Denoising Diffusion Implicit Models for Laser-Plasma Accelerator Simulation Trained With Physical Constraint Loss

Matěj Jech[1,2] ✉, Gabriele Maria Grittani[2], Carlo Maria Lazzarini[2], and Alexander Kovalenko[1]

[1] Czech Technical University in Prague, Thákurova 9, Prague, Czechia
{jechmate,alexander.kovalenko}@fit.cvut.cz
[2] The Extreme Light Infrastructure ERIC, Za Radnicí 835, Dolní Břežany, Czechia
{matej.jech,gabrielemaria.grittani,carlomaria.lazzarini}@eli-beams.eu

Abstract. The paper addresses simulating laser wakefield electron acceleration experiments by using a novel training methodology for probabilistic diffusion models with adherence to the foundational physical principles. The methodology allows overcoming the inability of the common generative machine learning models to capture external data manifold constraints due to the nature of training. Laser wakefield electron acceleration is a highly complex non-linear phenomenon with a developed approximation theory, which, however, falls short in many extreme cases. Applying a model trained with physical constraint loss to simulate these experiments demonstrates strong performance and produces results that are positively evaluated by experts in the field. Moreover, due to the embedded physical information, it can extrapolate outside the range of training input data, based on the known physics of the process. This approach shows immense potential for using generative models in the modeling of complex scientific experiments, which helps in efficient experiment planning and optimization. We evaluate the generative models using Wasserstein distance calculated between distributions of charge of electrons at corresponding energies. This metric provides a robust quantification of the similarity between the predicted and reference electron energy spectra.

Keywords: Diffusion models · Laser wakefield electron acceleration · Physics-informed machine learning · Generative modeling

1 Introduction

Despite the overall success of probabilistic diffusion models in image generation [1], their application remains largely confined to entertainment and design, possessing significant underrepresentation in the broader practical domain. The limitations that hinder the application of probabilistic diffusion models in other domains arise from learning solely from visual and textual representations that often do not reflect in learning of an actual world model. One of such limitations

is the use of physical constraints in generated data [2]. Physics-informed neural networks [3, 4, 5] (PINNs) are a powerful tool for efficient learning the distribution of data adherent to the foundational physical principles. PINNs help in reducing the required amount of data by embedding physics as part of the loss function. This helps achieve more scientifically reasonable results given fewer experimental observations, as we do not rely solely on passive information present in the dataset. By incorporating physical constraints and specific we can enhance the model's ability to generate data that adheres to real-world physical phenomena.

One promising application of such physics-constrained diffusion models is in the optimization of particle accelerators. These powerful scientific instruments enable us to study the fundamental properties of matter and the universe [6]. One of the most significant discoveries made using particle accelerators is the Higgs boson, a fundamental particle that gives other particles mass. The Large Hadron Collider at CERN [7], the world's largest and most powerful particle accelerator, discovered the Higgs boson in 2012 [8], confirming a key prediction of the Standard Model of particle physics [9]. However, the development of laser-plasma accelerators has enabled significant miniaturization of particle accelerators, making them indispensable in various scientific fields. These devices are workhorses in myriad fields of science, blasting out fundamental particles and generating intense beams of X-rays for studies of biomolecules and materials. Using laser wakefield acceleration (LWFA) [10, 11], kilometers-long devices [12] can be scaled down into a centimeters scale and, for example, the Large Hadron Collider, the biggest and most powerful of them all, potentially could fit in a classroom.

However, despite the main advantage - the compact size - laser wakefield electron accelerators are characterized by a plethora of non-linearly dependent parameters along with a high sensitivity to their changes [13]. Furthermore, these experiments are influenced by stochastic-like influences from temperature, lighting conditions, material deterioration or even other unknown sources. This makes the development and optimization of them a difficult task [14]. Therefore at the moment obtaining high-quality laser-plasma accelerated electron beams poses a significant challenge, given its crucial role in numerous applications of healthcare, notably in medical imaging [15, 16], radiation medicine [17, 18] and cancer treatment [19]. Additionally, they are invaluable in future research domains that would require accelerated particles, offering more compact and efficient alternatives to traditional accelerators. This process requires the careful manipulation of various interconnected physical phenomena, including plasma wave generation, electron injection, betatron radiation, and beam phase space evolution. The simulation of these processes demands sophisticated numerical methods and high-performance computational resources to solve the governing equations, namely Maxwell's and relativistic fluid equations. Conducting experiments to optimize these beams is complex and costly due to the nonlinear, transient nature of laser-plasma interactions and the plethora of instabilities that arise. Furthermore, the experimental parameter space is extremely high-dimensional, encompassing nu-

merous variables from both the laser system (such as pulse energy, duration, and focusing conditions) and electron acceleration (including plasma density, interaction length, and injection conditions). This multitude of interconnected parameters creates a prohibitively large search space for optimizing experimental outcomes.

To address these challenges, the adoption of artificial intelligence and generative deep learning models, such as Denoising Diffusion Probabilistic Models (DDPM) or those based on Nonequilibrium Thermodynamics, can prove beneficial. Diffusion models, which employ physics-inspired diffusion processes to iteratively refine a random initial state into a sample from the target distribution, have gained significant recognition for their capability to model complex, inherently probabilistic, high-dimensional distributions. In their comprehensive review, Döpp et al. [20] provided a detailed analysis of machine learning applications in laser-plasma science, demonstrating how various data-driven techniques have transformed experimental design, theoretical modeling, and diagnostic analysis

These models can be used to conditionally generate novel electron beam spectra based on the input parameters, effectively augmenting the search space without necessitating additional expensive and time-consuming experiments. This approach allows for more efficient exploration of the parameter space and increases the probability of identifying optimal conditions.

