

Efficient 6-dimensional phase space reconstruction from experimental measurements using generative machine learning

Ryan Roussel and Auralee Edelen

SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

Juan Pablo Gonzalez-Aguilera and Young-Kee Kim

Department of Physics and Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA

Eric Wisniewski, Alex Ody, Wanming Liu, and John Power

Argonne National Laboratory, Lemont, IL 60439, USA

(Dated: April 18, 2024)

Next-generation accelerator concepts which hinge on the precise shaping of beam distributions, demand equally precise diagnostic methods capable of reconstructing beam distributions within 6-dimensional position-momentum spaces. However, the characterization of intricate features within 6-dimensional beam distributions using conventional diagnostic techniques necessitates hundreds of measurements, using many hours of valuable beam time. Novel phase space reconstruction techniques are needed to substantially reduce the number of measurements required to reconstruct detailed, high-dimensional beam features in order to resolve complex beam phenomena, and as feedback in precision beam shaping applications. In this study, we present a novel approach to reconstructing detailed 6-dimensional phase space distributions from experimental measurements using generative machine learning and differentiable beam dynamics simulations. We demonstrate that for a collection of synthetic beam distribution test cases that this approach can be used to resolve 6-dimensional phase space distributions using basic beam manipulations and as few as 20 2-dimensional measurements of the beam profile, without the need for prior data collection or model training. We also demonstrate an application of the reconstruction method in an experimental setting at the Argonne Wakefield Accelerator, where it is able to reconstruct the beam distribution and accurately predict previously unseen measurements 75x faster than previous methods.

I. INTRODUCTION

Current and future particle accelerators rely on the precise control of beam distributions in six-dimensional position-momentum phase space, especially as accelerators push the boundaries of intensity and brightness for collider and light source applications [1]. For example, magnetized electron beams for hadron beam cooling [2], two bunch operations in free electron lasers [3], and drive/witness bunch pairs for high efficiency wakefield acceleration [4], are enabled by the precise control of 6-dimensional phase spaces beyond macroscopic (RMS) beam properties.

Achieving this level of control requires measurement techniques that provide detailed information about the beam distribution in all 6 phase space coordinates, including cross-correlations between the different phase space coordinates. Measurements of this type are needed to provide feedback to accelerator operators or autonomous control algorithms to tune accelerator parameters such that the measured beam distribution matches what is needed by applications. Additionally, detailed characterization of 6-dimensional phase space distributions is necessary to resolve complex beam dynamics phenomena that couples particle motion along multiple axes, such as coherent synchrotron radiation [5] or plasma wakefield accelerators [6].

A wide variety of beam manipulation and diagnostic techniques have been developed to measure and pre-

dict detailed characteristics of phase space distributions. These techniques can involve rotating the beam in phase space (tomography) [7–10], using masks or meshes to isolate and observe the dynamics of portions of the transverse beam [11], or using specialized, non-destructive beam diagnostics such as laser wires [12], or machine learning models [13–15] to predict the beam distribution from experimental measurements.

Reconstructing detailed 5- or 6-dimensional phase space distributions from experimental data has proven to be substantially more difficult than reconstructing lower dimensional spaces (4 or less phase space coordinates). Experimental 5-dimensional phase space characterization has been done in limited instances, once at the ARES beamline with a polarizable X-band transverse deflecting cavity [16, 17] using the Simultaneous Algebraic Reconstruction Technique (SART) algorithm [18], and once at the Spallation Neutron Source (SNS) using a set of movable masking slits [19]. Full 6-dimensional reconstruction of a single beam distribution has also only been done once, by combining multiple, scanning masking slits with a dipole spectrometer and a bunch shape monitor at the SNS beamline [20]. However, these measurements required a significant amount beam time resources to carry out (960 measurements over ~ 28 hrs for the 5-dimensional ARES case, over 5 million measurements over ~ 32 hrs for the 6-dimensional SNS measurement), making these measurement procedures impractical for regular use as feedback for online accelerator tun-

ing or for understanding complex beam dynamics phenomena. Reconstruction methods that significantly reduce the number of measurements needed to reconstruct the 6-dimensional beam distribution in order to be used during regular accelerator operations.

In previous work, we introduced and demonstrated a novel method for reconstructing detailed beam distributions using generative machine learning models and differentiable beam dynamics simulations [21], which we refer to here as generative phase space reconstruction (GPSR). The generative machine learning model creates nearly arbitrary beam distributions in 6-dimensional phase space by transforming random samples drawn from a fixed probability distribution into macroparticles with realistic position-momentum coordinates using a small neural network. A differentiable beam dynamics simulation, which allows for analytical calculation of derivatives during each computational step, is then used to train parameters of the generative model (which in turn modifies the generated beam distribution) by minimizing the difference between experimental measurements and simulated predictions. Combining generative models with differentiable beam dynamics simulations allows the generative model to be trained solely from experimental measurements in a short amount of time, removing the need for large initial training sets usually required by machine learning workflows. This method was used to reconstruct detailed, correlated, 4-dimensional phase space distributions using a single quadrupole scan (10 measurements of the beam distribution) by taking advantage of the information contained in conventional measurements of the transverse beam profile using a YAG screen [22]. More recently, this method was also used to reconstruct beams with large angular momentum components (magnetization) and large emittance ratios [23].

