinates (x_c, y_c, z_c) , index c is added to indicate a notation as related to the coordinate space, its energy γ and per-particle divergence in X and Y dimensions(transverse plane). Further we call phase space representation of electron bunch an electron/particle/point cloud, where each electron is characterized by 6 coordinates. Dimension Z (longitudinal coordinates) is a direction of bunch propagation, the bunch is centered around $z_c = 0$.

Each electron bunch consists of approximately 100.000 - 200.000 particles. During the transport particles with only a limited range of energies are propagated to the undulator in order to gain coherent emission, for this reason some particles will be filtered out in the beamline.

In total there are 32 simulations, each simulation consists of 8 electron clouds(1 per diagnostic screen) and identified by statistical emittance(area occupied by a particle cloud) - ϵ_x, ϵ_y and beta function(transverse size of the particle cloud) - β_x, β_y of the initial electron cloud(figure 1, screen 1, before passing any optical element in the line).

An initial distribution for all simulations defined to be a normal distribution for simplicity of method evaluation. In future the initial distribution will be extended to more complex shapes.

Normalizing Flows. Normalizing Flows(NFs) is a family of generative models, first introduced as a non-parametric method to estimate probability density by Tabak and Turner [2013]. Further development of the approach into a deep learning based method was done by L. Dinh, who suggested suitable architectures (Dinh et al. [2014], Dinh et al. [2016]) to apply it to higher dimensional data without limitation of network's complexity.

The goal is to find an invertible mapping between an electron \mathbf{p} and a sampled point $\mathbf{z} \sim \mathcal{N}(0, 1)$:

$$\mathbf{p} = f^{-1}(\mathbf{z}, \epsilon_x, \epsilon_y, \beta_x, \beta_y, \mathbf{s}; \theta),$$

where $\mathbf{x} = (x_c, y_c, z_c, x_p, y_p, \gamma)$ - phase-space coordinates of each particle and $\mathbf{z} = (z_1, \dots, z_6)$, $z_k \sim \mathcal{N}(0, 1)$, $k = 1 \dots 6$, \mathbf{s} - number of a diagnostic screen. Trainable parameters of the network are defined by θ . In our work we have a fixed finite number of diagnostic screens and we pass it in the format of a one-hot vector. Further $(\epsilon_x, \epsilon_y, \beta_x, \beta_y, \mathbf{s})$ will be called a condition and denoted by \mathbf{c} .

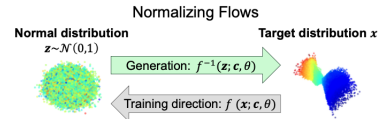


Figure 3: Scheme of normalizing flows

Mapping f is trained in the forward direction to minimize the negative log-likelihood loss function(Ardizzone et al. [2019]):
$$L = \max_{\theta} \sum_{i=1}^N -\frac{\|f(\mathbf{p}_i, \mathbf{c}; \theta)\|_2^2}{2} + \log \left| \det \frac{\partial f(\mathbf{p}_i, \mathbf{c}; \theta)}{\partial \mathbf{p}_i^T} \right|$$

Enhancement of normalizing flows by a masked autoregressive model was suggested by Papamakarios et al. [2017], where instead of pure samples z authors suggest to add dependence on all previous samples: $p_i = z_i \exp(\mu_i) + \alpha_i$, $\mu_i = f_{\mu_i}(p_{1:i-1}, c; \theta)$, $\alpha_i = f_{\alpha_i}(p_{1:i-1}, c; \theta)$, where i is a number of a sampled point, f_{μ_i}, f_{α_i} - MADE networks from Germain et al. [2015]. This architecture allows not to compute μ_i and α_i recurrently but by a single forward pass for each i . N is a total number of samples in a training set.

This re-formulation generalizes the original approach and has a potential to higher expressiveness than conventional NFs.

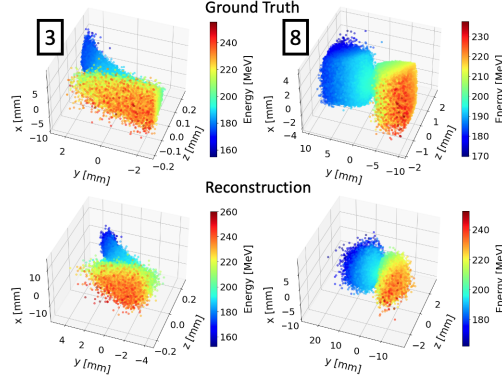


Figure 4: Approximation of point clouds from validation set

Model training. In this work we used two models: a conditional normalizing flow(from Ardizzone et al. [2019]) and masked autoregressive flow(MAF) from Papamakarios et al. [2017] with a condition. For the conditional NF we use implementation of GLOW blocks(Kingma and Dhariwal [2018]) provided by FrEIA package(Ardizzone et al. [2018-2022]) and masked autoregressive flow with affine blocks is implemented in *nflows*(Durkan et al. [2020]). Both models consist of 7 coupling blocks with MLP subnetworks.