By integrating reinforcement learning agents [21] or other optimization algorithms [22] with the diffusion model, the search process can be guided more efficiently, learning the complex dependencies within the data. Feedback from the experiments, based on the parameters suggested by the model, can be utilized to further refine the model, thereby enhancing its predictive power and the quality of the generated samples. Consequently, this approach establishes a virtuous cycle of learning and optimization, demonstrating the potential of combining physics and AI to tackle this complex challenge. Bayesian optimization has already proven to be a viable tool for LWFA experimental parameter tuning achieving an increase in both stability and electron energy [14, 13].

Therefore the present article is focused on developing a probabilistic diffusion model that takes into account physical theory related to particle acceleration induced by laser. Given the complexity of the data from the experiments and, often, the high cost of scientific trials, generative models based on physics-informed probabilistic diffusion models for spectral generation can be extremely beneficial for experimental design as they can support and accelerate scientific discovery by augmenting a search space for the conducted experiments.

2 Laser-Plasma Accelerator Setup

The key apparatus in these experiments was the laser system, operating at a frequency of 1 kHz and delivering high pulse energy at approximately 26 mJ. This multi-cycle, 15 fs pulse laser, based on optical parametric chirped pulse amplification (OPCPA) technology, was pivotal in enabling laser wakefield acceleration

(LWFA). Unlike earlier systems requiring single-cycle pulses, the design proved capable of achieving electron beam energies up to 50 MeV at a high repetition rate, showcasing a 25% energy spread and 2 mrad beam divergence. [23]

The laser was directed toward a gas target, composed either of pure nitrogen or a nitrogen-helium mixture of gases. Upon laser interaction, the gas undergoes ionization, forming plasma. As the laser traverses the plasma, it generates a wakefield in its trail. Electrons within the plasma are then drawn into this wakefield and accelerated to high energies, essentially "surfing" [24] in a trajectory parallel to the laser's path. The accelerated electron beam is directed toward a spectrometer equipped with a LANEX Fast Back scintillator screen. The interaction of electrons with the screen results in luminescence at the points of impact, enabling their detection and capture via a camera [25].

Another aspect of this setup is the incorporation of a 5mm aluminum slit and a magnetic dipole positioned before the screen. The magnet's role is to deflect the electron beam, causing a displacement relative to the beam's original path. By analyzing this displacement and applying a deflection curve specific to the experimental setup geometry, it is possible to measure the energy levels of the electrons striking the screen. The slit is used to center the spectrometer on the average electron beam pointing. A detailed illustration of the experimental setup is depicted in Figure 1, providing a comprehensive visual overview of the entire process and apparatus.

During the experiments, the electron beams were generated with energies reaching up to 50 MeV, additionally characterized by a low energy spread. It was observed that small changes in the laser and plasma parameters significantly affected the energy distribution of accelerated electrons. [23]

3 Data Processing Pipeline

The dataset obtained from the scientific group exhibited a high level of noise, resulting in grainy images. To mitigate this, the data was preprocessed using a median filter with a kernel size of 5. A larger kernel did not appear to improve noise removal significantly. Additionally, the images were taken by the camera with a damaged sensor making parts of the images contain irrelevant outlying values. These parts were manually set to 0 as is usually the background value. Fortunately, these errors never appear near the electron beams and therefore such treatment is sufficient.

The Lanex screen, which is hit by accelerated electrons, has black dots arranged in a cross shape to facilitate distance calculations. These dots obscured the image and needed to be removed. This process was automated using a dot-detection algorithm. The image is first converted to binary using adaptive thresholding. The main parameters are *blockSize*, which controls the size of the neighborhood used to calculate the segmentation threshold, and C, a constant subtracted from the mean of the neighborhood to obtain the final threshold value. After experimentation, these parameters were set to 51 and 4 respectively. The dots were then detected using the circle Hough transform

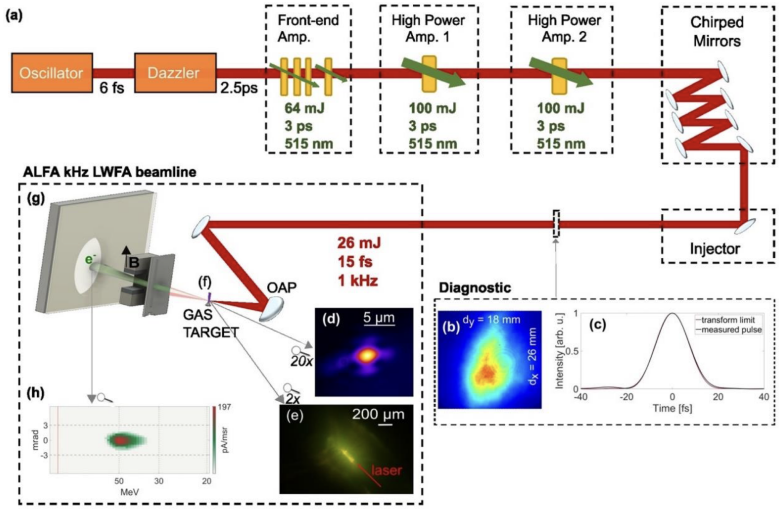


Fig. 1: From [23]: Experimental setup of the laser-plasma accelerator. (a) The laser system. (b, c, e) Various laser-related diagnostics. (d) View of the focal spot of the laser. (f) The gas-jet target of the laser. (g) Electron spectrometer. (h) Electron-beam trace projected on the Lanex screen.