In this work, we extend the use of GPSR to reconstructing 6-dimensional phase space distributions from experimental measurements. We demonstrate in simulation that a diagnostic beam line with a single quadrupole, transverse deflecting cavity, and dipole spectrometer can be used to resolve detailed characteristics of 6-dimensional phase space distributions using as little as 20 measurements, reducing the beam time necessary to reconstruct the beam distribution from many hours to several minutes. We show that the reconstruction technique accurately reconstructs a variety of different beam distributions including pure Gaussian distributions, correlated distributions, and beam distributions similar to those produced by the emittance exchange (EEX) beamline at the Argonne Wakefield Accelerator [24]. We then apply the GPSR algorithm to reconstructing a beam distribution from experimental measurements and demonstrate that it accurately predicts additional measurements in experiment. Finally, we discuss the current limitations and advantages of the reconstruction technique.

II. METHODS

The method we use to reconstruct high-dimensional phase space distributions targets two issues encountered by conventional reconstruction techniques.

First, these measurement techniques use low-dimensional projections of the 6-dimensional beam to infer distribution structure. Conventional algebraic reconstruction techniques such as Maximum Entropy Tomography (MENT) [25] and SART typically use 1-dimensional projections of screen images to produce a higher dimensional reconstruction. Alternatively, measuring the intensity of the beam distribution through a set of slits, as done in [20], further reduces the dimensionality of the measurement down to a scalar quantity. As a result, these methods lose most of the information about the beam distribution by projecting them to lower-dimensional measurements, requiring significantly more individual measurements of the beam distribution to resolve higher-dimensional features.

Second, conventional techniques for describing beam distributions scale extremely poorly when describing 6-dimensional beam distributions. Both of the methods listed above use a high-dimensional histogramming approach to describe the beam distribution, i.e., they solve for the beam density inside a number of high-dimensional voxels along a 6-dimensional mesh. While this formalism has been successfully used to describe 1-2 dimensional distributions, it becomes extremely expensive to describe beam distributions in 6-dimensional space as the number of bins grows exponentially with the number of dimensions. For example, resolving a beam distribution with 100 bins per dimension results in 10^{12} voxels needed to describe the full 6-dimensional distribution. Alternative methods for representing beam distributions, such as distribution moments or principal component analysis techniques, suffer from similar dimensionality issues.

To overcome these challenges we use two recent developments in machine learning and differentiable beam dynamics modeling.

First, we use a generative machine learning model to represent a distribution of macro particles in 6-dimensional space, inspired by the development of a technique known as neural radiance fields [26], which uses neural networks to represent mass and color density functions in 3-dimensional space. This model uses a small neural network to transform randomly generated samples from a 6-dimensional multivariate Normal distribution into real macroparticle phase space coordinates in position-momentum space. Neural networks of sufficient complexity are universal function approximators [27], enabling this transformation network to produce particle distributions with nearly arbitrary structure in high-dimensional phase space. For the work here, we found that a fully connected neural network with 2 layers, 50 neurons each, connected by Tanh activation functions was sufficient to represent nearly arbitrary beams with high enough detail. With this method, the distribution

