

Latent diffusion can map beam loss to two-dimensional phase-space projections

Alexander Scheinker^{*} and Alan Williams

Applied Electrodynamics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA



(Received 1 March 2025; accepted 29 August 2025; published 22 September 2025)

Beam loss monitors (BLMs) and beam current monitors (BCMs) are ubiquitous at particle accelerators around the world. These simple devices provide noninvasive high-level beam measurements but give no insight into the detailed 6D (x, y, z, p_x, p_y, p_z) beam phase-space distributions or dynamics. We show that generative conditional latent diffusion models can learn intricate patterns to solve the extreme inverse problem of mapping waveforms of tens of BLMs or BCMs along an accelerator to detailed 2D projections of a charged particle beam's 6D phase-space density. This transformational method can be used at any particle accelerator to transform simple noninvasive devices into detailed beam phase-space diagnostics. We demonstrate this concept via multiparticle simulations of the high-intensity beam in the kilometer-long Los Alamos Neutron Science Center linear proton accelerator.

DOI: [10.1103/rqg9-g3dp](https://doi.org/10.1103/rqg9-g3dp)

I. INTRODUCTION

Control of the 6D (x, y, z, p_x, p_y, p_z) phase-space distribution of beams is important for all particle accelerators over a wide range of applications including beam-based imaging for material science, accelerator-based light sources, and plasma wakefield-based accelerators.

At all large high-intensity beam particle accelerators, such as the Los Alamos Neutron Science Center (LANSCE) [1] or the Spallation Neutron Source (SNS) [2,3], the beam's phase distribution must be controlled for proper acceleration and to prevent beam loss by matching to the accelerator's magnetic focusing lattice. At particle accelerator-based free electron laser (FEL) light sources, such as the Linac Coherent Light Source [4], the European X-ray FEL [5,6], and the SwissFEL [7–9], the beam's phase space defines the properties of the generated light and must be adjusted between different experiments. For plasma wakefield acceleration (PWA), such as the Advanced Proton Driven Plasma Wakefield Acceleration Experiment (AWAKE) at a CERN [10] or the Facility for Advanced Accelerator Experimental Tests (FACET and FACET-II) at SLAC [11,12], the details of the beam's 6D phase-space density influence the PWA process. For laser-driven PWA [13], controlling the quality of the created beam's 6D phase space is challenging. Nondestructive

views of a beam's 6D phase space for PWA has the potential to help in the understanding and optimization of complex PWA processes.

When detailed beam phase-space measurements are available, advanced control theory-based adaptive machine learning (ML) techniques have been developed to quickly automatically shape the beam's phase space [14]. For noninvasive beam measurements, beam loss monitors (BLMs) and beam current monitors (BCMs) are ubiquitous, providing a measure of beam quality but give no insight into a beam's detailed 6D phase-space density. They simply measure 1D signals proportional to the total number of particles at a given accelerator location [15].

Recently, the world's first 6D beam measurement was conducted at the SNS Beam Test Facility, but this required 32 h (now reduced to 18 h) and 5,675,740 measurements [16]. Recent developments in high-dimensional beam diagnostics include 4D phase-space tomography from 1D measurements [17], 4D tomographic transverse phase-space reconstructions [18], 4D reconstructions using differentiable simulations [19], 5D tomographic phase-space reconstructions [20], single shot 4D core reconstructions [21], and ND maximum-entropy tomography via particle sampling [22]. An innovative tomographic algorithm has also been developed, based on the measurement data of the accumulated beam profile obtained from the wall current monitor in a synchrotron, to reconstruct the longitudinal phase space of the injected beam [23]. Efforts to develop ML-based virtual beam diagnostics including adaptive latent space tuning of generative autoencoders for tracking time-varying beams [24,25], combining generative ML models with physics models [26–28], latent evolution models [29], and normalizing flows [30].

^{*}Contact author: ascheink@lanl.gov

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

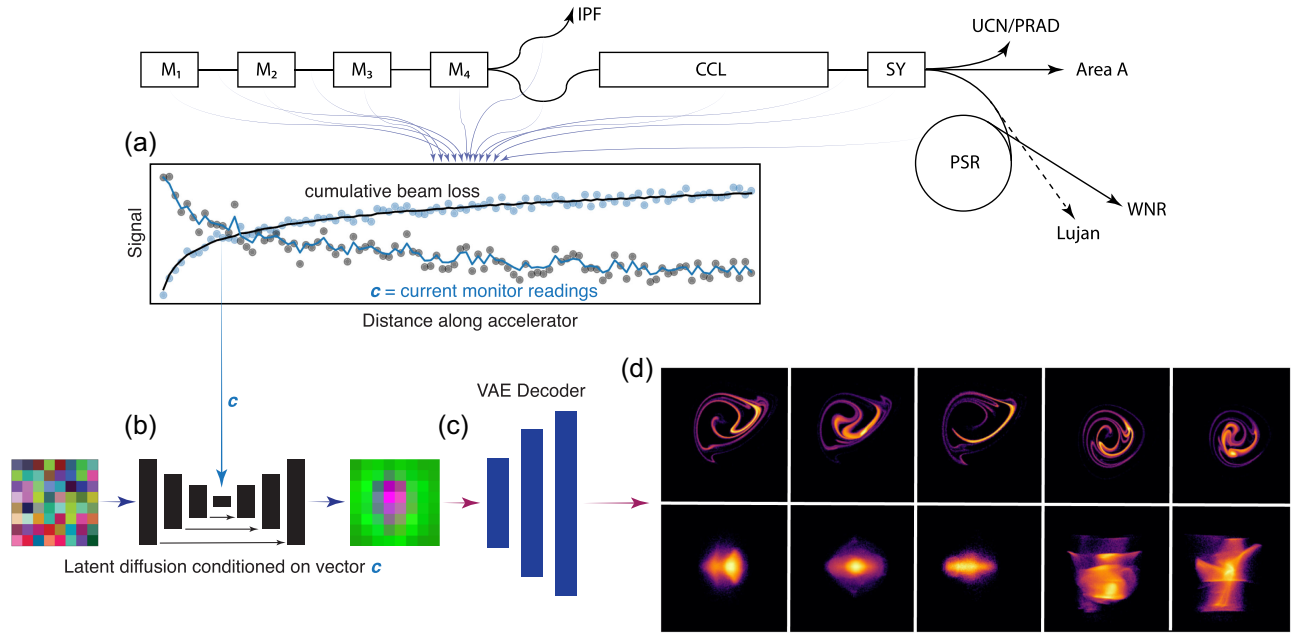


FIG. 1. Overview of the latent diffusion approach for mapping beam current or loss monitor readings to 2D phase-space projections. (a) Noninvasive beam current or loss monitor readings are recorded along the length of the accelerator. (b) The vector $\mathbf{c} = (c_1, \dots, c_n)$ of beam current at various accelerator locations together with accelerator parameters is used to condition the diffusion process. The conditional diffusion process generates a latent embedding representation of the beam's 2D phase-space projection. (c) The decoder of a variational autoencoder generates the desired 2D phase-space projection from its latent representation. (d) Various examples of 2D beam projections are shown. In this setup, we can generate any one of the 15 unique 2D projections of the beam's phase space and here various examples of the (z, E) and (x, y) projections are shown.

In the ML/AI community, diffusion-based models are the state-of-the-art for generating accurate representations of high-dimensional complex data [31–36]. Recently, the first application of diffusion was demonstrated as a virtual diagnostic to generate megapixel resolution virtual views of the (z, E) longitudinal phase space of the electron beam in the European x-ray FEL [37]. That method was then generalized to an adaptively tuned conditional diffusion approach for generating high-resolution representations of all the unique 2D projections of the HiRES UED beam's 6D phase-space distribution [38]. A physics-constrained superresolution diffusion-based method has been developed in which a beam's entire 6D phase-space density is generated as a 6D tensor from which 2D views are projected [39].

In this paper, we demonstrate that conditional generative latent diffusion can find subtle patterns in waveforms BLM or BCM measurements along an accelerator to solve an extreme inverse problem to map these vectors to highly detailed representations of the various 2D projections of a beam's 6D phase-space distribution, as shown in Fig. 1. This transformational result provides a method for transforming simple noninvasive BLM and BCM devices into detailed beam phase-space diagnostics. This same approach can be applied at any particle, completely transforming the amount of noninvasive beam diagnostics information that is available for tuning and setup. As a

proof of principal, we demonstrate this generative AI concept via multiparticle simulations of high-intensity beams in the kilometer-long LANSCE linear proton accelerator as shown in Fig. 1.

Figure 2 shows an overview of an inductive BCM whose operation is described by the Maxwell-Faraday equation

$$\oint_{\partial \Sigma} \mathbf{E} \cdot d\mathbf{l} = - \int_{\Sigma} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{A}, \quad \mathbf{B}(r, t) = \frac{\mu_0 I(t)}{2\pi r} \hat{\theta}, \quad (1)$$

where $I(t)$ is beam current, r is transverse distance from the beam center, $\mathbf{B}(r, t)$ is the beam's magnetic field, and $\mathbf{E}$ is the electric field induced by the beam passing through a

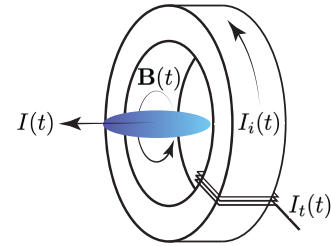


FIG. 2. In an inductive beam current monitor, the time-varying flux of a passing beam's magnetic field $\mathbf{B}(t)$ induces a current $I_i(t)$ in a conductor, which is measured as a current $I_r(t)$ proportional to the beam's current $I(t)$.

conducting loop. The induced electric field creates a current in the loop which is then coupled through a simple ac transformer to a current measurement pickup, giving a signal proportional to beam current.

II. DATA

For this work, we utilize the particle tracking code HPSim which includes 3D space charge and is implemented in CUDA to run extremely fast (few seconds for a million macroparticle simulation through the kilometer-long LANSCE accelerator with space charge) on a GPU [40]. The code simulates the transport of a charged particle beam from the source to the end of the coupled cavity linac section of the LANSCE accelerator through approximately 1 km as it traverses 48 resonant frequency (RF) accelerating cavities. The data are generated by setting an initial beam condition and then randomly perturbing the amplitude and phase of the first 4 RF cavities of LANSCE within uniform distributions over a range of $\pm 3^\circ$ and $\pm 3\%$ of the RF phase and amplitude set point, respectively, relative to a design tune of the RF system. Note that this is a significant change as operators typically adjust any one of these parameters at most $\pm 1^\circ$ or $\pm 1\%$ to maintain minimal beam loss while the system slowly varies with time due to factors such as temperature changes. For a fixed initial beam condition, 500 such simulations are run, each with 10^6 macroparticles, and for each simulation, 2D histograms are made for each of the unique 2D projections of the beam's 6D phase space at the exit of every single RF module over a 256×256 pixel grid. This process is repeated for 4 different initial conditions resulting in a dataset of 2000 simulations where each simulation results in 15 images of 256×256 pixels each at 48 module locations.

The final dataset is $2000 \times 48 \times 15 = 1.44$ million 256×256 pixel images. Of these simulations, from the 4 different runs having the same initial beam conditions, the training data consist of 400 simulations from each of the first 3 runs. The test set consists of 100 simulations from each of the first 3 runs and all 500 simulations of the 4th run. This split allows us to test the model's predictive ability relative to a fixed initial condition and varying RF parameters as well as a completely unseen initial condition together with varying RF parameters. The reasoning behind this dataset is that in practice the beam's initial conditions change on a slower timescale than the RF parameters. The LANSCE ion source is a complicated device, the H^- beam is created inside a high frequency (120 Hz), high duty factor (10%, 833 μ s) tungsten filament-plasma driven, cesiated surface-conversion, multicusp ion source. While operationally the source is well understood, there are many uncertainties known about many of its internal parameters such as a quantitative measure of cesium density and H^- beam density inside the source [41]. During operations, the filaments of the source gradually wear out via a complicated process, resulting in a regular source replacements

multiple times during each run cycle. Every time that the beam source is replaced, the beam's initial conditions experience a step change and then while that source is run and the beam's initial conditions are changing very slowly, the entire RF system's characteristics continuously drift (on timescales of minutes-hours), requiring continuous manual retuning by the operators and beam physicists.

At each module $m \in \{1, 2, \dots, 48\}$, the 15 unique beam projections of the beam's 6D phase-space density $\rho^m(x, y, z, p_x, p_y, p_z)$ are

$$\begin{aligned} &\rho_1^m(x, y), \rho_2^m(x, z), \rho_3^m(x, p_x), \rho_4^m(x, p_y), \rho_5^m(x, p_z) \\ &\rho_6^m(y, z), \rho_7^m(y, p_x), \rho_8^m(y, p_y), \rho_9^m(y, p_z) \\ &\rho_{10}^m(z, p_x), \rho_{11}^m(z, p_y), \rho_{12}^m(z, p_z) \cdot \\ &\rho_{13}^m(p_x, p_y), \rho_{14}^m(p_x, p_z) \\ &\rho_{15}^m(p_y, p_z) \end{aligned}$$

We calculate a number proportional to the surviving beam current at each module by simply summing over the $\rho_{12}^m(z, p_z)$ projection

$$b_m = \int_z \int_{p_z} \rho^m(z, p_z) dp_z dz. \quad (2)$$

Note that any projection can be used, they should all give identical answers. In practice, this one was chosen because the projection bounds were carefully selected along the modules to capture as much beam as possible for a wide range of RF settings while not making the bounds too wide so that details of the beam would be missed. Because of this, the edges of the beam are sometimes clipped by the size of the projection window, but this is rare and has negligible effects on the overall analysis.

In high-energy particle accelerators, the beam's longitudinal momentum typically dominates the other components with $p_z \gg p_x, p_y$, therefore instead of p_x and p_y , we typically consider the quantities $x' = p_x/p_z$ and $y' = p_y/p_z$ which provide a measure of the beam's divergence from straight line motion. Furthermore, because of the resonant RF structures used to accelerate the beam, instead of considering the global z position of particles in the accelerator, it is also common to consider the phase φ of a particle in a bunch relative to a design particle, where φ is the phase shift of the RF accelerating field. In particular, for a given RF cavity amplitude set point A_s and phase set point θ_s , HPSim considers a design particle that arrives when the RF field's phase is at θ_s so that the electric field in the cavity is given by

$$E = A_s \sin(\omega t + \theta_s), \quad (3)$$

then the "phase" of any particle in the bunch is defined by considering the distance Δz between the particle and the design particle, considering the particle velocity v at the

entrance to the RF cavity, and calculating the amount of phase shift of the RF field between the two particles according to

$$\Delta\theta = \omega\Delta t = \omega \frac{v}{\Delta z}, \quad (4)$$

and assigning a phase $\theta = \theta_s + \Delta\theta$. For the 2D projection of the bunch at the RF cavity, the absolute phase θ_s is not important, and so we consider just the relative phase $\varphi = \Delta\theta$ to get the total phase spread of the bunch. One final change is that we typically consider the total kinetic energy of the beam, E , assuming that it is almost entirely due to the p_z momentum component.