[26]. The OpenCV [27] implementation was used with the following parameters: *minDist=20, param1=50, param2=7, minRadius=0, maxRadius=10*. This resulted in more detected dark dots in the image than actually present, which is preferable to the alternative. The black dots were then removed by interpolating along the x axis, meaning that the "removal" of non-existent dots had minimal effect on the resulting image given the gradient-like nature of the images.

In addition to the data with the magnet in the aperture, which is used to measure the energy of the electrons, there are corresponding datasets without the magnet used to acquire the position of the electron beam. This position is then used as a reference point when calculating the energy of the accelerated particles. The position of the electron beam was estimated by summing all images of a given experiment and then searching for the position of the maximum in the summed image. This point of highest intensity is taken as the center of the electron beam.

The original images contained large areas either outside of the Lanex itself or far from the electron beam pointing pixel, that did not contain any valuable information. The images were cropped to a height of 256 pixels and width of 512 pixels, with the electron pointing pixel fixed at coordinates [128, 62] (vertical and horizontal position). This position was selected because the vertical axis is similarly important, and on the horizontal axis, the deflection by the magnet occurs towards positive values of x. However, the image size was still excessive to be used as an input for training a DDIM with many convolutional layers and

would require more VRAM than was available for this project. Cropping the images to a smaller size of 64 by 128 pixels resulted a significant information loss as many training data contained only lower energy electrons which would be deflected outside of the new horizontal image range. To overcome this issue, the images were resized using bilinear interpolation to a final size of 64 by 128 pixels before being used for training. At the end of the sampling process used for image generation, the images were resized back to the original size. Due to the gradient-like nature of most of the images, this process of repeating interpolation does not result in much loss in the fine detail of the images. The preprocessing pipeline demonstrated on one data sample can be seen in Figure 2. The preprocessed dataset can be downloaded from this link.

The dataset consists of folders of images of electron spectra from 22 experiments with varying values of the following parameters:

- **E** - energy of the laser (12–26 mJ)
- **P** - pressure of the gas jet (10–38 Bar)
- **gas** - type of gas (Nitrogen or Helium/Nitrogen mixture)
- **ms** - acquisition time of the camera/opening time of the gas valve (10–40 ms)
- **gain** - gain setting on the camera

In the process of training the model, only a subset of the parameters, specifically **E**, **P**, and **ms**, were incorporated into the conditional vector. The rationale behind this selection was based on the relevance and variability of these parameters in the context of the experiments.

The parameter **gas**, which denotes the type of gas used in the experiments, was deliberately excluded from the conditional vector. This decision was driven by the fact that only two experiments were conducted using a gas mixture, while the rest employed Nitrogen. Consequently, the **gas** parameter exhibited very little variation across the dataset, rendering it less informative for the purpose of training the model.

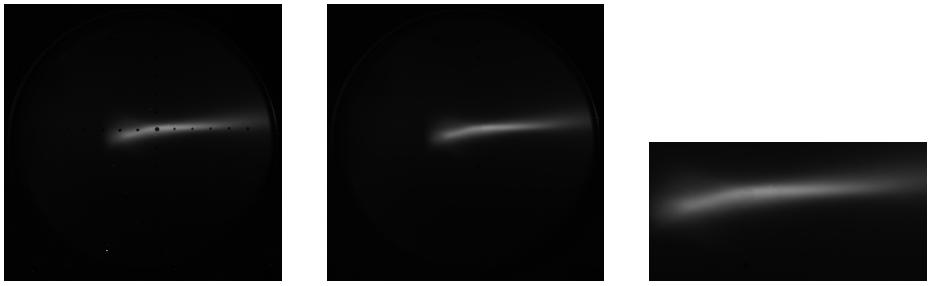
Furthermore, both **gain** and **ms** are parameters that influence the brightness of the captured images. However, the **gain** parameter was not included in the conditional vector. The influence of the **gain** setting on the image brightness is deterministic, meaning it follows a predictable and consistent pattern. Therefore, its effect can be accounted for by subtracting it from the images during the preprocessing stage. This allows the model to focus on learning from the more variable and informative aspects of the data, potentially improving its performance and generalization ability.

The data was collected using a Basler aca2040-25gm camera. According to the official documentation [28], the formula for calculating gain is:

$$\text{Gain} = 20 * \log_{10}(\text{GainRaw}/32) \tag{1}$$

Here, *GainRaw* is the value set within the Basler Pylon Viewer through which camera settings were adjusted during the experiment. After recalculating the value of *Gain* from decibels (dB) to linear units, the gain value is used to adjust the brightness of the image. However, it was experimentally found that

subtracting the gain value from the dataset causes the network to perform poorly during training. It is assumed to be caused by weak signal from dataset features. Therefore, the gain value was left as is. This may result in some inaccuracies when accessing the intensity of the generated samples which corresponds to the total charge of the accelerated electrons. However, this inaccuracy was deemed unsubstantial and a better alternative to poor training performance.



(a) Original image (b) Dots removal (c) Preprocessed image

Fig. 2: Data preprocessing pipeline

4 Methodology

The full implementation can be found in the github repository including training, validation, sampling and all hyperparameters.

4.1 Denoising Diffusion Implicit Model

The proposed diffusion process adheres to the following distribution as introduced by Song et al. [29]:

$$q(\mathbf{X}_{t-1}|\mathbf{X}_t, \mathbf{X}_0) = \mathcal{N}(\sqrt{\alpha_{t-1}}\mathbf{X}_0 + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\frac{\mathbf{X}_t - \sqrt{\alpha_t}\mathbf{X}_0}{\sqrt{1-\alpha_t}}, \sigma_t^2 I) \quad (2)$$

For the α_t noising parameters the cosine schedule was chosen based on its superior performance over a linear schedule [30].