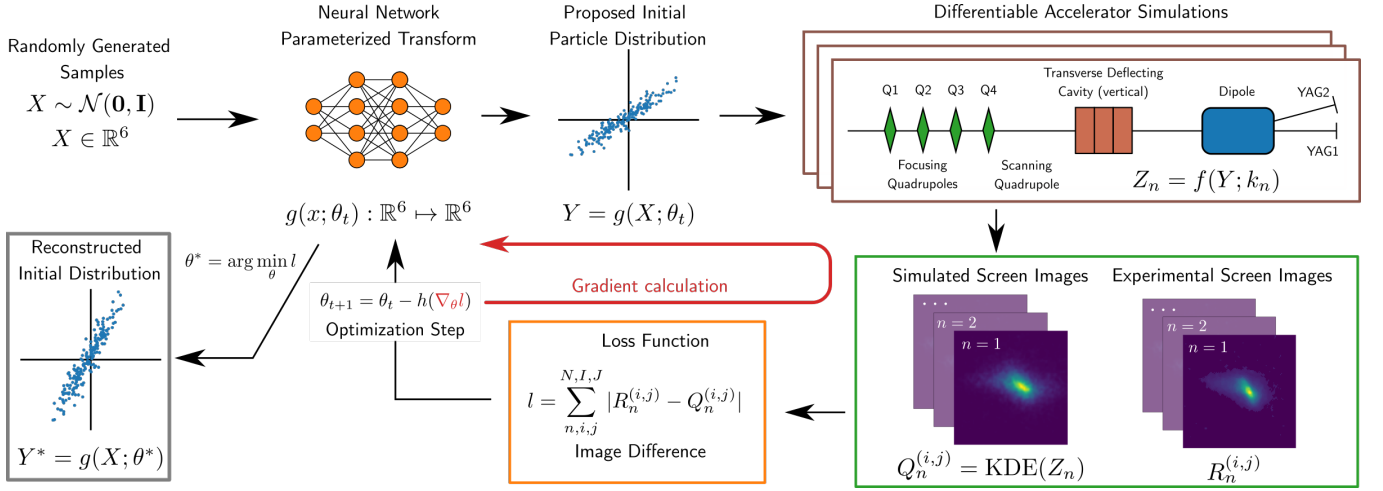


FIG. 1. Description of the GPSR approach for reconstructing phase space beam distributions from experimental data. A 6-dimensional beam distribution is parameterized via a generative machine learning model, where randomly generated samples drawn from a Multivariate Normal distribution are transformed using a neural network into macroparticle coordinates in position-momentum space. The beam distribution is then transported through a backwards differentiable simulation of the diagnostic beamline to make measurement predictions at the diagnostic screens. The total per-pixel difference in intensity between simulated predictions $Q_n^{(i,j)}$ and experimental measurements $R_n^{(i,j)}$ is calculated as a loss function. The neural network parameters θ are then optimized to minimize the loss function using gradients calculated by the differentiable beam dynamics simulation. The distribution generated once the loss function has been minimized (simulation predictions match experimental measurements) is the reconstructed beam distribution.

of macroparticles in 6-dimensional phase space is entirely controlled by the parameters of the neural network, resulting in a parameterization using approximately 1000 free parameters, as opposed to the millions of parameters needed for mesh-based representations of 6-dimensional distributions.

The reconstruction process determines the parameters of the neural network at the center of the generative beam model by solving an optimization problem, shown in Fig. 1. Optimization is used to minimize a loss function that is proportional to the mean squared discrepancy between simulated measurement predictions and experimental measurements, in this case the per-pixel intensity of a transverse YAG screen. Due to the number of free parameters contained inside the generative beam distribution model, solving this optimization problem in a reasonable amount of time is beyond the capabilities of black box optimization algorithms. Black box optimization algorithms, such as Bayesian optimization [28], use evaluations of the loss function to iteratively solve for the ideal parameter set that reaches the minimum objective value. However, due to the so-called “curse-of-dimensionality” [29], these algorithms require an exponentially increasing number of iterations to reach a minimum value when so many free parameters are present.

Given a large number of free parameters, it is necessary to use gradients of the loss function in addition to the loss function value to inform optimization and use gradient-descent-based algorithms to solve optimization problems. Unfortunately, this is also prohibitively difficult when using conventional beam dynamics simulations to predict

experimental measurements, since calculating the gradients numerically in these cases requires finite-difference methods that also scale poorly with the number of free parameters used in optimization.

To address this challenge, we take advantage of recent developments in automatic differentiation programming to substantially reduce the cost of evaluating gradients of simulation outputs with respect to many input parameters. Automatic differentiation is the technique of tracking derivative information alongside each computation step during physics simulations. This in turn, allows analytical evaluation of simulation output derivatives by using the chain rule in reverse to determine derivatives with respect to input parameters, as process commonly referred to as “backpropagation” [30] or “adjoint-differentiation” [31]. Computational costs of calculating the derivatives in this manner is roughly equivalent to the cost of evaluating the simulation itself, and more importantly, is *independent* of the number of input parameters the derivative is taken with respect to, thus making it substantially cheaper to compute than finite-difference methods when optimizing with respect to a large number of input parameters.

It should be noted that this process is similar to, but distinct from, previous uses of automatic differentiation in accelerator physics, often referred to as “forward-mode” differentiation [31], “differential algebra” [32], or, when computing higher derivatives, “truncated power series algebra” [33]. These techniques are well suited for calculating particle transport dynamics up to arbitrary order, however they do not scale well to calculating gra-