With these new phase-space coordinates, the 15 projections that we typically care about are the 15 unique beam projections of the beam's 6D phase-space density $\rho^m(x, y, z, x', y', E)$ given by

$$\begin{aligned} &\rho_1^m(x, y)\rho_2^m(x, \varphi)\rho_3^m(x, x')\rho_4^m(x, y')\rho_5^m(x, E) \\ &\rho_6^m(y, \varphi)\rho_7^m(y, x')\rho_8^m(y, y')\rho_9^m(y, E) \\ &\rho_{10}^m(\varphi, x')\rho_{11}^m(\varphi, y')\rho_{12}^m(\varphi, E) \\ &\rho_{13}^m(x', y')\rho_{14}^m(x', E) \\ &\rho_{15}^m(y', E) \end{aligned}$$

represented as a 256×256 pixel images.

The importance of specific 2D projections depends on position within the accelerator. Near the front end, the Los Alamos Neutron Science Center (LANSCE) linear proton accelerator uses 100 MeV proton beams to create medical isotopes, for which the transverse beam size σ_r , average

beam current, and beam energy are important as the beam is impacted onto a target for which the beam's kinetic energy controls the nuclear process of isotope creation and the beam must be rastered across the sample based on its size and current to prevent overheating [42]. Near the end of LANSCE, an 800 MeV proton beam is impacted onto a tungsten target producing a high-neutron flux for neutron-diffraction material-science studies, for which only the average beam energy is important. At the end of LANSCE, proton radiography requires a precisely controlled longitudinal phase-space (z, p_z) distribution of beam energy and bunch length σ_z for beam-based dynamic imaging [43]. Throughout the LANSCE accelerator, the beam's transverse (x, y) and (p_x, p_y) distributions must be controlled to keep the beam matched to the quadrupole magnet-based focusing lattice so that the beam's space charge does not rip it apart causing it to impact the accelerator's beam pipe.

III. VARIATIONAL AUTOENCODER

Our eventual goal is to create a map from RF settings and beam current measurements to the various 2D phase-space projections described above via a conditional guided latent diffusion approach. The first step of latent diffusion is to compress a large 256×256 pixel image ρ_j^m down to a much lower dimensional latent representation z_j^m to speed up the diffusion process that will then work directly in the latent space, as shown in Fig. 3. Besides the reduction in computational complexity, by working with latent representations z_j^m , the diffusion model is able to focus on abstract features that have been learned by the VAE, which keeps only the most essential details, rather than the raw

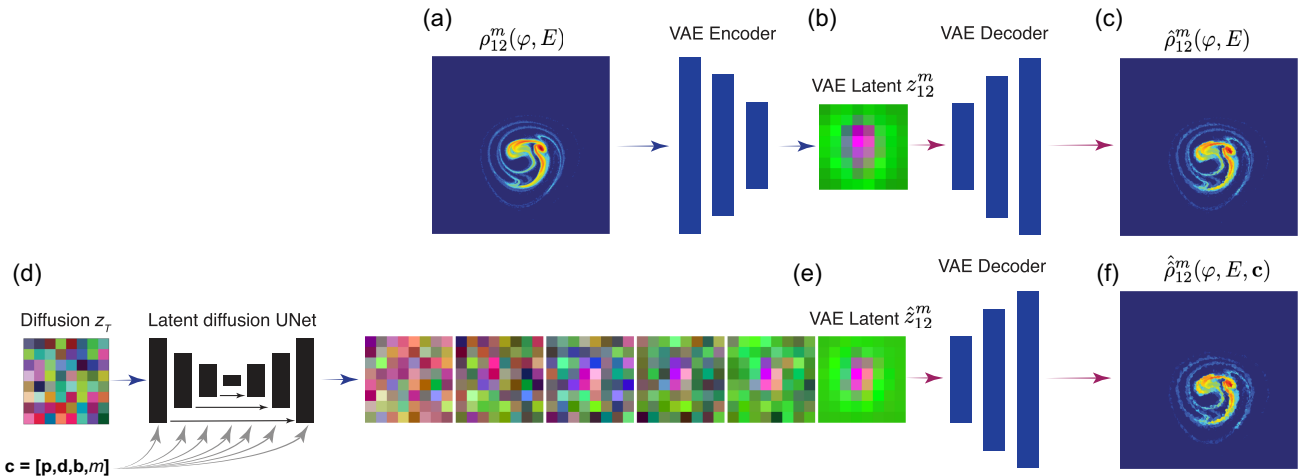


FIG. 3. Combining a variational autoencoder (VAE) with a diffusion model for latent diffusion. (a) The VAE is used to compress a projection $\rho_j^m \in \mathbb{R}^{256 \times 256}$ down to a much lower dimensional latent embedding $z_j^m \in \mathbb{R}^{8 \times 8 \times 3}$ (b) from which an estimate of the image is then reconstructed as $\hat{\rho}_j^m \in \mathbb{R}^{256 \times 256}$ (c). (d) A latent diffusion model, guided by a conditional vector $\mathbf{c}$ containing beam current measurements, is then trained to generate estimates of the latent embeddings $\hat{z}_j^m \in \mathbb{R}^{8 \times 8 \times 3}$ (e). (f) The generated latent embeddings $\hat{z}_j^m \in \mathbb{R}^{8 \times 8 \times 3}$ are then passed back into the decoder of the VAE resulting in conditional generation of the high-resolution phase-space projections $\hat{\hat{\rho}}_j^m \in \mathbb{R}^{256 \times 256}$.

pixel values. This is clear in the latent representations shown in Fig. 5 where the original phase-space images have large empty backgrounds of 0-value pixels while the latent representations are much more dense, using up the reduced pixel space in a more optimal way.

For this dimensionality reduction task, we first train a variational autoencoder (VAE) convolutional neural network which embeds the 2D phase-space projections $\rho_j^m \in \mathbb{R}^{256 \times 256}$ into latent representations $z_j^m \in \mathbb{R}^{8 \times 8 \times 3}$, from which estimates of the projections $\hat{\rho}_j^m \in \mathbb{R}^{256 \times 256}$ are then reconstructed, as shown in Fig. 3(a). By embedding a $256 \times 256 = 65536$ pixel image into a three-channel $8 \times 8 \times 3 = 192$ pixel image, the data are compressed by a factor of 341. Using additional channels allows the VAE's latent representation to take into account various image properties such as edges, as shown in Fig. 5. A high-level overview of the VAE is shown in Fig. 4, demonstrating how much smaller the latent embedding is for fast latent diffusion. The resolution decreasing layers shown in Fig. 4 each consist of a group of three 2D convolution layers with 3×3 filters and swish activation functions [44] followed by batch normalization [45]. The first group of 2D convolutions has 64 filters each. After each such set of three, the image resolution is reduced by a factor of 2 using a MaxPool 2D layer with a 2×2 pool size. This procedure is repeated until the images are reduced to 8×8 with the number of filters doubling at each stage. The network then splits into two branches, each of size $8 \times 8 \times 3$, with one defining a mean image and the other defining a log

variance. Together these define a $8 \times 8 \times 3$ normal distribution from which samples are passed through the decoder half of the VAE whose architecture mirrors that of the encoder, with MaxPool layers replaced by transposed 2D convolutions.

If we represent a 256×256 pixel input image as X , then the encoder function of the VAE, f_{en} maps X to mean z_μ and log variance z_Σ :

$$f_{\text{en}}: X \rightarrow [z_\mu(X), z_\Sigma(X)], \quad (5)$$

which are then used to define a normal distribution from which latent representations z are sampled

$$z(X) \sim \mathcal{N}(z_\mu(X) + \Sigma(X)\mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \Sigma(X) = \exp[z_\Sigma(X)/2].$$

The sampled latent representation is then passed through the decoder function, f_{de} , which attempts to regenerate X according to

$$f_{\text{de}}: z(X) \rightarrow \hat{X}[z(X)]. \quad (6)$$

The VAE's loss function has two main components, the first is the Kullback–Leibler (KL) divergence between the learned latent distribution $\mathcal{N}_\theta$ and a mean 0 unit variance normal distribution with diagonal covariance matrix according to

$$\text{KL}[\mathcal{N}_\theta, \mathcal{N}(\mathbf{0}, \mathbf{I})] = \frac{1}{2} \{ \text{Tr}[\Sigma] + z_\mu^T z_\mu - \log[\det(\Sigma)] \}. \quad (7)$$

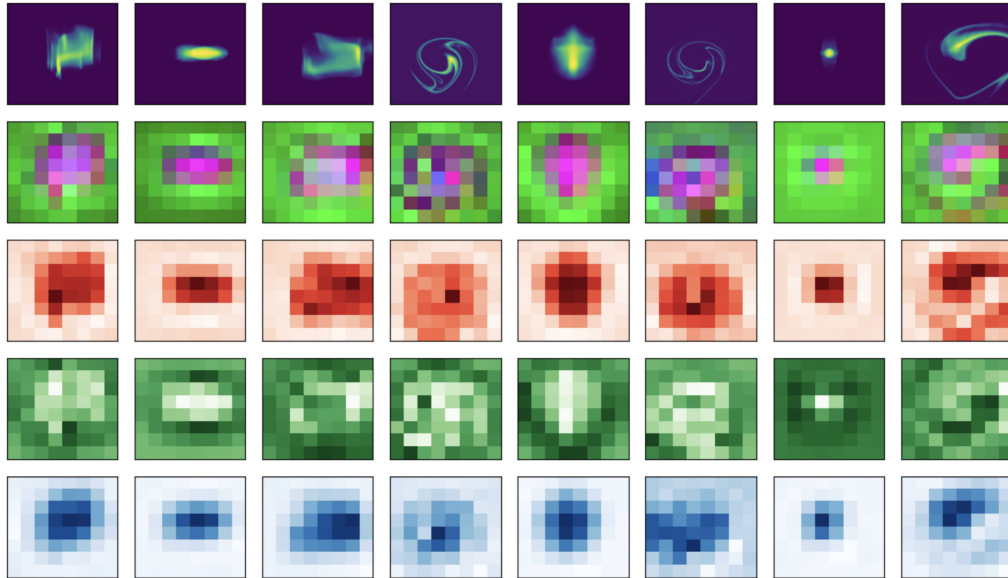


FIG. 4. The top row shows eight random examples of 256×256 pixel 2D phase-space projections of the beam. The second row shows the associated 8×8 pixel red, green, blue (RGB) three-channel latent embedding images of the VAE. The third through fifth rows show the separate RGB channels of the latent embeddings. The red channel seems to be activated by the brightest regions of the original image and its interior, the green channel is highly active at the image boundaries, and the blue channel looks like a down sampled 8×8 pixel version of the original image, uniformly covering the support of the image.

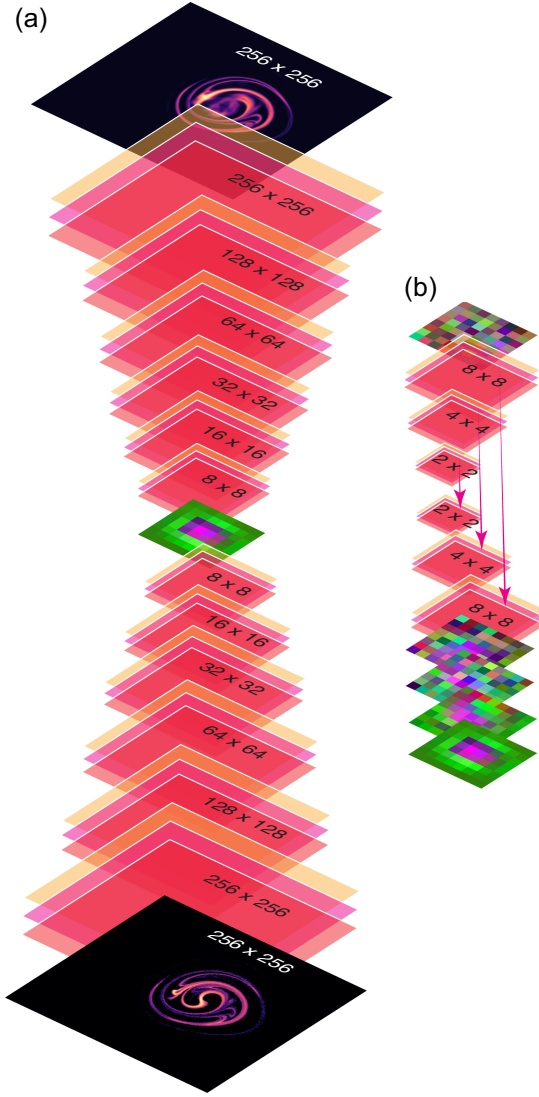


FIG. 5. (a) High-level overview of the VAE architecture (a) showing image size reduction from 256×256 pixels down to a $8 \times 8 \times 3$ latent embedding. (b) The much smaller diffusion UNet working directly in the latent embedding space.

This KL divergence regularizes the network's latent representation, pushing it toward a mean 0 unit variance distribution so that the latent space can be easily sampled and traversed without embeddings being placed in fragmented distant islands.

The second component of the loss function is the reconstruction loss defined as the negative log-likelihood of the regenerated images

$$R = -\log[Pr(X|z(X))], \quad (8)$$

which pushes the network toward accurately reconstructing the input images. The total VAE training loss is then defined as a weighted sum

$$C = R + wKL, \quad (9)$$

where the weight w is typically chosen to be < 1 to allow the network to develop some structure in the latent embedding.

Despite using a highly compressed representation, the VAE is able to reconstruct highly accurate, very detailed images, as shown in Figs. 6 and 7. The overall absolute percent error statistics for the training and test data of the VAE are shown in Fig. 8.