The objective function for this model is most commonly defined as the MSE between the removed noise, $\tilde{\epsilon}_t$ and the actual noise ϵ_t, that was added during step t in the forward diffusion process. This is calculated over a batch of size B, as shown in the following equation:

$$MSE = \frac{1}{B}\sum_{j=1}^{B}(\epsilon_{t_j} - \tilde{\epsilon}_{t_j})^2 \quad (3)$$

The input image is processed through a U-Net architecture [31], which is composed of three downsampling and upsampling blocks. The initial processing stage involves a double-convolution block, which applies a 2D convolutional layer, group normalization, and a GELU activation function twice in succession. Following this, the first downsampling block is applied, which consists of a 2D maxpooling layer and two subsequent double-convolution blocks.

In addition to images, the network processes the settings vector and the noising step number. The settings vector, denoted as y, is processed through a block of layers, which includes batch normalization, a linear layer, and a SiLU activation. The noising step number, denoted as t, is encoded using positional encoding to provide the model with more nuanced information about the noise step beyond a single integer value. The tensors y and t are combined and serve as inputs to the network's Down and Up blocks, where they undergo further processing through a SiLU activation [32] and a linear layer before being added to the tensors derived from the image data. This approach equips the model with both conditional information and the remaining number of noise steps to be eliminated.

At the network's core, there are three double-convolution blocks, succeeded by the upsampling phase. The Up block mirrors the structure of the Down block, except for an upsampling layer replacing the maxpooling layer to enlarge the tensor dimensions. Layers are concatenated with a skip connection, utilizing the output from the corresponding Down block.

Each Down and Up block is followed by a SelfAttention [33] block. Initially, the input is reshaped to conform to the data ordering expected by the MultiheadAttention layer and is then normalized using layer normalization. The output is subsequently combined with the original input, followed by a sequence of layer normalization, linear transformation, GELU activation [34], and another linear transformation. Prior to reshaping the output to maintain consistency with the network's data ordering, the output of the MultiheadAttention layer is added once more. Overall the architecture of the model was hand-tailored, however we leveraged transfer learning [35] and used a subset of pre-trained convolutional filters from previous implementation of a DDPM model [36]. A simplified scheme of the architecture can be seen in Figure 3.

The enhanced DDIM sampling sequence, as described in [29], was implemented utilizing the source code for the diffusion autoencoder [37]. This implementation leverages the underlying DDPM noising schedule to derive a new schedule with a reduced number of steps. Furthermore, it offers the flexibility to distribute the new steps non-uniformly, enabling a more precise focus on specific segments of the denoising process. A sample $\mathbf{X}_{t-1}$ can be generated from a sample $\mathbf{X}_t$ using the following equation [29]:

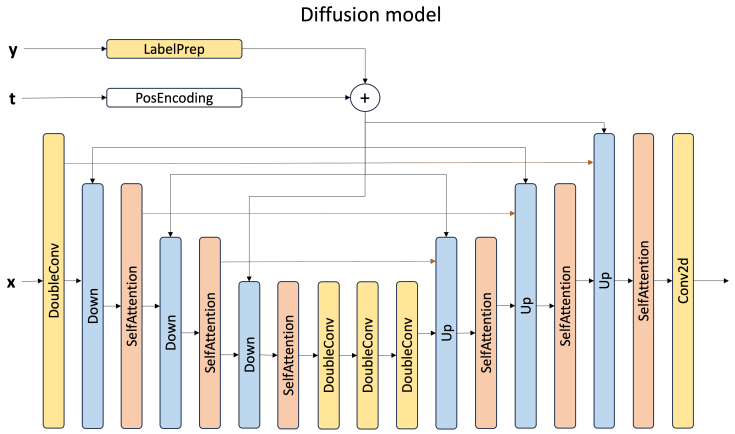


Fig. 3: A schematic of the model architecture. Black arrows show the flow of data, red arrows signify skip connections.

$$\mathbf{X}_{t-1} = \sqrt{\alpha_{t-1}} \underbrace{\left(\frac{\mathbf{X}_t - \sqrt{1-\alpha_t}\epsilon_\theta^{(t)}(\mathbf{X}_t)}{\sqrt{\alpha_t}} \right)}_{\text{predicted } \mathbf{X}_0} + \underbrace{\sqrt{1-\alpha_{t-1}-\sigma_t^2} \cdot \epsilon_\theta^{(t)}(\mathbf{X}_t)}_{\text{direction pointing to } \mathbf{X}_t} + \underbrace{\sigma_t \epsilon_t}_{\text{random noise}} \tag{4}$$

Upon completion of the sampling process, the output image is reshaped to the desired dimensions. In the final implementation, this step is crucial for generating the image in its original size, as the training dataset was previously rescaled to minimize computational demands.

4.2 Physical constraint loss

The physical constrained loss is defined in two parts corresponding to different aspects of the experiments. First, the ability of the model to precisely model the energy and charge of electrons is tested. This loss function is further referred to as L_{phys1}. Second, the model is penalized for generating electrons left of the electron-pointing beam. Such particles would have beyond infinite energy as their distance to the beam-pointing pixel relative to the magnet is negative. The loss corresponding to this constraint is termed L_{phys2}.

Both L_{phys1} and L_{phys2} use a partially noised image. L_{phys1} compares an image partially altered by the noise distribution learned by the model and an image noised by the forward diffusion process. L_{phys2} uses an image partially denoised by the model. Both L_{phys1} and L_{phys2} are controlled during training by a scheduled weight parameter. This weight is parameterized by the current denoising step number t. During the first steps where t is high, the constraint losses are virtually unused since the image is too noisy to discern any real features. For the

schedule of the weight, a cosine schedule was chosen as it mirrors the approach used in the forward diffusion process. The formula for the scheduling is defined as:

$$\text{weight}_{\text{phys}}(t) = 0.5 \times \left(1 + \cos\left(\frac{t}{1000} \times \pi\right)\right) \tag{5}$$

Electron beam spectra simulation From each image of the accelerated electrons the total charge of electrons at a certain energy level can be calculated. This is based on the distance of the deflected electrons from the original position of the electron beam. A training image, along with its calculated electron beam spectrum, is visualized in Figure 4.