All passed data are normalized to lie within $[0, 1]$. We sample 10.000 particles from each point cloud per iteration, 70 point clouds are processed per batch. Conditional NF model is trained until convergence over 10.000 iterations using Adam optimizer

with learning rate of 10^{-5} . The reached loss on training data is -14.474 , and on validation data: -11.705 . The model was trained on a single NVIDIA Tesla P100 16Gb node over 26 hours.

Masked autoregressive model was trained over 13.750 iterations using Adam optimizer with learning rate of 10^{-3} . The training was distributed(horovod package for distributed learning: Sergeev and Balso [2018]) over 4 NVIDIA Tesla V100 32Gb nodes. Training took 16 hours, reached training loss of -18.657 and loss on validation data: -18.485 . Overfitting was not observed.

3 Numerical results

A disadvantage of MAF is a high computational demand, though in the context of our problem input and output have low enough dimensionality.

The models are compared by the following metrics: MMD(Gretton et al. [2012]) and Sinkhorn distance(upper bound of EMD, introduced in Cuturi [2013]), quantitative results are presented in table 1. Metrics are given in average value(among all clouds in validation set) and the worst reconstructed cloud. The masked autoregressive model has shown significantly better reconstruction of point clouds, even the lowest reconstruction quality is higher than a corresponding average value for conditional normalizing flows.

Figure 4 demonstrates reconstructed point clouds made by MAF model. Shapes of distributions at different positions in the beamline were recognized and approximated correctly. Figure 5 shows projections of point cloud on different planes. The top row corresponds to the number of particles per a slice: slices are taken in direction Z , each slice has the same width, bottom - 2D projection of particles on plane Z against energy. As we see, approximated distribution by the conditional NF model does not cover the complete volume of the cloud and most of particles are located around 0, while the MAF model successfully tackles this issue.

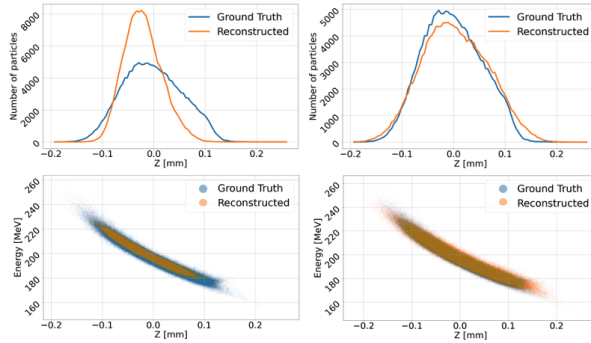


Figure 5: Projections of an approximated point cloud: left column - conditional NFs, right column - MAF

Table 1: Statistical distance on validation data

Model	MMD, average	MMD, worst	Sinkhorn, average	Sinkhorn, worst
NF	0.03293	0.06590	0.29618	0.7245
MAF	0.00104	0.00180	0.08257	0.1343

Limitations. The dataset has to contain smoothly changing electron clouds to achieve meaningful interpolation. The number of points in each cloud has to be large enough to represent distribution and allow to sample at each training iteration. A low number of particles per cloud cause not stable approximation in a point-wise mapping manner. Optimization of hyperparameters plays significant role due to invertibility of the model: parameters of subnetworks have to lie within a given range such that results of forward and backward passes do not explode or vanish. This limits complexity of transformations, the balance between width of that range and model complexity has to be found. An example of this problem one can observe at figure 4. The reconstruction has certain outliers e.g. in YX-plane: range of values in y axes for ground truth data is around $[-4, 4]$, while in a reconstructed one: $[-20, 20]$.

4 Conclusions

Surrogate modeling can help to optimize experimental research, increase degree of understanding of physical processes in FELs and simplify control over them. High costs of numerical simulations motivate to use deep learning based methods in order to derive fast and reliable simulation based inference. We suggest a masked autoregressive flow based model for conditional generation of electron clouds, that is an important part in the beamline inversion: in order to extract the simulation parameters from diagnostic measurements. Results indicate better performance than conditional NFs due to higher expressivity that was theoretically proven by Papamakarios et al. [2017].

5 Impact statement

The developed surrogate model could significantly accelerate the experimental research of free electron lasers. It provides an opportunity to replace simulations by a faster model and optimize parameter space to achieve desired results in experimental companion. The negative impact might come from reliability of the model for cases when experimental results do not correspond data used for model training or model was not verified well enough. It might cause inconsistent conclusions from collected data during the experiment and bring to wrong interpretations. Development of reliable digital twins is a crucial step towards an AI-guided experiments, that can reduce number of experimental runs to explore design space as well as amount of computational resources to produces required simulations. It will cause a significant decrease of energy consumption and make research and development of FELs more sustainable.