IV. CONDITIONAL GUIDED LATENT DIFFUSION

In what follows, we sometimes drop phase-space coordinate arguments and refer to the projections above as

$$\rho_j^m(\mathbf{p}, \rho^0) \in \{\rho_1^m(\mathbf{p}, \rho^0), \rho_2^m(\mathbf{p}, \rho^0), \dots, \rho_{15}^m(\mathbf{p}, \rho^0)\}, \quad (10)$$

to emphasize their dependence on the initial beam density ρ^0 and on the RF parameter settings

$$\mathbf{p} = [A_1, \theta_1, A_2, \theta_2, A_3, \theta_3, A_4, \theta_4], \quad (11)$$

which are the amplitude set points (A_i) and the phase set points (θ_i) of the first 4 RF modules which make up the 201.25 MHz drift tube linac section of the LANSCE accelerator (all other RF settings are fixed).

We focus on this section of the accelerator because the beam's characteristics are most sensitive to these RF modules, which have the greatest impact on the overall downstream performance. In practice, these are the first knobs that the operators tune in order to minimize beam loss or adjust other beam properties. For any given initial beam distribution ρ^0 and any RF settings $\mathbf{p}$, we test whether we can reconstruct the 2D beam phase-space projections from a vector of beam currents as measured at all 48 RF module locations

$$\mathbf{b} = [b_1, b_2, \dots, b_{48}], \quad (12)$$

where the individual beam current measurements are as defined above in Eq. (2). We assume that we do not have access to ρ^0 which in practice is a time-varying $\rho^0(t)$ which undergoes step changes with every source change. Our overall virtual diagnostic is then tasked with the mapping:

$$\mathbf{c} = [\mathbf{p}, \mathbf{b}, \mathbf{d}, m] \rightarrow \rho_j^m, \quad (13)$$

where $\mathbf{d}$ is a vector which defines which of the 15 projections we would like to generate.

As discussed in [38], a good physics-informed choice for $\mathbf{d}$ is to use a six-dimensional vector whose positions correspond to the six phase-space dimensions $(x, y, \varphi, x', y', E)$ so that, for example, when generating $\rho_1^m(x, y)$, then $\mathbf{d} = [1, 1, 0, 0, 0, 0]$ and when generating $\rho_{15}^m(y', E)$, then $\mathbf{d} = [0, 0, 0, 0, 1, 1]$.

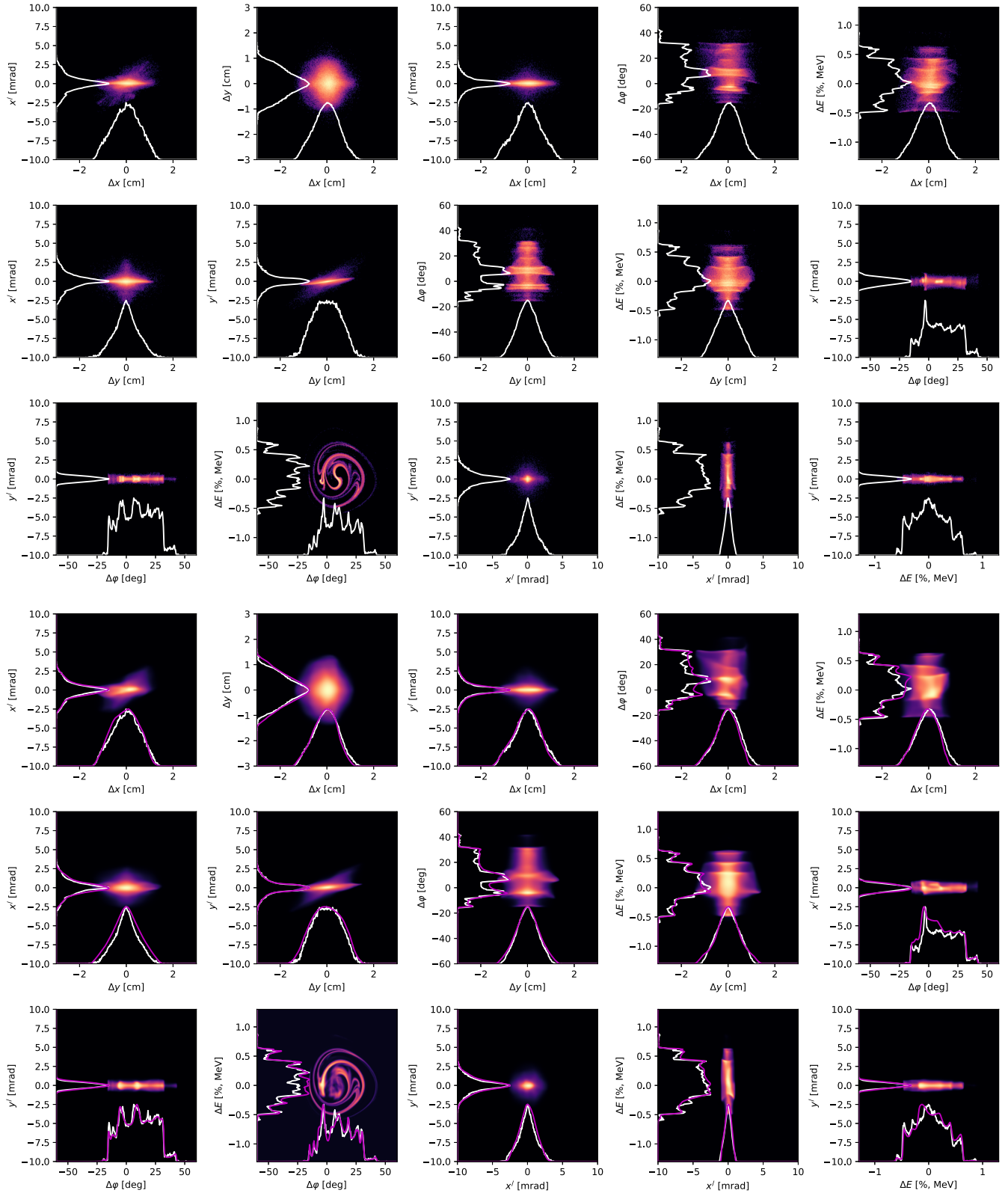


FIG. 6. The top 15 images show one example of the true phase-space projections at module 48 for a test data point and the bottom 15 images show the generated images. The white curves are projections of the true data on their respective axes and the magenta curves are those of the generated data.

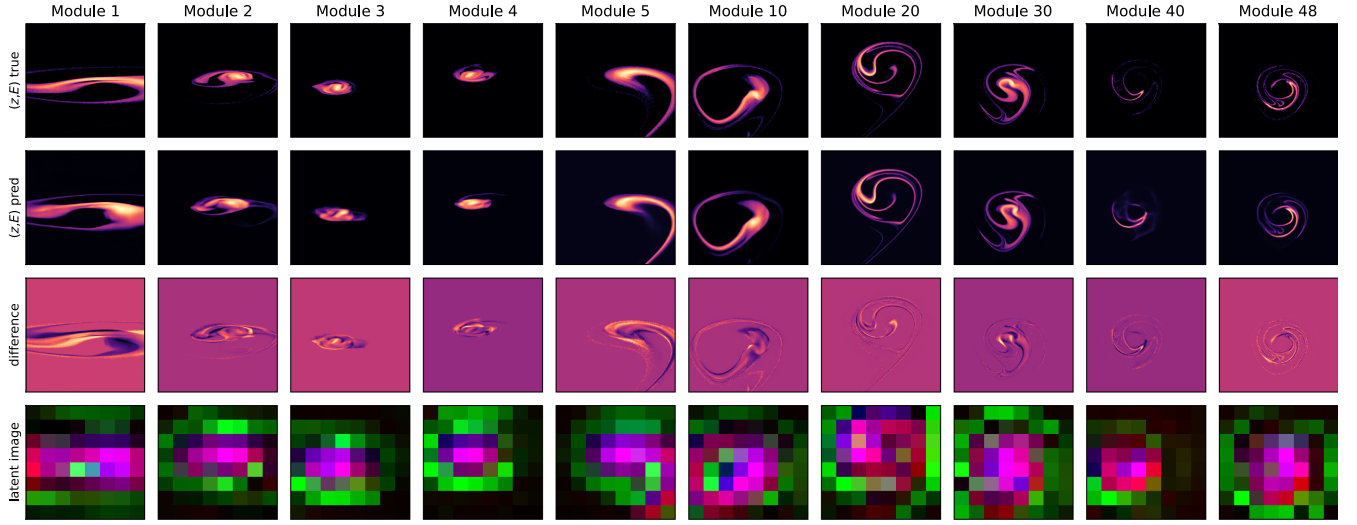


FIG. 7. The true and predicted $(\Delta\varphi, \Delta E)$ projections are shown at several modules along LANSCE along with the difference and the latent embeddings.

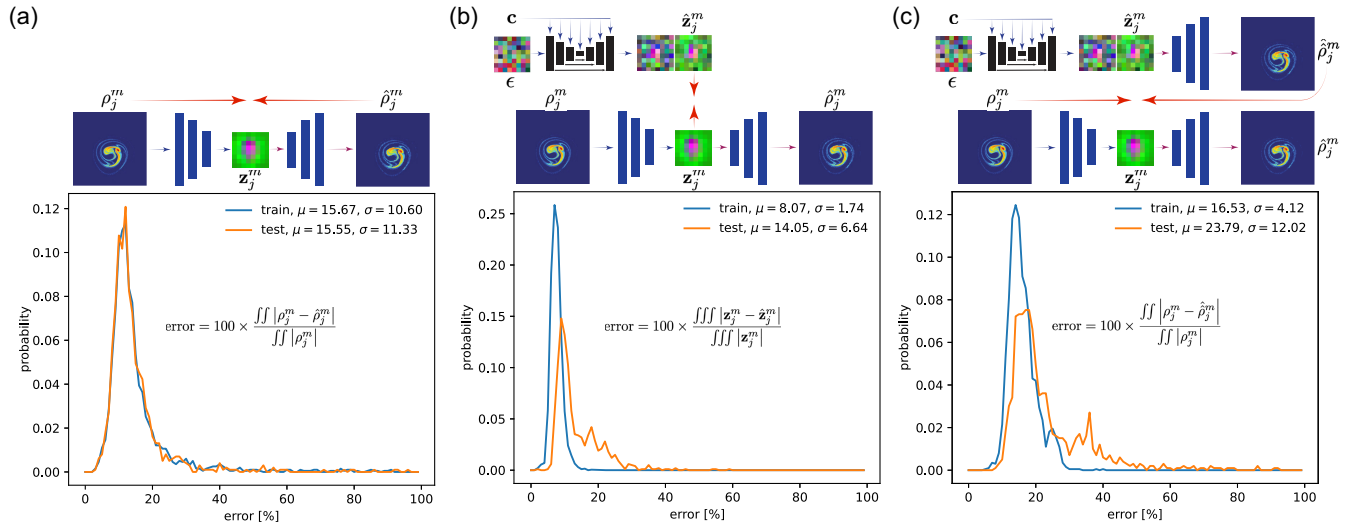


FIG. 8. Percent absolute error statistics for VAE reconstructions over the training and test data (a), for conditional diffusion reconstructions of the latent representations (b), and for the VAE decoder reconstructing the original images using the diffusion-based latent reconstructions (c).

This physics-informed method is found to work better than simply choosing a scalar $d \in \{1/15, 2/15, \dots, 1\}$, which arbitrarily places certain projections closer to each other than others. For example, such an approach would condition the $\rho_1^m(x, y)$ projection with $1/15$, the $\rho_2^m(x, \varphi)$ projection as $2/15$, and the $\rho_3^m(x, x')$ projection as $3/15$, thus placing $\rho_2^m(x, \varphi)$ twice as close as $\rho_3^m(x, x')$ to $\rho_1^m(x, y)$, although all three projections share just the single x axis. In contrast to this, for the physics-informed 6D vector approach conditions, the projections $\rho_1^m(x, y)$, $\rho_2^m(x, \varphi)$, and $\rho_3^m(x, x')$ with the vectors $\mathbf{d}_1 = [1, 1, 0, 0, 0, 0]$, $\mathbf{d}_2 = [1, 0, 1, 0, 0, 0]$, and $\mathbf{d}_3 = [1, 0, 0, 1, 0, 0]$, respectively, so that all of these projections that share the x axis have the same pairwise inner

products $\mathbf{d}_i \cdot \mathbf{d}_j = 1$. Furthermore, completely independent projections, such as $\rho_{12}^m(\varphi, E)$ represented by $\mathbf{d}_{12} = [0, 0, 1, 0, 0, 1]$ and $\rho_{13}^m(x', y')$ represented by $\mathbf{d}_{13} = [0, 0, 0, 1, 1, 0]$ are orthogonal with $\mathbf{d}_{12} \cdot \mathbf{d}_{13} = 0$. This method also outperforms one-hot encoding $\mathbf{d}$ as 15-dimensional unit basis vectors of the form $\mathbf{d}_1 = [1, 0, \dots, 0]$, $\mathbf{d}_2 = [0, 1, 0, \dots, 0]$, which forces each projection to be orthogonal, ignoring the shared axes.

Details of the diffusion process are available in the literature [31–36]. Here we provide only a quick overview. To generate latent images z , generative diffusion works by learning how to reverse a diffusion stochastic differential equation (SDE). One can approximate the SDE with finite differences, yielding an iterative (diffusion based) noising

process, converting z into mean zero, unit variance Gaussian noise:

$$\begin{aligned} z_1 &= \sqrt{1 - \beta_1}z + \sqrt{\beta_1}\epsilon_1 \\ z_2 &= \sqrt{1 - \beta_2}z_1 + \sqrt{\beta_2}\epsilon_2 \\ &\vdots \\ z_T &= \sqrt{1 - \beta_T}z_{T-1} + \sqrt{\beta_T}\epsilon_T, \quad \epsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \end{aligned} \quad (14)$$

The noise schedule $\{\beta_k, k \in 1, \dots, T\}$ is chosen as an increasing sequence such that $0 < \beta_{k-1} < \beta_k < 1$ with typical starting and ending values of $\beta_1 = 1 \times 10^{-4}$ and $\beta_T = 1 \times 10^{-2}$. For large high-resolution images (1024×1024), the sequence length is typically $T = 1000$, but for the much smaller latent images that we use for latent diffusion, we can work with $T = 100$. This is one additional factor that greatly speeds up latent diffusion relative to standard image space diffusion, besides working with a smaller UNet, we also need fewer diffusion steps.