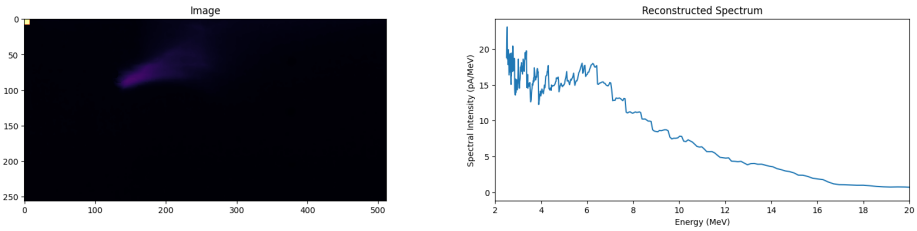


Fig. 4: An image from the training dataset with the 1D spectrum calculation. The image is rendered in monochrome, with a colormap applied to enhance visual clarity and facilitate easier interpretation.

This calculation is used for the first part of the physics constraint loss by calculating it from an image noised by the forward diffusion process and an image noised by the distribution learned by the model. This encourages the model to learn the distribution of noise more precisely by ensuring the noising process has little or no effect on the calculation of the spectra. The spectrum in MeV is calculated by using the deflection curve of the magnet inserted into the aperture. In the original script, this is done by interpolating between measured points of the curve of the magnet. For ease of loss computation we use a double exponential decay function:

$$f(d) = a_1 e^{-\lambda_1 d} + a_2 e^{-\lambda_2 d} \tag{6}$$

fitted as:

$$\hat{f}(d) \approx 77.86 e^{-0.47 d} + 19.91 e^{-0.04 d} \tag{7}$$

The spectrum is then calculated by taking the horizontal sum of the image denoted as x_{sum} into:

$$S(j) = \left\{ \frac{x_{\text{sum}}[j]}{\hat{f}(j-1) - \hat{f}(j)} \mid j \in \mathbb{Z}, 62 \leq i < \text{len}(\mathbf{X}) \right\} \tag{8}$$

where 62 is the position of the electron pointing pixel (point of infinity) and $\text{len}(\mathbf{X})$ is the horizontal size of the image. The spectrum is then calibrated by $C = 3.706$ which is a conversion value from pixel intensity to nA/MeV, the acquisition time t_{acq} and gain g by

$$S_{\text{MeV}}(x) = \frac{S(x) \cdot C}{t_{\text{acq}} \cdot g} \tag{9}$$

for a final loss of:

$$L_{\text{phys1}} = \text{MSE}\left(S_{\text{MeV}}(\mathbf{X}_t), S_{\text{MeV}}(\hat{\mathbf{X}}_t)\right) \tag{10}$$

where $\hat{\mathbf{X}}_t$ is the image noised by the reverse process learned by the model.

Electron beam pointing pixel The second part of the loss uses the fact that no electrons should appear left of the electron-pointing pixel as they would have an impossible higher than infinite energy. It uses a transition smoothed by a sigmoid function applied over the image. First, for defining the sigmoid transition at the correct area we need to define a calculation of horizontal distance as:

$$d_{X_i} = \{(i - 62) \cdot 0.137 \mid i \in \mathbb{Z}, -62 \leq i < \text{len}(\mathbf{X}) - 62\} \tag{11}$$

where 62 is the position of the electron beam, 0.137 is the ratio of pixels to millimeters specific to the experimental camera setup, i is the horizontal index of the pixel and $\text{len}(\mathbf{X})$ is the horizontal size of the image. This is then used in the final loss function as:

$$L_{phys2} = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{X} \left(1 - \frac{1}{e^{-2 \cdot d_{\mathbf{X}_i}}} \right) \right) \tag{12}$$

This approach to penalization, which is informed by the physical properties of the system under study, contributes to the robustness and accuracy of the model's performance. This enhances the model's ability to generate meaningful and reliable outputs. This condition could have been also fulfilled by a hard constraint where this area would not be generated by the model. However, the training data shows that some electrons do seldom appear in this region due to specific interactions with the magnetic field and therefore their distribution is also worthwhile to simulate.

5 Evaluation metrics

We implemented a targeted cross-validation approach to evaluate our models, wherein each model underwent five training iterations. For each iteration, one

experiment was withheld as a validation set. The five experiments were selected to maximize diversity in both experimental parameters and electron spectra distributions, ensuring robust evaluation across varied conditions.

Following the standard procedure for evaluating generative models we first used the Fréchet Inception Distance (FID) [38] metrics. Nevertheless, we argue that since the feature embeddings captured by the Inception model are designed for natural images they are consequently not suitable for evaluation of the nuanced physical properties present in electron spectra data.

Therefore, to address this limitation, we propose a more dataset specific evaluation approach based on the Wasserstein distance [39], which enables direct comparison of charge distributions across energy bins. For each validation experiment and its generated counterpart, we compute the one-dimensional spectra of charge versus electron energy according to Equations 8 and 9. Our methodology then calculates the Wasserstein distance between real and generated charge distributions for each energy bin, as formulated in Equation 13:

$$W_w(\text{val}, \text{gen}) = \frac{1}{N} \sum_{i=1}^{N} w_i \cdot W_1(P_{\text{val}_i}, P_{\text{gen}_i}) \tag{13}$$

where W_w represents the weighted Wasserstein metric, N is the total number of energy bins, w_i is the weight assigned to the i-th energy bin, and $W_1(P_{\text{val}_i}, P_{\text{gen}_i})$ is the Wasserstein distance between the probability distributions of charge in the i-th energy bin for validation and generated samples, respectively.