References

- T. Andre, I. Andriyash, A. Loulergue, M. Labat, E. Roussel, A. Ghaith, M. Khojoyan, C. Thauray, M. Valléau, F. Briquez, F. Marteau, K. Tavakoli, P. N’Gotta, Y. Dietrich, G. Lambert, V. Malka, C. Benabderrahmane, J. Veteran, L. Chapuis, and M.-E. Couprie. Control of laser plasma accelerated electrons for light sources. *Nature Communications*, 9, 04 2018. doi: 10.1038/s41467-018-03776-x.
- L. Ardizzone, T. Bungert, F. Draxler, U. Köthe, J. Kruse, R. Schmier, and P. Sorrenson. Framework for Easily Invertible Architectures (FrEIA), 2018-2022. URL <https://github.com/VLL-HD/FrEIA>.
- L. Ardizzone, C. Lüth, J. Kruse, C. Rother, and U. Köthe. Guided image generation with conditional invertible neural networks. *arXiv preprint arXiv:1907.02392*, 2019.
- M. Borland. Elegant: A flexible sdds-compliant code for accelerator simulation. 8 2000. doi: 10.2172/761286. URL <https://www.osti.gov/biblio/761286>.
- K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.
- M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- L. Dinh, D. Krueger, and Y. Bengio. Nice: Non-linear independent components estimation, 2014. URL <https://arxiv.org/abs/1410.8516>.
- L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using real nvp, 2016. URL <https://arxiv.org/abs/1605.08803>.
- C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. nflows: normalizing flows in PyTorch, Nov. 2020. URL <https://doi.org/10.5281/zenodo.4296287>.
- M. Germain, K. Gregor, I. Murray, and H. Larochelle. Made: Masked autoencoder for distribution estimation. In *International conference on machine learning*, pages 881–889. PMLR, 2015.
- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.
- N. Hoffmann, E. Koch, G. Steiner, U. Petersohn, and M. Kirsch. Learning thermal process representations for intraoperative analysis of cortical perfusion during ischemic strokes. In G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, and J. Cornebise, editors, *Deep Learning and Data Labeling for Medical Applications*, pages 152–160, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46976-8.
- D. P. Kingma and P. Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018.
- D. P. Kingma, M. Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.
- M. Labat, J. Couperus Cabadağ, A. Ghaith, A. Irman, A. Berlioux, P. Berteaud, F. Blache, S. Bock, F. Bouvet, F. Briquez, Y.-Y. Chang, S. Corde, A. Debus, C. Oliveira, J.-P. Duval, Y. Dietrich, M. Ajjour, C. Eisenmann, J. gautier, and M. Couprie. Seeded free-electron laser driven by a compact laser plasma accelerator. 2022. doi: 10.21203/rs.3.rs-1692828/v1.
- R. Lehe, A. Hanuka, A. Edelen, X. Huang, C. Mayes, N. Cook, A. Huebl, C. Emma, B. Djordjevic, A. Kemp, et al. Machine learning and surrogates models for simulation-based optimization of accelerator design. *Snowmass21 LOI*, 2020.

- G. Papamakarios, T. Pavlakou, and I. Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017.
- A. Sergeev and M. D. Balso. Horovod: fast and easy distributed deep learning in TensorFlow. *arXiv preprint arXiv:1802.05799*, 2018.
- P. Stiller, F. Bethke, M. Böhme, R. Pausch, S. Torge, A. Debus, J. Vorberger, M. Bussmann, and N. Hoffmann. Large-scale neural solvers for partial differential equations. In J. Nichols, B. Verrastegui, A. B. Maccabe, O. Hernandez, S. Parete-Koon, and T. Ahearn, editors, *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI*, pages 20–34, Cham, 2020. Springer International Publishing. ISBN 978-3-030-63393-6.
- E. Tabak and C. Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66, 02 2013. doi: 10.1002/cpa.21423.

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
 - (b) Did you describe the limitations of your work? [Yes] section 3
 - (c) Did you discuss any potential negative societal impacts of your work? [Yes] section 5
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
 - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
<https://github.com/Photon-AI-Research/NF4ElectronDistribution>
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] section 3
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] section 3
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? [N/A] we use simulated data that we created on our own, creators of simulation code are cited
 - (b) Did you mention the license of the assets? [N/A]
 - (c) Did you include any new assets either in the supplemental material or as a URL? [No]
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] used data only contains simulated physical process
5. If you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]