For an image evolving according to Eq. (14), a sample z_k can be generated in a single shot from z without having to go through the Markov chain one step at a time as it is easily calculated that

$$z_k = \sqrt{\alpha_k}z + \sqrt{1 - \alpha_k}\epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (15)$$

where

$$\alpha_k = \prod_{j=1}^k (1 - \beta_j). \quad (16)$$

Equation (16) and $0 < \beta_j < 1$ implies that α_k converges toward 0 as k increases and therefore from Eq. (15), it is clear that z_k approaches a sample from $\mathcal{N}(\mathbf{0}, \mathbf{I})$. Equation (15) makes it easy to sample the state of an image z at an arbitrary diffusion time k . We can analytically approximate the conditional diffusion distribution of a reverse diffusion step of moving from z_k to z_{k-1} by a combination of Bayes' rule and a change of variables, as

$$q(z_{k-1}|z_k, z) = \mathcal{N}\left[\frac{1 - \alpha_{k-1}}{1 - \alpha_k} \sqrt{1 - \beta_k}z_k + \frac{\sqrt{\alpha_{k-1}\beta_k}}{1 - \alpha_k}z, \frac{(1 - \alpha_{k-1})\beta_k}{1 - \alpha_k}\mathbf{I}\right]. \quad (17)$$

Having this analytic form (and making several additional simplifying assumptions as described in the literature) allows us to design the diffusion UNet to perform iterative denoising steps by predicting and removing the noise that is added during the diffusion process D_k which we model using a UNet D_θ according to

$$D_k[z_t, \mathbf{c}] = \frac{1}{\sqrt{1 - \beta_k}}z_k - \frac{\beta_k}{\sqrt{1 - \alpha_k}\sqrt{1 - \beta_k}}D_\theta[z_t, t, \mathbf{c}]. \quad (18)$$

The overall architecture uses a U-net [46], with the setup that was developed for the PixelCNN [47], which uses group normalization instead of regular normalization [48], two residual CNN blocks at each resolution of 8×8 , 4×4 , and 2×2 , and attention layers at the 4×4 and 2×2 pixel embedding levels [49–54].

Reconstruction error statistics for both the training and test data are shown in Figs. 8–10. In Fig. 8, we can see a trade-off of the latent diffusion approach as the overall test error grows along the encode, generate, decode sequence. Although it is faster than diffusion in image space, in the latent diffusion approach, each generative model, and especially compression, adds its own error to the overall final result. In our case, and in most use cases, this trade-off is acceptable if a large/powerful enough autoencoder is used. In Fig. 9, we see a comparison between training and test data of the accuracy of the predicted first moments of the beam after it has been regenerated from its latent representation by the latent diffusion process. As expected, performance is worse for test data than for training data, but the results remain highly accurate, with σ errors much smaller than the σ values of the beams themselves. This is expected because here the diffusion model is solving a difficult ill-defined inverse problem of mapping a 1D distributed measurement back to detailed 2D phase-space projections. For such problems, although a validation error is used to stop training once overfitting is detected, the overall validation error is greater than the training error, those two curves do not overlap as they do in much more simple problems. This can be clearly seen in Fig. 8. While the VAE is solving the easier self-supervised task of compressing and reconstructing images, the train and test error distributions look almost identical (A). However, once we reach the diffusion process, which addresses the ill-defined inverse problem, even with early stopping of training to prevent overfitting, we immediately see a split in distributions with the test data error increasing (B). In Fig. 10, we see a more detailed breakdown of the test error as the dependence of MSE test error is shown relative to module number and projection number. As expected, we see that the method's performance is related to the variance of the data (when variance is low, a model can easily get away with just learning the mean).

V. ROBUSTNESS TO NOISE

For a study of how feasible this approach is, it is important to test the robustness of the approach to measurement noise. In simple BCMs, such as the design shown in Fig. 2, the induced current $I_i(t)$ in the body of the BCM itself excites a measured current $I_r(t)$ through a transformerlike coupling. The actual measured current $I_r(t)$

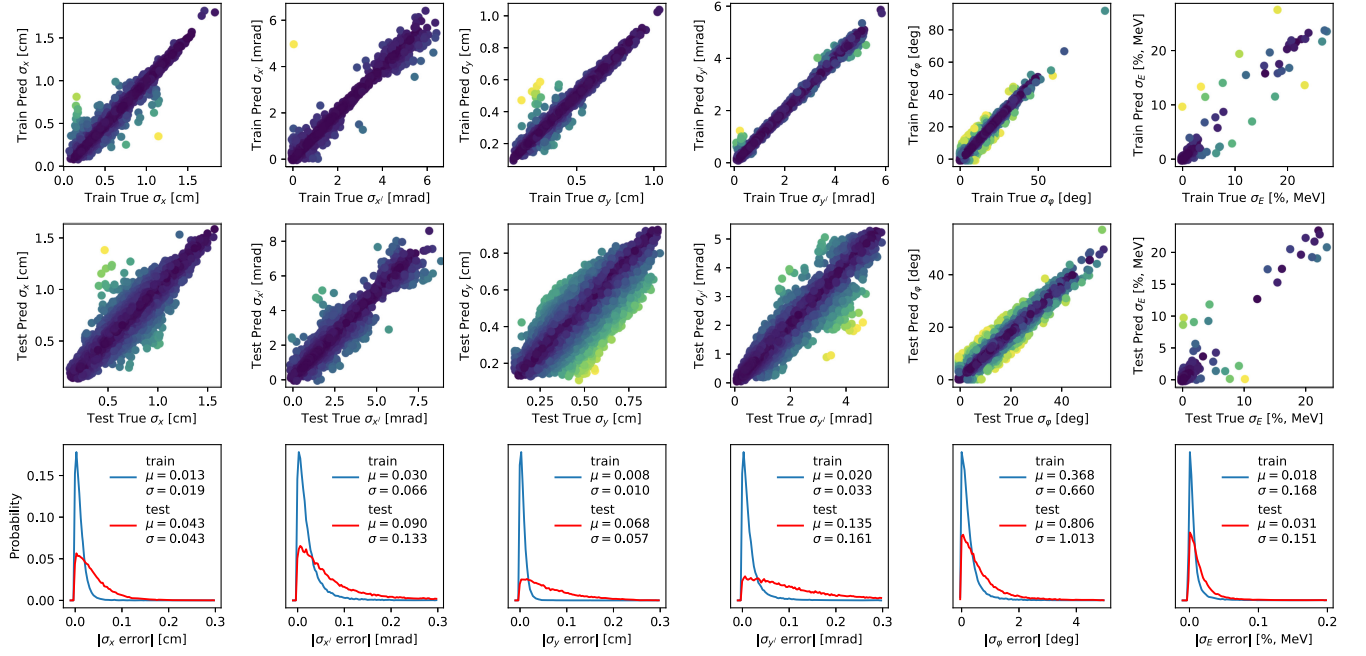


FIG. 9. Error statistics calculated by fitting Gaussians to each 1D projection of each image and then comparing the σ values for both training and test data predictions. The top row shows true and predicted results for the training data, the middle row shows results for test data, and the bottom row shows histograms of the absolute errors.

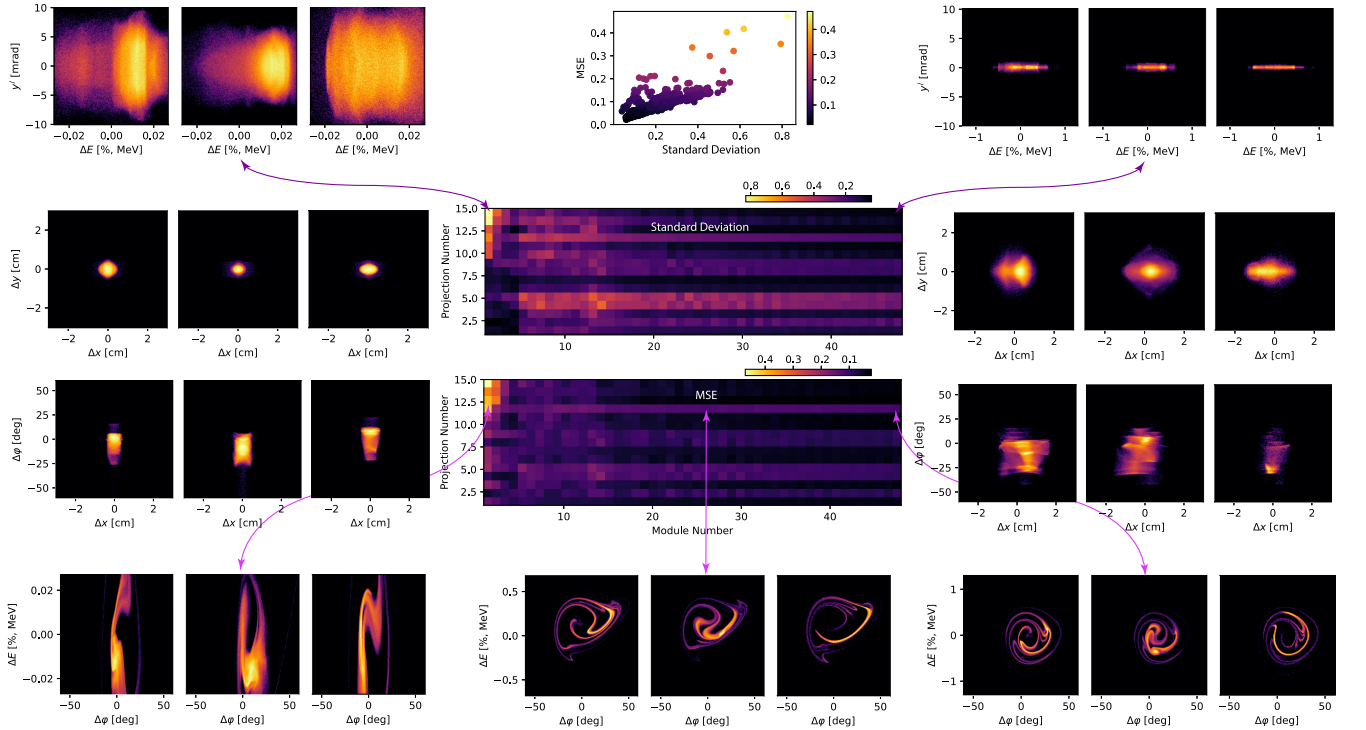


FIG. 10. Prediction error compared to variation of the data as a function of location in LANSCE (module number) and projection number. The top three figures in the middle column show that as expected MSE of the predictions rises with variation of the data (if there was no variation at all, then the model could just memorize 1 image). In the first and third columns, three random examples of projections are shown at module 1 and also at the end of the accelerator at module 48, the bottom row shows three random longitudinal phase-space projections of the beam including at the middle of the accelerator.

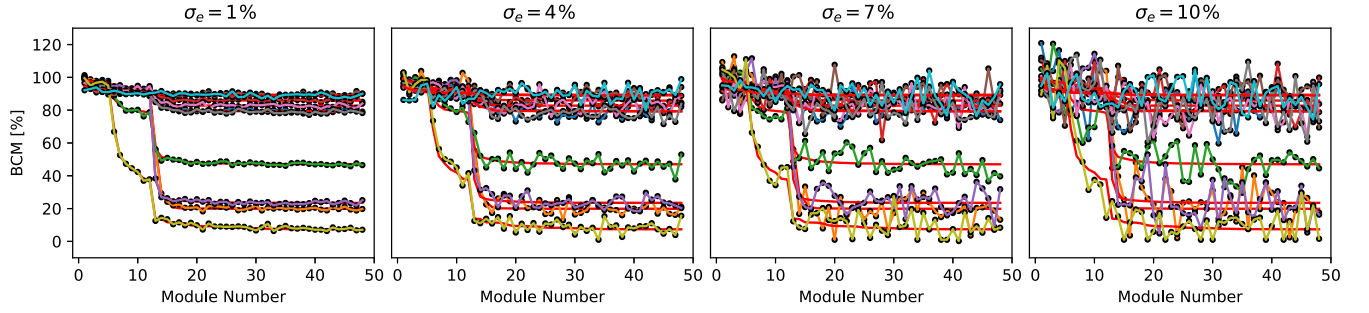


FIG. 11. Examples of 10 noisy BCM readings at various noise levels used for testing the robustness of the approach.

depends on $I_i(t)$ as a function of the impedance of the measurement circuit through which $I_i(t)$ passes and on the transformer coupling and can also be perturbed by external noise. The result is that the overall BCM circuit must be carefully calibrated during installation, and despite such an effort, the actual measured beam current as provided by the BCM will be a different function of $I_i(t)$ at each specific BCM. The result of all of this is that we can see unphysical measurements such as the BCM readings increasing at certain regions of the accelerator, which corresponds to more beam being created as it travels through the machine, which is of course the opposite of what really happens in that beam is continuously lost along the accelerator.

To take into account such effects in our simulation-based synthetic data, and to study the robustness of our generative AI to such measurement noise, we add mean-zero Gaussian noise $\mathcal{N}(0, \sigma)$ of varying standard deviation σ to our BCM signals, where we test sigma values of $\sigma \in \{1\%, 2\%, \dots, 10\%\}$ with the percentage corresponding to the ideal initial 100% beam current at the accelerator entrance, as shown in Fig. 11 for 10 random examples of the beam moving through all 48 modules of LANSCE, with only 4 of the noise levels shown. Recall that our BCM simulation is based on a surviving particle count, which counts the total number of particles at each of the 48 accelerator modules, based on the (φ, E) phase-space projection. Therefore, the red lines shown in Fig. 11 are strictly decreasing as the simulation removes lost particles that intercept accelerator components or the beam pipe. By adding random noise, we can add the AI's robustness to the nonphysical effects that we discuss of BCM readings sometimes slightly increasing.