This formulation offers significant flexibility for physics-specific evaluation. By adjusting the weight vector $\{w_i\}_{i=1}^{N}$, researchers can emphasize particular regions of interest in the energy spectrum. For instance, setting $w_i = 0$ for low-energy bins and $w_i = 1$ for high-energy bins enables focused evaluation of peak acceleration performance, which is often critical in electron acceleration experiments. The validation metric for each experiment is computed as the weighted mean across all bins, providing a comprehensive yet customizable assessment of model performance.

6 Results

The evaluation process was performed for a vanilla DDIM training approach and for the PCDDIM (physically constrained DDIM) introduced in this paper. The only difference between these two models and their training and sampling regime is the presence of the physical constraint loss in the PCDDIM. All of the training and validation models can be accessed from this link. The results can be seen in Table 1 over various classifier-free guidance [40] values. The results are averaged across various step counts ranging from 10 to 30 totaling 10 different runs per model and CFG value. The Wasserstein distance metric as described in Section 5 was used for energy bins corresponding to 30 MeV or higher, disregarding the lower energy electrons and focusing more on peak energy. Across both models and metrics we can see a decrease in sampling quality as CFG increases. This is attributed to the recent findings in [41].

Table 1: Comparison of Wasserstein distance and FID metrics for DDIM and PCDDIM models across different CFG values.

Metric	Model	CFG Values							
		1	2	3	4	5	6	7	8
Wasserstein	DDIM	**5.43**	9.56	13.95	16.34	19.33	21.46	22.13	23.47
	PCDDIM	17.98	16.94	17.17	19.38	21.01	24.4	23.62	27.73
FID	DDIM	107.11	111.56	111.74	114.20	115.75	116.03	118.41	121.11
	PCDDIM	**97.12**	101.32	103.49	104.99	106.62	107.57	108.70	109.29

Furthermore, the increased accuracy of the PCDDIM can be seen in its ability to produce samples outside of the training data parameters range. From Figure 5 it is apparent that the PCDDIM suggests electrons are being accelerated at laser energies of 40 mJ and above which is well aligned with other LWFA experimental data even though such experiments were not included in the training dataset. The proposed method can be further tested and validated outside the scope of available experimental data by using Particle-in-Cell (PIC) simulation data and comparing them to the generated results. [42]

(a) DDIM (b) PCDDIM

Fig. 5: Comparison of average results between DDIM and PCDDIM when sampling at higher laser energies.

7 Conclusion

This paper presents a novel methodology that integrates physical principles into generative denoising diffusion models applied for the optimization of laser-plasma accelerators. The approach demonstrates notable improvements in modeling complex, high-dimensional distributions associated with laser wakefield electron acceleration experiments. By embedding physical constraints into the model's training process, PCDDIM not only improves the FID metric but also offers robust extrapolation capabilities beyond the range of training data. Additionally, due to the ability of the PCDDIM to operate outside of the training inputs,

the outputs from the model were recognized by the experts in the field as more credible overall.

The above-described approach can be used for the conditional generation of multidimensional spectral data under physics constraints in various fields. This approach is believed to accelerate scientific discoveries by augmenting the search space of the experiments. Combination of PCDDIM with agent-based or other type of optimization can be a step forward in the efficient experimental design. The PCDDIM can be adapted to be used in other applications modeling highly non-linear interactions and parameter dependencies. The physical constraints need to be tailored specifically for each application and used within the proposed training scheme.

A specific metric tailored for electron spectra simulation was developed by comparing the distributions of charge over energy levels between generated and real samples. This offers a flexible approach to evaluation as one can assign significance to parts of the spectra. This can be used in optimization tasks by utilizing sampling settings which more closely model desirable parts of the spectrum.

Acknowledgments. This work was supported by the Student Summer Research Program 2023 of FIT CTU in Prague. This work was supported by the National Science Foundation and Czech Science Foundation under NSF-GACR collaborative Grant No. 2206059 from the NSF and Czech Science Foundation Grant No. 22-42963L.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. CROITORU, Florinel-Alin; HONDRU, Vlad; IONESCU, Radu Tudor; SHAH, Mubarak. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023.
2. GORI, Marco; BETTI, Alessandro; MELACCI, Stefano. *Machine Learning: A constraint-based approach*. Elsevier, 2023.
3. KARNIADAKIS, George Em; KEVREKIDIS, Ioannis G; LU, Lu; PERDIKARIS, Paris; WANG, Sifan; YANG, Liu. Physics-informed machine learning. *Nature Reviews Physics*. 2021, vol. 3, no. 6, pp. 422–440.
4. SHU, Dule; LI, Zijie; FARIMANI, Amir Barati. A Physics-informed Diffusion Model for High-fidelity Flow Field Reconstruction. *Journal of Computational Physics*. 2022, vol. 478, p. 111972. ISSN 0021-9991. Available from DOI: 10.1016/j.jcp.2023.111972.
5. BASTEK, Jan-Hendrik; SUN, WaiChing; KOCHMANN, Dennis M. *Physics-Informed Diffusion Models*. arXiv, 2024. Available from DOI: 10.48550/ARXIV.2403.14404.
6. SESSLER, Andrew. *Engines of discovery: a century of particle accelerators*. World Scientific, 2014.
7. MYERS, Stephen. The large hadron collider 2008–2013. *International Journal of Modern Physics A*. 2013, vol. 28, no. 25, p. 1330035.