Figure 12 shows the resulting mean squared error (MSE) of the latent representations generated by the diffusion model when they are conditioned on the true BCM readings and also when the BCM part of the conditional vector is corrupted by increasing levels of noise. The bottom part of the plot shows that there is a clear correlation between noise level and latent image prediction $\hat{z}_j^m$ error, but the impact of the noise even at 10%, is not dramatic, as seen in the top part of the image. Figure 13 then takes those latent images that were generated based on noisy BCM readings and

sends them through the decoder of the VAE to generate approximations of the beam projections $\hat{\rho}_j^m$. Here there is again a clear correlation between noise level and prediction error with no significant jump in error until a 6% noise level is reached. Figure 14 shows one example of the true and

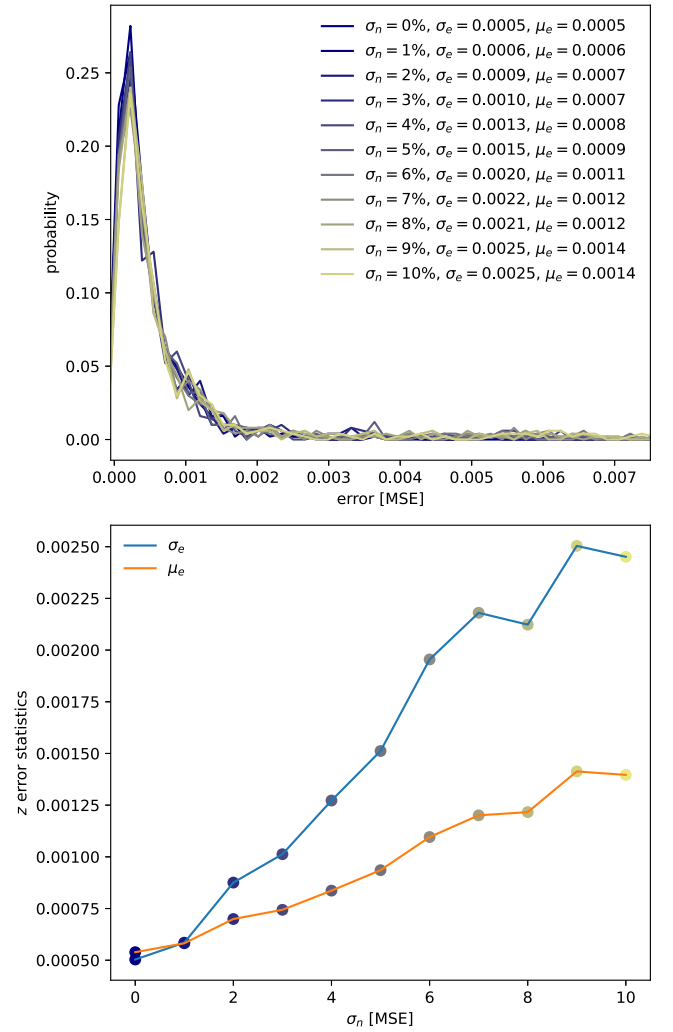


FIG. 12. Error statistics for predicting the latent space representation z at module 48 for the test data when varying levels of noise are added to the beam current measurement.

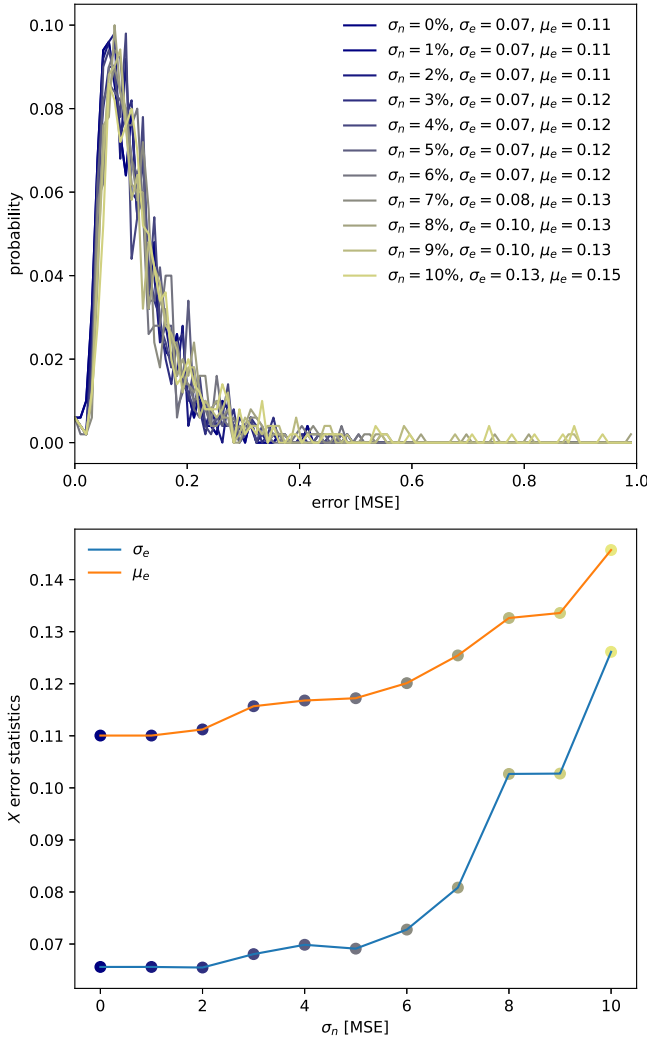


FIG. 13. Error statistics for predicting the 2D (z, φ) projections at module 48 for the test data when varying levels of noise are added to the beam current measurement.

generated 2D (φ, E) projection at module 48 in LANSCE for 0%, 5%, and 10% signal noise levels. The fact that the diffusion model and therefore the final generated image are not extremely sensitive to such noise is promising for testing this method experimentally and are explained by the fact that despite additive noise, overall trend of the BCM reading is on-average preserved, until extremely large noise levels are reached.

VI. ADAPTIVE LATENT DIFFUSION

ML for nonstationary systems is an open problem and an active field of research [55–58]. Typical approaches for nonstationary systems rely on detecting significant changes after which the weights of neural networks are updated/retrained with new information or are continuously trained to keep up with continuous changes [59–61]. Approaches have been developed for the case of covariate shift, where

the input distribution $P(\mathbf{x})$ is different for training and test data, but the conditional distribution of output values $P(y|\mathbf{x})$ remains unchanged [62], based on importance-weighting (IW) techniques [63]. IW methods have also been developed using kernel mean matching methods [64] and by minimizing the Kullback-Leibler divergence between a test data density distribution and its estimate [65,66]. Methods have also been developed for extracting frequencies and amplitudes from time-series data [67], and Bayesian methods have been used with sinusoidal kernels for periodic time-varying systems [68].

Recently, an adaptive machine learning approach has been developed which combines deep learning with the state-of-the-art robust adaptive feedback mechanism from control theory, known as bounded extremum seeking, for stabilizing and optimizing analytically unknown dynamic systems [69–73]. This approach was first demonstrated for automatic control of the (z, E) longitudinal phase space of the LCLS electron beam [14], it has been incorporated within the architecture of generative models for real-time adaptive latent space tuning for real-time tracking of time-varying accelerator beams based on limited measurements [24–26], for tracking time-varying 3D electron densities in coherent diffraction imaging [74], and it has been used for adaptive tuning of conditional vectors that guide generative diffusion models as virtual particle accelerator beam diagnostics [38,39].

For our adaptive application, we consider the case where we have access to a 1D projection of one of the 2D images that we are trying to generate. For example, the projection of the longitudinal phase-space distribution along the phase or time axis, as defined by

$$\rho^m(\varphi) = P_E \rho^m(\varphi, E) = \int_E \rho^m(\varphi, E) dE, \quad (19)$$

is a measure of the bunch current profile at module m . Sometimes, the beam can also be projected along the energy axis, as defined by

$$\rho^m(E) = P_\varphi \rho^m(\varphi, E) = \int_\varphi \rho^m(\varphi, E) d\varphi. \quad (20)$$

In the above, $P_\bullet$ represents the projection operator so that $P_\bullet \rho(\circ, \bullet) = \int \rho(\circ, \bullet) d\bullet$. Such a measurement is performed by passing the beam through a dispersive element, such as a dipole bending magnet, thus correlating transverse position with energy, and then impacting a phosphor screen or scintillator. For any such projection, $P_\bullet$ of a 2D beam phase-space density $X(\tau)$, we can simulate the same projection, $\hat{P}_\bullet$, and apply it to the diffusion-generated version of the 2D distribution $\hat{X}(\tau)$. Comparing the measured and simulated projections, we can define a simple cost function

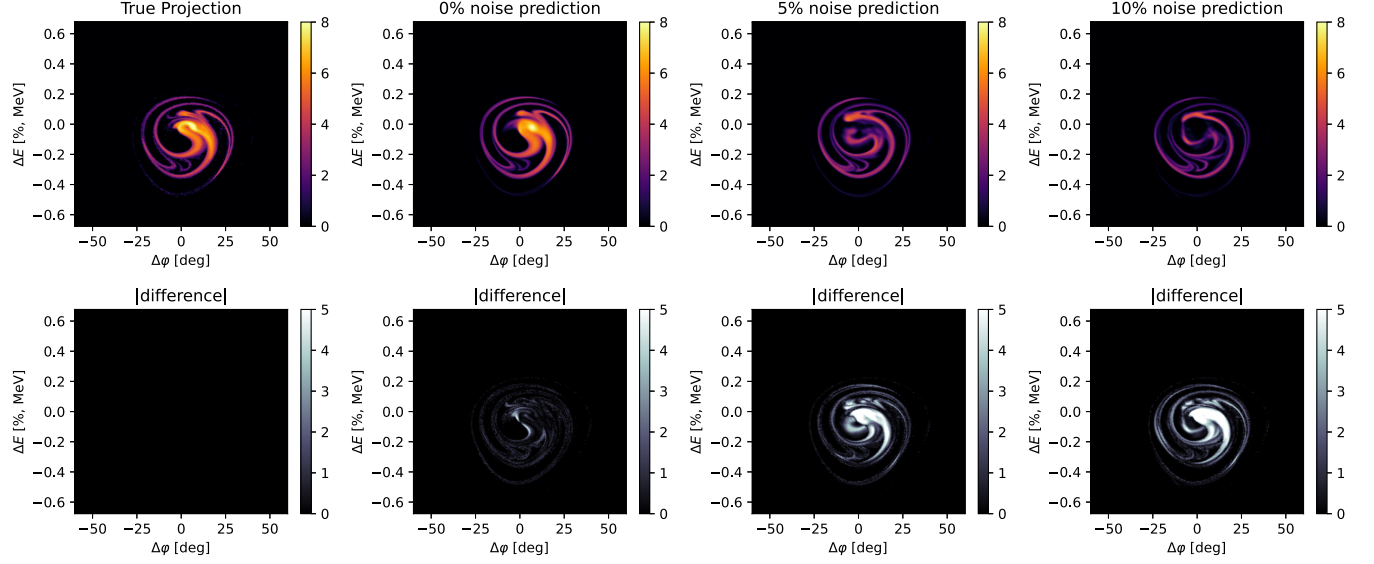


FIG. 14. Comparing predictions to true image of the 2D (ϕ, E) projection at module 48 for the test data when varying levels of noise are added to the beam current measurement. As described in the text, the noise is mean zero Gaussian with only the standard deviation increasing, there is no set offset to the measured signal, it is just perturbed with a higher and higher variance mean-zero random noise.

$$C(\tau, \mathbf{c}) = \|\hat{P}\hat{X}(\mathbf{c}, \tau) - PX(\tau)\|$$

$$= \int |\hat{P}\hat{X}(\mathbf{c}, \tau) - PX(\tau)|, \quad (21)$$

where we have highlighted the fact that the generated projection $\hat{X}$ depends on the conditional vector $\mathbf{c}$. Our goal is then to apply bounded extremum seeking to adaptively

tune the conditional vector $\mathbf{c}$ to continuously track a time-varying $PX(\tau)$.

For this kind of time-varying optimization problem, the bounded extremum seeking method, as first described in [70], simply follows the dynamics

$$\dot{c}_j = \sqrt{\alpha\omega_j} \cos[\omega_j\tau + kC(\tau, \mathbf{c})], \quad (22)$$

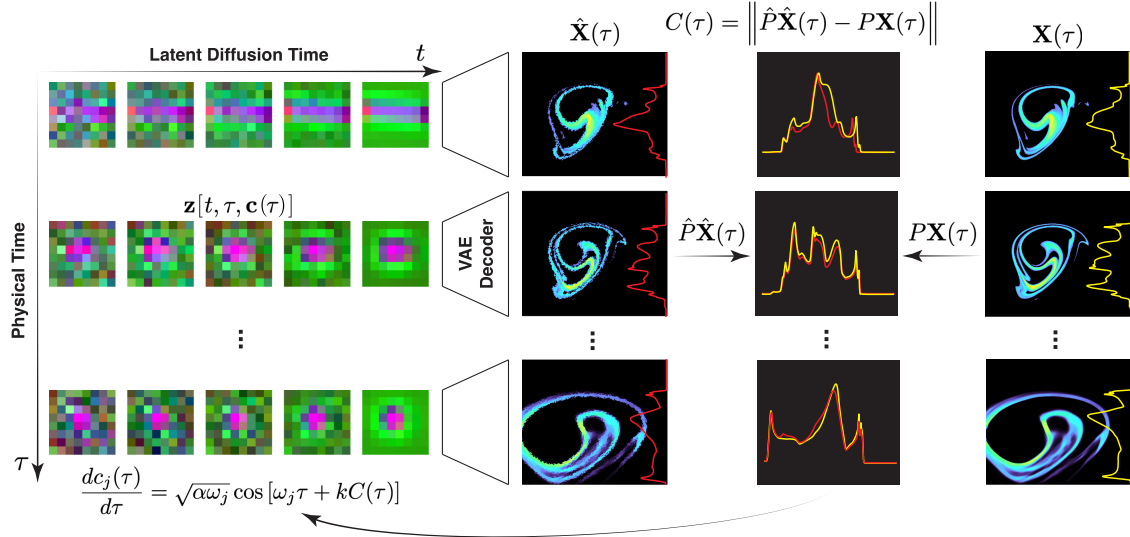


FIG. 15. Overview of adaptive conditional tuning for a virtual 2D diagnostic that tracks a time-varying beam based on 1D projections. Here, τ represents physical time, relative to which accelerator components and initial conditions of the beam are changing, while t represents generative diffusion process time. If a 1D projection $PX(\tau)$ of one of the beam's 2D distributions is available for measurement, it can be compared to a simulated projection $\hat{P}$ of the generated representation of the same 2D distribution given as $\hat{P}\hat{X}(\tau)$. A cost function defined by the difference between these projections is then used to adjust the conditional vector $\mathbf{c}(\tau)$ to adaptively track the time-varying beam.

where $\omega_i \neq \omega_j$ are distinct dithering frequencies. The term α defines the steady-state oscillation amplitudes of the tuned parameters. Once a minimum of $C(\tau, \mathbf{c})$ is reached, $c_j(\tau)$ will oscillate about its minimizing value $c_j^*(\tau)$ with a trajectory of the form

$$c_j(\tau) \approx c_j^*(\tau) + \sqrt{\frac{\alpha}{\omega_j}} \sin(\omega_j \tau + \theta),$$

where θ is some phase offset. To get close to the minimum and for the adaptive dynamics to be faster than the system's time variation, the ω_j must be sufficiently large relative to $\|\partial C / \partial t\|$. The term α also acts as a feedback gain together with the term k as on average, the system dynamics of Eq. (22) can be approximated arbitrarily accurately by the gradient flow dynamics

$$\dot{\mathbf{c}} = -\frac{k\alpha}{2} \nabla_{\mathbf{c}} C(\tau, \mathbf{c}), \quad (23)$$

which for $k\alpha > 0$ finds and tracks the minimum of the analytically unknown function $C(\tau, \mathbf{c})$. Note that, on average, the dynamics are a gradient flow; the adaptive feedback, as defined in Eq. (22), is not attempting any kind of discrete gradient estimation, which makes the method incredibly robust for high-dimensional and noisy dynamical systems. The method has been demonstrated in hardware for automatic particle accelerator tuning including automatic optimization of high-voltage converter modulator rise time [75], automatic FEL output power maximization at both the LCLS and the EuXFEL [76], for real-time multi-objective optimization of the AWAKE electron beamline [77], for automatic beam loss minimization at LANSCE [78,79], and for vertical stabilization of plasmas in a tokamak [80].