8. AABOUD, Morad; AAD, Georges; ABBOTT, Brad; ABELOOS, B; ABIDI, SH; ABOUZEID, OS; ABRAHAM, NL; ABRAMOWICZ, H; ABREU, H; ABULAITI, Y, et al. Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector. *Physical review D*. 2018, vol. 97, no. 7, p. 072003.
9. GAILLARD, Mary K; GRANNIS, Paul D; SCIULLI, Frank J. The standard model of particle physics. *Reviews of Modern Physics*. 1999, vol. 71, no. 2, S96.
10. TAJIMA, T.; DAWSON, J. M. Laser Electron Accelerator. *Physical Review Letters*. 1979, vol. 43, no. 4, pp. 267–270. ISSN 0031-9007. Available from DOI: 10.1103/physrevlett.43.267.
11. SPRANGLE, P; ESAREY, E; TING, A; JOYCE, G. Laser wakefield acceleration and relativistic optical guiding. *Applied Physics Letters*. 1988, vol. 53, no. 22, pp. 2146–2148.
12. GIBNEY, Elizabeth. How the revamped Large Hadron Collider will hunt for new physics. *Nature*. 2022, vol. 605, no. 7911, pp. 604–607. ISSN 1476-4687. Available from DOI: 10.1038/d41586-022-01388-6.
13. JALAS, Sören; KIRCHEN, Manuel; MESSNER, Philipp; WINKLER, Paul; HÜBNER, Lars; DIRKWINKEL, Julian; SCHNEPP, Matthias; LEHE, Remi; MAIER, Andreas R. Bayesian Optimization of a Laser-Plasma Accelerator. *Physical Review Letters*. 2021, vol. 126, no. 10, p. 104801. ISSN 1079-7114. Available from DOI: 10.1103/physrevlett.126.104801.
14. SHALLOO, R. J.; DANN, S. J. D.; GRUSE, J.-N.; UNDERWOOD, C. I. D.; ANTOINE, A. F.; ARRAN, C.; BACKHOUSE, M.; BAIRD, C. D.; BALCAZAR, M. D.; BOURGEOIS, N.; CARDARELLI, J. A.; HATFIELD, P.; KANG, J.; KRUSHELNICK, K.; MANGLES, S. P. D.; MURPHY, C. D.; LU, N.; OSTERHOFF, J.; PÕDER, K.; RAJEEV, P. P.; RIDGERS, C. P.; ROZARIO, S.; SELWOOD, M. P.; SHAHANI, A. J.; SYMES, D. R.; THOMAS, A. G. R.; THORNTON, C.; NAJMUDIN, Z.; STREETER, M. J. V. Automation and control of laser wakefield accelerators using Bayesian optimization. *Nature Communications*. 2020, vol. 11, no. 1. ISSN 2041-1723. Available from DOI: 10.1038/s41467-020-20245-6.
15. BRÜMMER, Theresa; DEBUS, Alexander; PAUSCH, Richard; OSTERHOFF, Jens; GRÜNER, Florian. Design study for a compact laser-driven source for medical x-ray fluorescence imaging. *Physical Review Accelerators and Beams*. 2020, vol. 23, no. 3, p. 031601. ISSN 2469-9888. Available from DOI: 10.1103/physrevaccelbeams.23.031601.
16. COLE, J. M.; WOOD, J. C.; LOPES, N. C.; PODER, K.; ABEL, R. L.; ALATABI, S.; BRYANT, J. S. J.; JIN, A.; KNEIP, S.; MECSEKI, K.; SYMES, D. R.; MANGLES, S. P. D.; NAJMUDIN, Z. Laser-wakefield accelerators as hard x-ray sources for 3D medical imaging of human bone. *Scientific Reports*. 2015, vol. 5, no. 1. ISSN 2045-2322. Available from DOI: 10.1038/srep13244.
17. CHIU, Charles; FOMYTSKYI, Mykhailo; GRIGSBY, Franklin; RAISCHEL, Frank; DOWNER, Michael C.; TAJIMA, Toshiki. Laser electron acceler-

ators for radiation medicine: A feasibility study. *Medical Physics*. 2004, vol. 31, no. 7, pp. 2042–2052. ISSN 2473-4209. Available from DOI: 10.1118/1.1739301.

18. GIULIETTI, Antonio. *Laser-Driven Particle Acceleration Towards Radiobiology and Medicine*. Ed. by GIULIETTI, Antonio. Springer International Publishing, 2016. ISBN 9783319315638. ISSN 2197-5647. Available from DOI: 10.1007/978-3-319-31563-8.
19. NICKS, B. S.; TAJIMA, T.; ROA, D.; NEČAS, A.; MOUROU, G. Laser-wakefield application to oncology. *International Journal of Modern Physics A*. 2019, vol. 34, no. 34, p. 1943016. ISSN 1793-656X. Available from DOI: 10.1142/s0217751x19430164.
20. DÖPP, Andreas; EBERLE, Christoph; HOWARD, Sunny; IRSHAD, Faran; LIN, Jinpu; STREETER, Matthew. Data-driven Science and Machine Learning Methods in Laser-Plasma Physics. *High Power Laser Science and Engineering (2023)*. 2022. Available from DOI: 10.48550/ARXIV.2212.00026.
21. MEHTA, Viraj; PARIA, Biswajit; SCHNEIDER, Jeff; ERMON, Stefano; NEISWANGER, Willie. An experimental design perspective on model-based reinforcement learning. *arXiv preprint arXiv:2112.05244*. 2021.
22. GREENHILL, Stewart; RANA, Santu; GUPTA, Sunil; VELLANKI, Pratibha; VENKATESH, Svetha. Bayesian optimization for adaptive experimental design: A review. *IEEE access*. 2020, vol. 8, pp. 13937–13948.
23. LAZZARINI, C. M.; GRITTANI, G. M.; VALENTA, P.; ZYMAK, I.; ANTIPENKOV, R.; CHAULAGAIN, U.; GONCALVES, L. V. N.; GRENFELL, A.; LAMAČ, M.; LORENZ, S.; NEVRKLA, M.; ŠPAČEK, A.; ŠOBR, V.; SZUBA, W.; BAKULE, P.; KORN, G.; BULANOV, S. V. Ultrarelativistic electron beams accelerated by terawatt scalable kHz laser. *Physics of Plasmas*. 2024, vol. 31, no. 3. ISSN 1089-7674. Available from DOI: 10.1063/5.0189051.
24. TAJIMA, T.; YAN, X. Q.; EBISUZAKI, T. Wakefield acceleration. *Reviews of Modern Plasma Physics*. 2020, vol. 4, no. 1. ISSN 2367-3192. Available from DOI: 10.1007/s41614-020-0043-z.
25. ZYMAK, Illia; FAVETTA, Marco; GRITTANI, Gabriele Maria; LAZZARINI, Carlo Maria; TASSIELLI, Gianfranco; GRENFELL, Annika; GONCALVES, Leonardo; LORENZ, Sebastian; SLUKOVÁ, Vanda; VITHA, Filip; VERSACI, Roberto; CHACON-GOLCHER, Edwin; NEVRKLA, Michal; ŠIŠMA, Jiří; ANTIPENKOV, Roman; ŠOBR, Václav; SZUBA, Wojciech; STAUFER, Theresa; GRÜNER, Florian; LAPADULA, Loredana; RANIERI, Ezio; PIOMBINO, Michele; HAFZ, Nasr A. M.; KAMPERIDIS, Christos; PAPP, Daniel; MONDAL, Sudipta; BAKULE, Pavel; BULANOV, Sergei V. Characterization of kHz Repetition Rate Laser-Driven Electron Beams by an Inhomogeneous Field Dipole Magnet Spectrometer. *Photonics*. 2024, vol. 11, no. 12, p. 1208. ISSN 2304-6732. Available from DOI: 10.3390/photonics11121208.
26. HOUGH, Paul V C. *Method and means for recognizing complex patterns*. Inventor: Paul V C HOUGH. Publ.: 1962. US3069654A. Available also from: https://patents.google.com/patent/US3069654A/en.

27. BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. 2000.
28. BASLER AG. *Gain - Basler Product Documentation* [https://docs.baslerweb.com/gain]. 2023. Accessed: 2024.
29. SONG, Jiaming; MENG, Chenlin; ERMON, Stefano. Denoising Diffusion Implicit Models. 2020. Available from DOI: 10.48550/ARXIV.2010.02502.
30. NICHOL, Alex; DHARIWAL, Prafulla. Improved Denoising Diffusion Probabilistic Models. 2021. Available from DOI: 10.48550/ARXIV.2102.09672.
31. RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. Available from DOI: 10.48550/ARXIV.1505.04597.
32. NWANKPA, Chigozie; IJOMAH, Winifred; GACHAGAN, Anthony; MARSHALL, Stephen. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*. 2018.
33. VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Lukasz; POLOSUKHIN, Illia. Attention Is All You Need. 2017. Available from DOI: 10.48550/ARXIV.1706.03762.
34. HENDRYCKS, Dan; GIMPEL, Kevin. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*. 2016.
35. SHIN, Hoo-Chang; ROTH, Holger R; GAO, Mingchen; LU, Le; XU, Ziyue; NOGUES, Isabella; YAO, Jianhua; MOLLURA, Daniel; SUMMERS, Ronald M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*. 2016, vol. 35, no. 5, pp. 1285–1298.
36. KARRAS, Tero; AITTALA, Miika; AILA, Timo; LAINE, Samuli. Elucidating the Design Space of Diffusion-Based Generative Models. 2022. Available from DOI: 10.48550/ARXIV.2206.00364.
37. PREECHAKUL, Konpat; CHATTHEE, Nattanat; WIZADWONGSA, Suttisak; SUWAJANAKORN, Supasorn. Diffusion Autoencoders: Toward a Meaningful and Decodable Representation. 2021. Available from DOI: 10.48550/ARXIV.2111.15640.
38. HEUSEL, Martin; RAMSAUER, Hubert; UNTERTHINER, Thomas; NESSLER, Bernhard; HOCHREITER, Sepp. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 2017. Available from DOI: 10.48550/ARXIV.1706.08500.
39. KANTOROVICH, L. V. Mathematical Methods of Organizing and Planning Production. *Management Science*. 1960, vol. 6, no. 4, pp. 366–422. ISSN 1526-5501. Available from DOI: 10.1287/mnsc.6.4.366.
40. HO, Jonathan; SALIMANS, Tim. Classifier-Free Diffusion Guidance. 2022. Available from DOI: 10.48550/ARXIV.2207.12598.
41. KYNKÄÄNNIEMI, Tuomas; AITTALA, Miika; KARRAS, Tero; LAINE, Samuli; AILA, Timo; LEHTINEN, Jaakko. Applying Guidance in a Limited

Interval Improves Sample and Distribution Quality in Diffusion Models. In: *Proc. NeurIPS*. 2024.

42. VALENTA, P.; ESIRKEPOV, T. Zh.; LUDWIG, J. D.; WILKS, S. C.; BULANOV, S. V. Bayesian optimization of electron energy from laser wakefield accelerator. 2025. Available from DOI: 10.48550/ARXIV.2501.06069.