An overview of the adaptive setup is shown in Fig. 15, which emphasizes that there are two timescales to consider. The time τ is related to physical time relative to which the beam and accelerator are changing and evolving and the conditional vector $\mathbf{c}(\tau)$ must be adjusted to track the time-varying object. The second timescale t is the diffusion

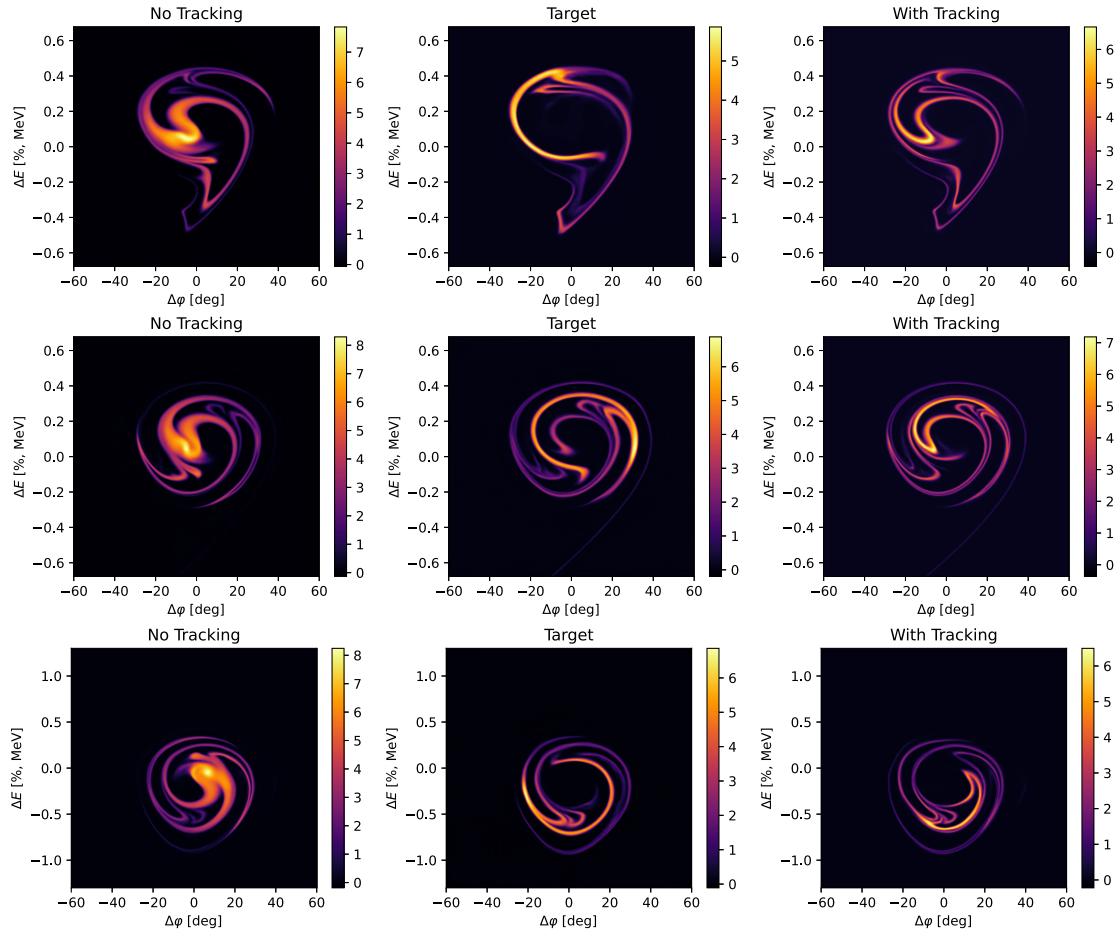


FIG. 16. Demonstration of adaptive longitudinal phase-space tracking at LANSCE. The left columns show initial beam distributions, the middle columns show the distributions that the beams evolve to due to time-varying RF module settings, and the right columns show the adaptively tracked predictions. The top row is for module 28, the next is module 38, and the bottom row is module 48.

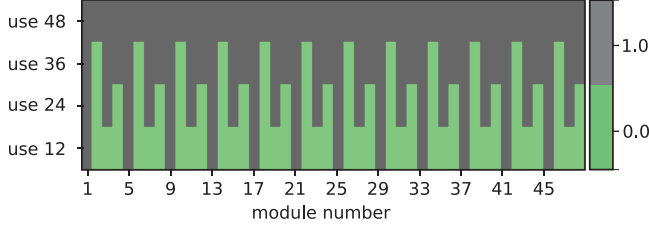


FIG. 17. Mask used to remove beam measurements. Locations of used beam measurements shown in gray, those that are thrown out are shown in green.

process time representing the diffusion-based image generation process.

While this process can, in principal, be applied for any beam projection at any accelerator location, in this demonstration, we focus on using the 1D energy projection as our measurement. We use the energy projection because we have locations in LANSCE where the beam goes around a bend and we can measure the energy spread on a phosphor. We focus on modules near the high-energy end of LANSCE because that is where the energy screens are located and that is also where we need the finest control over the beam's (φ, E) phase space because high-energy beam loss contributes most to equipment damage and radioactive activation of accelerator components and also

the final beam energy must be finely controlled for experiments. Figure 16 shows three examples of an initial beam distribution which then changes because of drifting RF parameters, where we assume that we do not have access to the true RF signals, but we are nevertheless able to track the time-varying beam distribution by continuously adaptively tuning the conditional vector of the generative diffusion process based only on the 1D energy projections. We choose the RF parameters to vary because during LANSCE operations, the RF systems are the fastest time-varying signals and the RF pickup loops have arbitrary temperature-dependent offsets. The eight tuned RF parameters of the accelerator changed approximately 2° and 2% for the phase and amplitude, respectively, and so did the conditional vectors in order to properly track them.

Because our adaptive feedback only has access to a 1D projection of a 2D image, the tracked reconstruction is not perfect as even an exact match of a 1D projection can come from uncountably many different 2D objects. For example, a square and a circle whose diameter is the same length as the side of the square can have the same 1D projection if their densities are defined correctly. Nevertheless, by starting at the correct solution, we see that we can accurately track the time-varying image (only missing minor interior details) based only on tracking the 1D projection because the beam is not just an arbitrary shape,

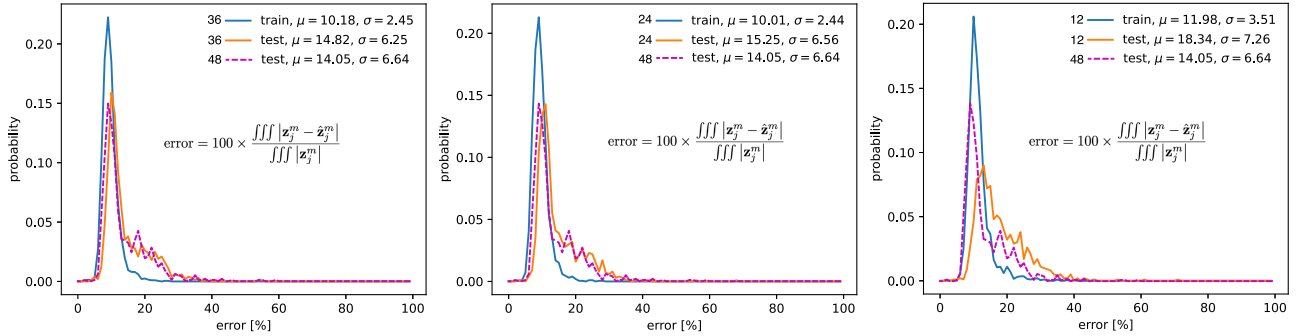


FIG. 18. Reconstruction accuracy of the latent representation as a function of number of beam measurements.

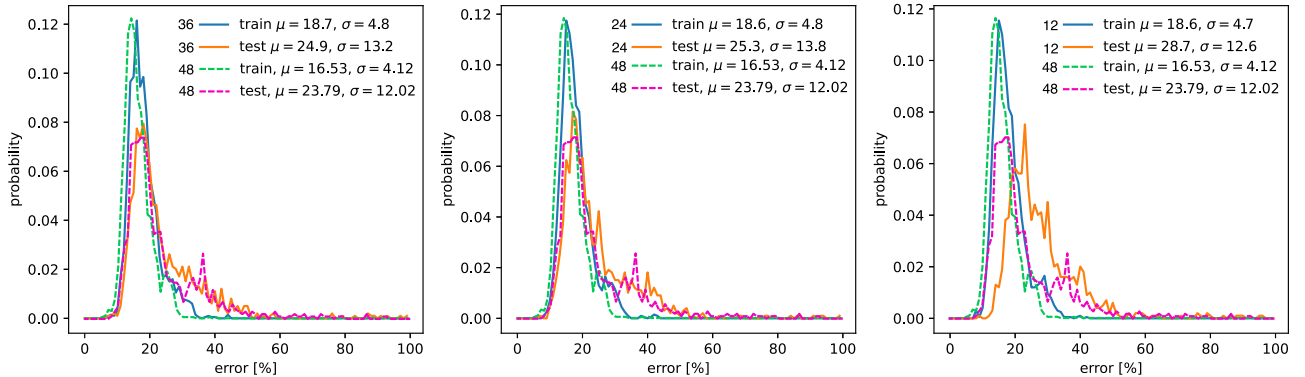


FIG. 19. Reconstruction accuracy of 2D beam phase-space projections as a function of number of beam measurements used to reconstruct that latent representations.

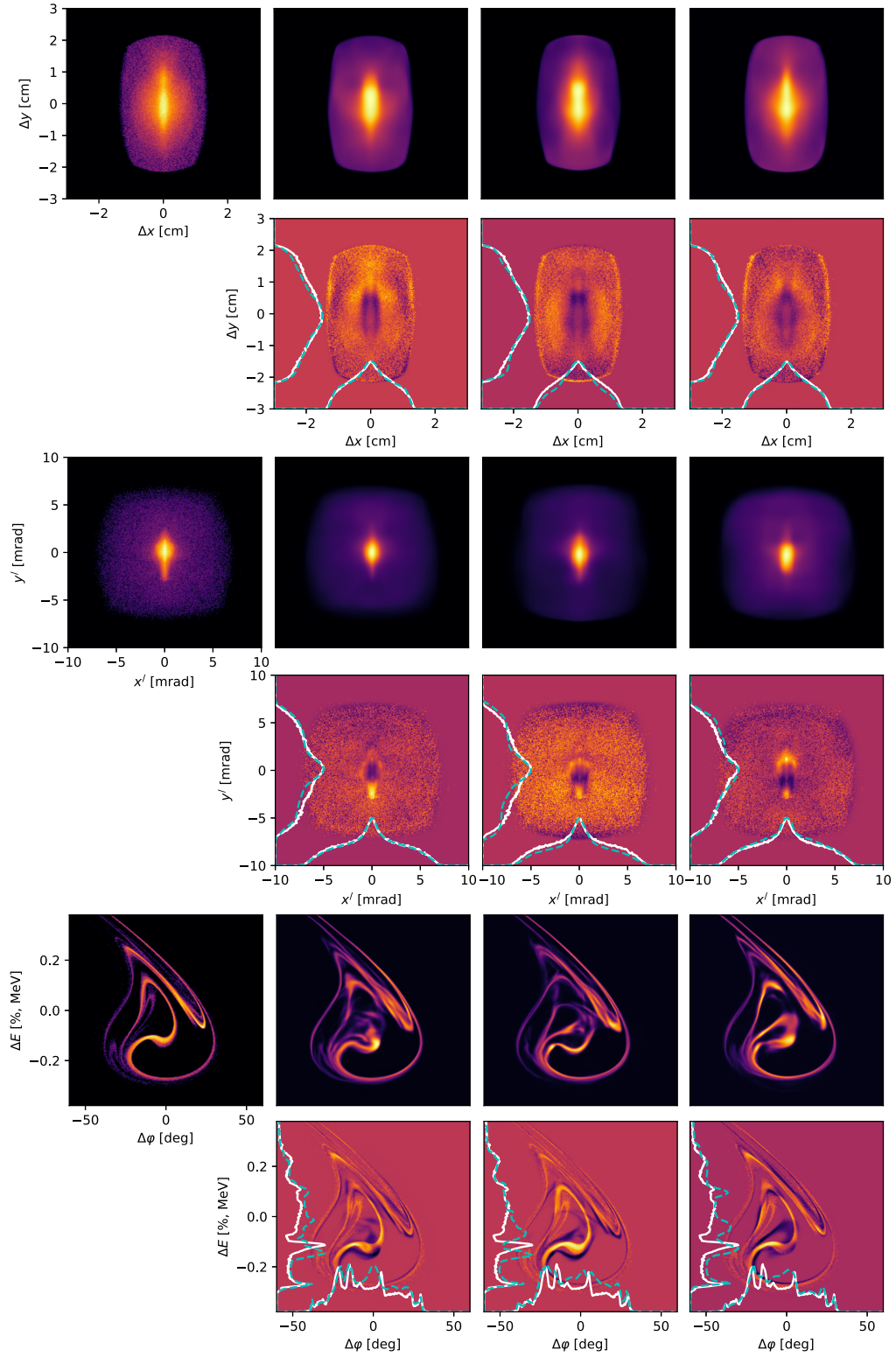


FIG. 20. Example of a good reconstruction by all models. The left column shows the true projection, the second, third, and fourth columns are being generated by the models using 36, 24, and 12 channels, respectively. The 2nd, 4th, and 6th rows show the true 1D projections in solid white and generated 1D projections as dashed lines superimposed on top of the difference between the true and the generated beam images.

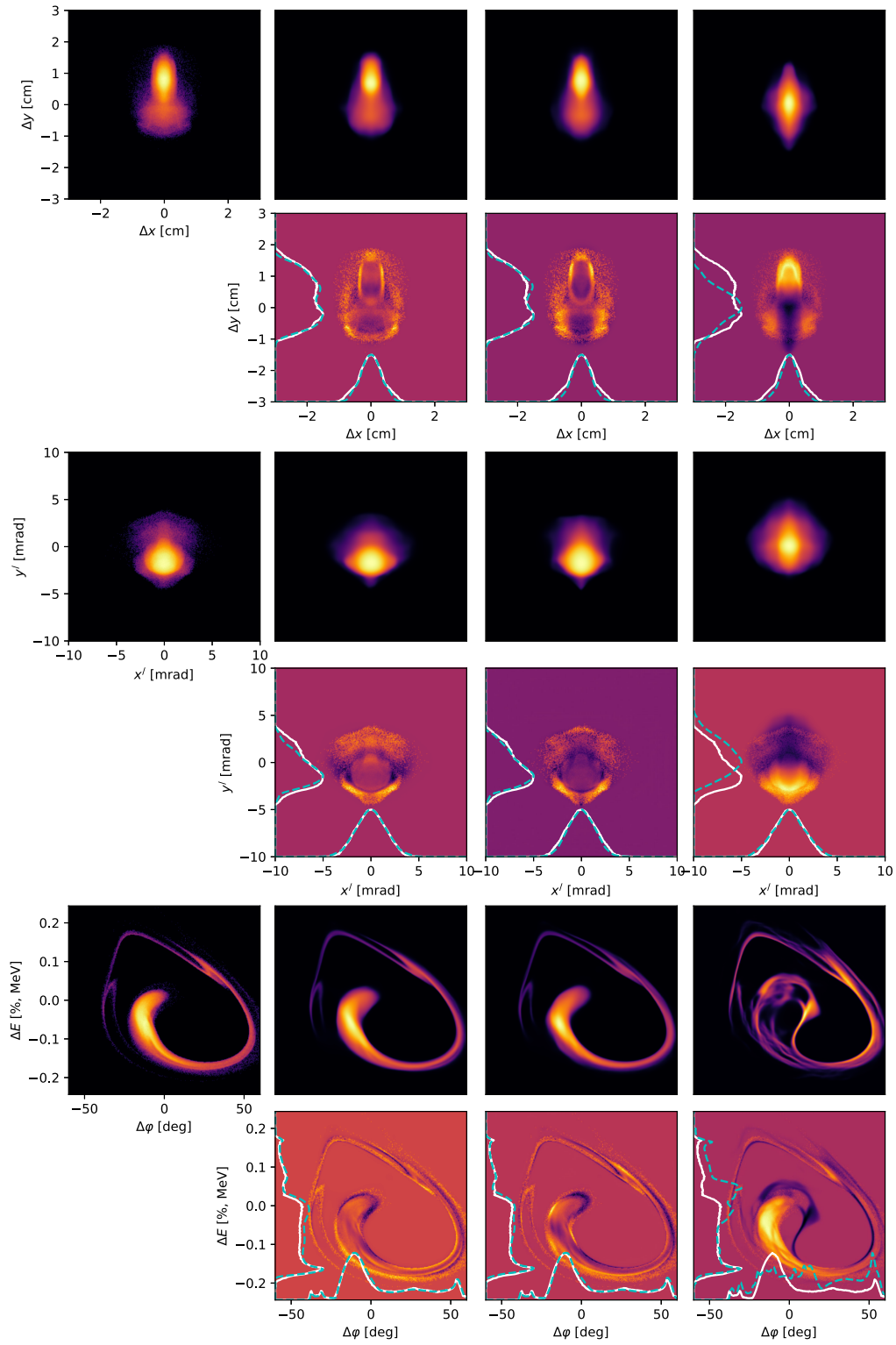


FIG. 21. Example of a bad reconstruction by the model using only 12 channels. The left column shows the true projection, the second, third, and fourth columns are being generated by the models using 36, 24, and 12 channels, respectively. The second, fourth, and sixth rows show the true 1D projections in solid white and generated 1D projections as dashed lines superimposed on top of the difference between the true and the generated beam images.

rather it is defined as a function of RF parameter settings according to a very specific physical process and the diffusion model has learned how to accurately move around in this parameter space.

VII. DEPENDENCE ON NUMBER OF BEAM MEASUREMENTS

The robustness of this approach relative to the number of beam measurements that are available is studied next. All of the analysis above used 48 beam measurements, one after each accelerating module. In order to study how applicable this approach might be to other accelerators, with fewer numbers of beam measurements, we retrain 3 new conditional diffusion models which use only 36, 24, or 12 beam monitors for the conditional vector of the diffusion model, as shown in Fig. 17.

By reducing the amount of detail that is available for the conditional model, the already ill-defined inverse problem should be much more difficult to solve uniquely as there is a greater chance of two different beams having much more similar beam loss or beam current-based signatures that are conditionally guiding the generative diffusion process. For this reason (similar to the discussion above), there should be a growing test error, which grows much faster than the training set error.

Figure 18 shows the statistics of training and test error for the conditionally generated latent images $\hat{z}$, and they are compared to the test error of the original diffusion model that uses all 48 beam signals, as shown in Fig. 8(b). When using only 36 or 24 measurements, the statistics are very similar to that of using all 48, once we go down to just 12 measurements, the error increase is more noticeable. Once the generated $\hat{z}$ images are passed through the same decoder that was used above, we show the error statistics for the reconstructed 2D beam phase-space projections in Fig. 19 alongside the errors when using all 48.

Figure 20 shows an example of a reconstruction that is very good when using all three conditional diffusion models, and Fig. 21 shows an example in which the diffusion model with only 12 signals is producing an inaccurate latent representation that is then being mapped to the wrong 2D phase-space projections by the decoder.

As shown in Figs. 18–21, reducing the number of beam measurements, which reduces the amount of information in the conditional vector of the diffusion model, leads to a reduction in predictive ability, as expected. However, it is also seen that the degradation in prediction accuracy is not severe, with a significant effect showing up only when the number of beam measurements has been reduced by a factor of 4. At that point, when only 12 beam measurements are being used, the prediction accuracy is a mix and more unreliable. The most obvious and easy way to fix this, in the case where only limited numbers of beam measurements are available, is to add additional complementary beam data, such as beam

position monitors, which were not studied in this work. This kind of study is expected to be very accelerator architecture dependent.

VIII. CONCLUSIONS

We have developed a conditionally guided latent diffusion model that can map waveforms of beam loss or beam current measurements along an accelerator to detailed 2D projections of the beam's 6D phase-space density. We showed that the method can interpolate to unseen data points for different initial beam conditions and RF parameter settings. We studied the robustness of the method to noisy BCM readings and showed that it is not prohibitively sensitive to additive noise. We have also shown that if calibrated then adaptive condition tuning can track a time-varying beam based only on 1D projections even when the RF parameters begin to change and are unknown. This transformational method has the potential to be used at any particle accelerator to transform simple noninvasive devices into detailed beam phase-space diagnostics. It is worth noting that the applications studied above relied on the fact that the LANSCE accelerator has non-negligible beam loss and therefore non-negligible changes in current monitor readings so that those signals could inform the generative diffusion process.

In this work, we demonstrated this concept via multi-particle simulations of the high-intensity beam in the kilometer-long LANSCE linear proton accelerator. Future work will focus on expanding to a higher dimensional parameter study including additional RF cavities, magnets, and an experimental demonstration.

Not all particle accelerators operate under significant beam loss conditions. For example, in the nominal operation of the SwissFEL accelerator, nearly 100% transmission is achieved. Moreover, in such conditions, many different phase-space projection techniques can be implemented directly due to the high transmission, and a beam loss or beam current-based method would probably not be applicable in such scenarios. In such cases, it is possible that other rich noninvasive signals, such as beam position monitors, could be used as conditional vectors to inform the 2D phase-space reconstructions, but that is a topic of future work which has not been studied here.

We acknowledge that in order to develop such a generative tool for any of the 2D beam projections, one must have access to a well-calibrated model to train the generative AI and for even more accurate results, one would need a way to experimentally measure projections, such as sending the beam through a screen to measure the transverse (x, y) density or through a TCAV-based system to measure the (z, E) distribution. If such destructive diagnostics are available, they can be used to train the generative model, which can thereafter act as a noninvasive diagnostic that does not intercept the beam but instead relies on beam loss or beam current measurements.

For accelerators where the beam loss is primarily caused by trajectory deviations, it would be interesting to investigate this approach and use beam position monitor (BPM) readings as the conditional vectors that guide the diffusion process. It is well known that BPM signals encode some information about the beam size and shape through the image charge [81–85]. It would be interesting to study how much information a deep generative model can extract from such signals, but that is beyond the scope of this work.

Finally, our study of robustness to the number of beam measurements has shown that, at least for LANSCE, this method remains effective even when the number of measurements is reduced by a factor of 2, from 48 to 24. When reduced by a factor of 4, to just 12 measurements, the results are still decent overall, but with larger errors.

ACKNOWLEDGMENTS

This work was funded by the U.S. Department of Energy (DOE) Los Alamos National Laboratory LDRD Program Directed Research (DR) Project No. 20220074DR.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

-
- [1] P. W. Lisowski and K. F. Schoenberg, The Los Alamos Neutron Science Center, *Nucl. Instrum. Methods Phys. Res., Sect. A* **562**, 910 (2006).
 - [2] S. Henderson, W. Abraham, A. Aleksandrov, C. Allen, J. Alonso, D. Anderson, D. Arenius, T. Arthur, S. Assadi, J. Ayers *et al.*, The spallation neutron source accelerator system design, *Nucl. Instrum. Methods Phys. Res., Sect. A* **763**, 610 (2014).
 - [3] K. Rajput, M. Schram, W. Blokland, Y. Alanazi, P. Ramuhalli, A. Zhukov, C. Peters, and R. Vilalta, Robust errant beam prognostics with conditional modeling for particle accelerators, *Mach. Learn.* **5**, 015044 (2024).
 - [4] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker *et al.*, First lasing and operation of an Ångström-wavelength free-electron laser, *Nat. Photonics* **4**, 641 (2010).
 - [5] W. Decking, S. Abeghyan, P. Abramian, A. Abramsky, A. Aguirre, C. Albrecht, P. Alou, M. Altarelli, P. Altmann, K. Amyan *et al.*, A MHz-repetition-rate hard x-ray free-electron laser driven by a superconducting linear accelerator, *Nat. Photonics* **14**, 391 (2020).
 - [6] 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. Appl.* **16**, 024005 (2021).
 - [7] C. Milne, T. Schietinger, M. Aiba, A. Alarcon, J. Alex, A. Anghel, V. Arsov, C. Beard, P. Beaud, S. Bettoni *et al.*, SwissFEL: The Swiss x-ray free electron laser, *Appl. Sci.* **7**, 720 (2017).
 - [8] E. Prat, R. Abela, M. Aiba, A. Alarcon, J. Alex, Y. Arbelo, C. Arrell, V. Arsov, C. Bacellar, C. Beard *et al.*, A compact and cost-effective hard x-ray free-electron laser driven by a high-brightness and low-energy electron beam, *Nat. Photonics* **14**, 748 (2020).
 - [9] A. Malyzhenkov, Y. P. Arbelo, P. Craievich, P. Dijkstal, E. Ferrari, S. Reiche, T. Schietinger, P. Juranić, and E. Prat, Single- and two-color attosecond hard x-ray free-electron laser pulses with nonlinear compression, *Phys. Rev. Res.* **2**, 042018 (2020).
 - [10] M. Turner, E. Adli, A. Ahuja, O. Apsimon, R. Apsimon, A.-M. Bachmann, M. B. Marin, D. Barrientos, F. Batsch, J. Batkiewicz *et al.*, Experimental observation of plasma wakefield growth driven by the seeded self-modulation of a proton bunch, *Phys. Rev. Lett.* **122**, 054801 (2019).
 - [11] S. Gessner, E. Adli, J. M. Allen, W. An, C. I. Clarke, C. E. Clayton, S. Corde, J. Delahaye, J. Frederico, S. Z. Green *et al.*, Demonstration of a positron beam-driven hollow channel plasma wakefield accelerator, *Nat. Commun.* **7**, 1 (2016).
 - [12] M. J. Hogan, Electron and positron beam-driven plasma acceleration, *Rev. Accel. Sci. Technol.* **9**, 63 (2016).
 - [13] C. Schroeder, E. Esarey, C. Geddes, C. Benedetti, and W. Leemans, Physics considerations for laser-plasma linear colliders, *Phys. Rev. ST Accel. Beams* **13**, 101301 (2010).
 - [14] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and A. Lutman, Demonstration of model-independent control of the longitudinal phase space of electron beams in the linac-coherent light source with femtosecond resolution, *Phys. Rev. Lett.* **121**, 044801 (2018).
 - [15] R. C. Webber, Charged particle beam current monitoring tutorial, *AIP Conf. Proc.* **333**, 3 (1995).
 - [16] B. Cathey, S. Cousineau, A. Aleksandrov, and A. Zhukov, First six dimensional phase space measurement of an accelerator beam, *Phys. Rev. Lett.* **121**, 064804 (2018).
 - [17] A. Hoover, Four-dimensional phase space tomography from one-dimensional measurements of a hadron beam, *Phys. Rev. Accel. Beams* **27**, 122802 (2024).
 - [18] K. Hock and A. Wolski, Tomographic reconstruction of the full 4D transverse phase space, *Nucl. Instrum. Methods Phys. Res., Sect. A* **726**, 8 (2013).
 - [19] S. Kim, J. P. Gonzalez-Aguilera, P. Piot, G. Chen, S. Doran, Y.-K. Kim, W. Liu, C. Whiteford, E. Wisniewski, A. Edelen *et al.*, Four-dimensional phase-space reconstruction of flat and magnetized beams using neural networks and differentiable simulations, *Phys. Rev. Accel. Beams* **27**, 074601 (2024).
 - [20] S. Jaster-Merz, R. Assmann, R. Brinkmann, F. Burkart, W. Hillert, M. Stanitzki, and T. Vinatier, 5D tomographic phase-space reconstruction of particle bunches, *Phys. Rev. Accel. Beams* **27**, 072801 (2024).
 - [21] D. Marx, J. Giner Navarro, D. Cesar, J. Maxson, B. Marchetti, R. Assmann, and P. Musumeci, Single-shot reconstruction of core 4D phase space of high-brightness electron beams using metal grids, *Phys. Rev. Accel. Beams* **21**, 102802 (2018).
 - [22] A. Hoover, N-dimensional maximum-entropy tomography via particle sampling, *arXiv:2409.17915*.

- [23] H. Liu, S. Xu, M. Li, Y. Yuan, L. Huang, X. Feng, and S. Wang, Reconstructing phase space of injection beam based on the measurement of accumulated beam: A novel tomography algorithm and its applications, *Phys. Rev. Accel. Beams* **27**, 052804 (2024).
- [24] A. Scheinker, Adaptive machine learning for time-varying systems: Low dimensional latent space tuning, *J. Instrum.* **16**, P10008 (2021).
- [25] A. Scheinker, F. Cropp, and D. Filippetto, Adaptive autoencoder latent space tuning for more robust machine learning beyond the training set for six-dimensional phase space diagnostics of a time-varying ultrafast electron-diffraction compact accelerator, *Phys. Rev. E* **107**, 045302 (2023).
- [26] A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, An adaptive approach to machine learning for compact particle accelerators, *Sci. Rep.* **11**, 19187 (2021).
- [27] R. Roussel, A. Edelen, C. Mayes, D. Ratner, J. P. Gonzalez-Aguilera, S. Kim, E. Wisniewski, and J. Power, Phase space reconstruction from accelerator beam measurements using neural networks and differentiable simulations, *Phys. Rev. Lett.* **130**, 145001 (2023).
- [28] R. Roussel, J. P. Gonzalez-Aguilera, E. Wisniewski, A. Ody, W. Liu, J. Power, Y.-K. Kim, and A. Edelen, Efficient six-dimensional phase space reconstructions from experimental measurements using generative machine learning, *Phys. Rev. Accel. Beams* **27**, 094601 (2024).
- [29] M. Rautela, A. Williams, and A. Scheinker, A conditional latent autoregressive recurrent model for generation and forecasting of beam dynamics in particle accelerators, *Sci. Rep.* **14**, 18157 (2024).
- [30] A. Hoover and J. C. Wong, High-dimensional maximum-entropy phase space tomography using normalizing flows, *Phys. Rev. Res.* **6**, 033163 (2024).
- [31] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in *Proceedings of the International Conference on Machine Learning* (PMLR, 2015), pp. 2256–2265, <http://proceedings.mlr.press/v37/sohl-dickstein15.html>.
- [32] J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models, in *Proceedings of the 34th International Conference on Advances in Neural Information Processing System* (2020), Vol. 33, pp. 6840–6851, <https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html>.
- [33] J. Song, C. Meng, and S. Ermon, Denoising diffusion implicit models, [arXiv:2010.02502](https://arxiv.org/abs/2010.02502).
- [34] A. Q. Nichol and P. Dhariwal, Improved denoising diffusion probabilistic models, in *Proceedings of the International Conference on Machine Learning* (PMLR, 2021), pp. 8162–8171, <https://proceedings.mlr.press/v139/nichol21a.html>.
- [35] T. Karras, M. Aittala, T. Aila, and S. Laine, Elucidating the design space of diffusion-based generative models, in *Proceedings of the 36th International Conference on Advances in Neural Information Processing System* (2022), Vol. 35, pp. 26565–26577, https://proceedings.neurips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract-Conference.html.
- [36] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, High-resolution image synthesis with latent diffusion models, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 10684–10695, https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.
- [37] A. Scheinker, Conditional guided generative diffusion for particle accelerator beam diagnostics, *Sci. Rep.* **14**, 19210 (2024).
- [38] A. Scheinker, cDVAE: VAE-guided diffusion for particle accelerator beam 6D phase space projection diagnostics, *Sci. Rep.* **14**, 29303 (2024).
- [39] A. Scheinker, Physics-constrained superresolution diffusion for six-dimensional phase space diagnostics, *Phys. Rev. Res.* **7**, 023091 (2025).
- [40] X. Pang and L. Rybarczyk, GPU accelerated online multi-particle beam dynamics simulator for ion linear particle accelerators, *Comput. Phys. Commun.* **185**, 744 (2014).
- [41] D. Kleinjan, G. Rouleau, and L. Neukirch, Design of experiment to explore inside the LANSCE-H[−] ion source with laser absorption techniques, *AIP Conf. Proc.* **2373**, 040003 (2021).
- [42] E. M. O’Brien, S. A. Baily, E. R. Birnbaum, C. Chapman, E. A. Espinoza, J. A. Faucett, J. O. Hill, K. D. John, P. S. Marroquin, R. C. McCrady *et al.*, Novel design and diagnostics improvements for increased production capacity and improved reliability at the Los Alamos Isotope Production Facility, *Nucl. Instrum. Methods Phys. Res., Sect. A* **956**, 163316 (2020).
- [43] C. Morris, J. Allison, S. Cool, C. Cude-Woods, M. Freeman, K. McClellan, F. Mariam, W. Meijer, L. Neukirch, M. Schanz *et al.*, Large-grain scintillator screens for proton radiography, *Rev. Sci. Instrum.* **95**, 083707 (2024).
- [44] P. Ramachandran, B. Zoph, and Q. V. Le, Searching for activation functions, [arXiv:1710.05941](https://arxiv.org/abs/1710.05941).
- [45] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, [arXiv:1502.03167](https://arxiv.org/abs/1502.03167).
- [46] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany* (Springer, New York, 2015), pp. 234–241, [arXiv:1505.04597](https://arxiv.org/abs/1505.04597).
- [47] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, PixelCNN++: Improving the PixelCNN with discretized logistic mixture likelihood and other modifications, [arXiv:1701.05517](https://arxiv.org/abs/1701.05517).
- [48] Y. Wu and K. He, Group normalization, in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 3–19, https://openaccess.thecvf.com/content_ECCV_2018/html/Yuxin_Wu_Group_Normalization_ECCV_2018_paper.html.
- [49] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen, A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, *J. Neurosci.* **13**, 4700 (1993).

- [50] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, Modeling visual attention via selective tuning, *Artif. Intell.* **78**, 507 (1995).
- [51] L. Itti and C. Koch, Computational modelling of visual attention, *Nat. Rev. Neurosci.* **2**, 194 (2001).
- [52] V. Mnih, N. Heess, A. Graves *et al.*, Recurrent models of visual attention, in *Proceedings of the 28th International Conference on Advances in Neural Information Processing Systems* (2014), Vol. 27, pp. 2204–2212, https://proceedings.neurips.cc/paper_files/paper/2014/file/3e456b31302cf8210edd4029292a40ad-Paper.pdf.
- [53] M.-T. Luong, H. Pham, and C. D. Manning, Effective approaches to attention-based neural machine translation, [arXiv:1508.04025](https://arxiv.org/abs/1508.04025).
- [54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *Proceedings of the 31st International Conference on Advances in Neural Information Processing System* (2017), Vol. 30, pp. 6000–6010, <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.
- [55] T. Goehring, M. Keshavarzi, R. P. Carlyon, and B. C. Moore, Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants, *J. Acoust. Soc. Am.* **146**, 705 (2019).
- [56] M. A. M. Ramírez, E. Benetos, and J. D. Reiss, A general-purpose deep learning approach to model time-varying audio effects, [arXiv:1905.06148](https://arxiv.org/abs/1905.06148).
- [57] G. Javadi, M. N. N. To, S. Samadi, S. Bayat, S. Sojoudi, A. Hurtado, S. Chang, P. Black, P. Mousavi, and P. Abolmaesumi, Complex cancer detector: Complex neural networks on non-stationary time series for guiding systematic prostate biopsy, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, New York, 2020), pp. 524–533, [10.1007/978-3-030-59716-0_50](https://doi.org/10.1007/978-3-030-59716-0_50).
- [58] J. S. Dramsch, M. Lüthje, and A. N. Christensen, Complex-valued neural networks for machine learning on non-stationary physical data, *Comput. Geosci.* **146**, 104643 (2021).
- [59] R. Calandra, T. Raiko, M. P. Deisenroth, and F. M. Pouzols, Learning deep belief networks from non-stationary streams, in *Proceedings of the International Conference on Artificial Neural Networks* (Springer, New York, 2012), pp. 379–386, [10.1007/978-3-642-33266-1_47](https://doi.org/10.1007/978-3-642-33266-1_47).
- [60] A. Koesdwiady, S. Bedawi, C. Ou, and F. Karray, Non-stationary traffic flow prediction using deep learning, in *Proceedings of the 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)* (IEEE, New York, 2018), pp. 1–5, [10.1109/VTCFall.2018.8690756](https://doi.org/10.1109/VTCFall.2018.8690756).
- [61] R. Kurle, B. Cseke, A. Klushyn, P. van der Smagt, and S. Günnemann, Continual learning with Bayesian neural networks for non-stationary data, in *Proceedings of the International Conference on Learning Representations* (2019), <https://openreview.net/forum?id=SJlsFpVtDB>.
- [62] H. Shimodaira, Improving predictive inference under covariate shift by weighting the log-likelihood function, *J. Stat. Plann. Infer.* **90**, 227 (2000).
- [63] G. Fishman, *Monte Carlo: Concepts, Algorithms, and Applications* (Springer Science & Business Media, Berlin, 2013).
- [64] J. Huang, A. Gretton, K. Borgwardt, B. Schölkopf, and A. Smola, Correcting sample selection bias by unlabeled data, in *Proceedings of the 20th International Conference on Advances in Neural Information Processing Systems* (2006), Vol. 19, pp. 601–608, <https://proceedings.neurips.cc/paper/2006/hash/a2186aa7c086b46ad4e8bf81e2a3a19b-Abstract.html>.
- [65] M. Sugiyama, S. Nakajima, H. Kashima, P. Von Buenau, and M. Kawanabe, Direct importance estimation with model selection and its application to covariate shift adaptation, in *Proceedings of the NIPS* (Citeseer, 2007), Vol. 7, pp. 1433–1440, <https://proceedings.neurips.cc/paper/2007/hash/be83ab3ecd0db773eb2dc1b0a17836a1-Abstract.html>.
- [66] M. Sugiyama and M. Kawanabe, *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation* (MIT Press, Cambridge, MA, 2012).
- [67] D. E. Shea, R. Giridharagopal, D. S. Ginger, S. L. Brunton, and J. N. Kutz, Extraction of instantaneous frequencies and amplitudes in nonstationary time-series data, [arXiv:2104.01293](https://arxiv.org/abs/2104.01293).
- [68] N. Kuklev, Y. Sun, H. Shang, M. Borland, and G. Fystro, Time-aware Bayesian optimization for adaptive particle accelerator tuning, in *Proceedings of the Machine Learning and the Physical Sciences NeurIPS Workshop* (2022), https://ml4physicalsciences.github.io/2022/files/NeurIPS_ML4PS_2022_175.pdf.
- [69] A. Scheinker and M. Krstić, Minimum-seeking for CLFS: Universal semiglobally stabilizing feedback under unknown control directions, *IEEE Trans. Autom. Control* **58**, 1107 (2012).
- [70] A. Scheinker, Model independent beam tuning, in *Proceedings of the International Particle Accelerator Conference (IPAC-2013)*, Shanghai, China (JACoW, Geneva, Switzerland, 2013), pp. 1862–1864, <https://proceedings.jacow.org/IPAC2013/papers/tupwa068.pdf>.
- [71] A. Scheinker and D. Scheinker, Bounded extremum seeking with discontinuous dithers, *Automatica* **69**, 250 (2016).
- [72] A. Scheinker and D. Scheinker, Constrained extremum seeking stabilization of systems not affine in control, *Int. J. Robust Nonlinear Control* **28**, 568 (2018).
- [73] A. Scheinker and D. Scheinker, Extremum seeking for optimal control problems with unknown time-varying systems and unknown objective functions, *Int. J. Adapt. Control Signal Process.* **35**, 1143 (2021).
- [74] A. Scheinker and R. Pokharel, Enabling dynamic 3D coherent diffraction imaging via adaptive latent space tuning of generative autoencoders, *npj Comput. Mater.* **10**, 283 (2024).
- [75] A. Scheinker, M. Bland, M. Krstić, and J. Audia, Extremum seeking-based optimization of high voltage converter modulator rise-time, *IEEE Trans. Control Syst. Technol.* **22**, 34 (2013).
- [76] A. Scheinker, D. Bohler, S. Tomin, R. Kammering, I. Zagorodnov, H. Schlarb, M. Scholz, B. Beutner, and W. Decking, Model-independent tuning for maximizing free

- electron laser pulse energy, *Phys. Rev. Accel. Beams* **22**, 082802 (2019).
- [77] A. Scheinker, S. Hirlander, F. M. Velotti, S. Gessner, G. Z. Della Porta, V. Kain, B. Goddard, and R. Ramjiawan, Online multi-objective particle accelerator optimization of the AWAKE electron beam line for simultaneous emittance and orbit control, *AIP Adv.* **10**, 055320 (2020).
- [78] A. Scheinker, S. Baily, D. Young, J. S. Kolski, and M. Prokop, In-hardware demonstration of model-independent adaptive tuning of noisy systems with arbitrary phase drift, *Nucl. Instrum. Methods Phys. Res., Sect. A* **756**, 30 (2014).
- [79] A. Scheinker, E.-C. Huang, and C. Taylor, Extremum seeking-based control system for particle accelerator beam loss minimization, *IEEE Trans. Control Syst. Technol.* **30**, 2261 (2021).
- [80] S. Dubbioso, G. De Tommasi, C. Galperti, F. Felici, S. Marchioni, A. Mele, A. Merle, and A. Tenaglia, Simulation validation of an extremum seeking-based vertical stabilization system for TCV, *Fusion Eng. Des.* **219**, 115198 (2025).
- [81] R. H. Miller, J. Clendenin, M. James, and J. Sheppard, Nonintercepting emittance monitor, Technical Report No. SLAC-PUB-3186, SLAC National Accelerator Laboratory, Menlo Park, CA, 1983, [10.2172/5746569](#).
- [82] S. J. Russell and B. E. Carlsten, Measuring emittance using beam position monitors, in *Proceedings of the 15th Particle Accelerator Conference, PAC-1993, Washington, DC, 1993* (IEEE, New York, 1993), pp. 2537–2539.
- [83] S. J. Russell, J. D. Gilpatrick, J. F. Power, and R. B. Shurter, Characterization of beam position monitors for measurement of second moment, in *Proceedings of the Particle Accelerator Conference, Dallas, TX, 1995* (IEEE, New York, 1995), Vol. 4, pp. 2580–2582, [10.1109/PAC.1995.505624](#).
- [84] A. Jansson, Noninvasive single-bunch matching and emittance monitor, *Phys. Rev. ST Accel. Beams* **5**, 072803 (2002).
- [85] J. A. Tsemo Kamga, W. F. Müller, and T. Weiland, Analytical and numerical calculation of the second-order moment of the beam using a capacitive pickup, *Phys. Rev. Accel. Beams* **19**, 042801 (2016).