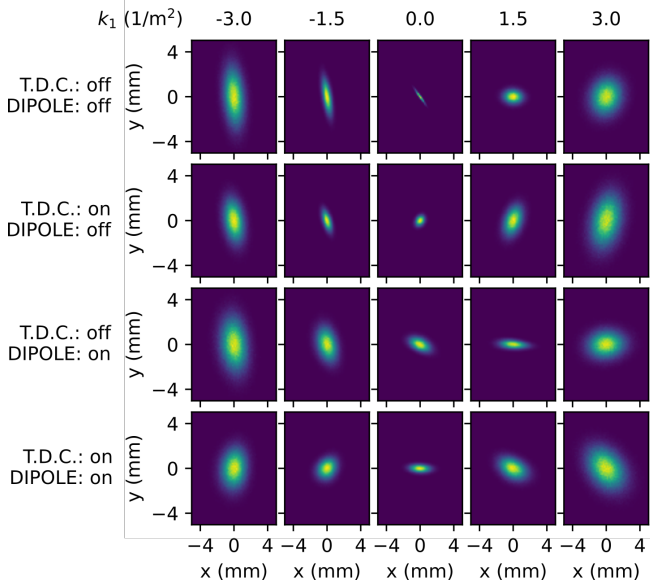


FIG. 2. Simulated screen images of the Gaussian beam distribution during the 6-dimensional reconstruction scan. Brighter colors denote higher beam intensity (arbitrary scale for each image).

dients with respect to thousands of input parameters.

Facilitating the use of backpropagation in the context of phase space reconstruction requires beam dynamics and diagnostic simulations that support the tracking of derivatives during evaluations. To this end, we have developed the simulation package *Bmad-X* [34] which re-implements beam transport through a number of simple accelerator elements as simulated by the software package *Bmad* [35]. *Bmad-X* implements the base mathematics of beam transport in Python using a package-agnostic approach, enabling arbitrary numerical computation packages to be used to do the underlying computation. This work uses the machine learning library *PyTorch* [36], which implements backpropagation, to perform beam dynamics computations. Additionally, we simulate the measurement of transverse beam profile intensity using a screen diagnostic by using kernel density estimation (KDE) [37] as opposed to normal histogramming to preserve differentiability. By using differentiable beam dynamics and diagnostic simulations we are able to cheaply compute derivatives for use in gradient descent optimization of beam distribution parameters to reconstruct the beam distribution. More recently, other, similar beam dynamics simulation packages have been developed, namely the Cheetah python package [38], although they were not used in this work.

III. 6-DIMENSIONAL DIAGNOSTIC BEAMLINE DESCRIPTION

In previous work [21], we demonstrated that this reconstruction technique is able to produce accurate pre-

dictions of the beam distribution using images gathered from a single quadrupole scan. To extend this work towards resolving 6-dimensional phase spaces, we add a transverse deflecting cavity (TDC) and a dipole spectrometer to the diagnostic beamline, shown in Fig. 1. In this case, quadrupole Q4 is scanned to measure the transverse phase space distribution, where quadrupoles Q1-Q3 are used to focus the beam onto YAG1 before Q4 is scanned. Focusing the beam on the diagnostic screen improves the measurement resolution and increases the range of Q4 strengths that can be scanned over, since the transverse beam profile needs to be kept within a region of interest on both of the diagnostic screens. Pairing a TDC that kicks the beam vertically to resolve the current profile of the beam, with a horizontally bending dipole magnet which measures beam energy spread, is a common approach taken to measure the longitudinal phase space distribution [39]. The diagnostic setup used in this work is motivated by the notion that combining transverse information from the quadrupole scan with the longitudinal phase space manipulations should provide enough information to resolve the full 6-dimensional phase space distribution.

We demonstrate the 6-dimensional reconstruction technique using a simulation of the 6-dimensional reconstruction diagnostic beamline with similar parameters to the diagnostic beamline used at the Argonne Wakefield Accelerator (AWA). The beam dynamics and screen diagnostics were simulated in the differentiable beam dynamics package *Bmad-X*. Transverse diagnostic screens YAG1 and YAG2 are placed along each beam path when the dipole is off and on respectively. These simulated screens have a region of interest that is 5 x 5 mm (or 200 x 200 pixels) in size, with a resolution of 25 $\mu\text{m}/\text{px}$. The L-band transverse deflecting cavity used in simulation (1.3 GHz, $L = 0.48$ m) operated with a peak field in a range from 0-3 MV depending on the beam distribution, consistent with the operational range of the AWA deflecting cavity [40]. The rectangular dipole spectrometer ($L = 0.3018$ m) has a bend angle of 20 degrees.

IV. SYNTHETIC RECONSTRUCTION EXAMPLES

In this section we evaluate the effectiveness of the GPSR technique on several different phase space distributions. For each case we describe below, the following procedure was followed. The first three quadrupoles (Q1, Q2, Q3) are tuned using Bayesian optimization implemented in Xopt [41], to minimize the transverse beam size at YAG2 with the scanning quadrupole (Q4) turned off. Once the range of quadrupole strengths is determined, we repeat a quadrupole scan four times, once for each combination of TDC and dipole states. Images from the quadrupole scans are then used to reconstruct the beam distribution. In the synthetic case studies that follow, the quadrupole strength was scanned over five steps,

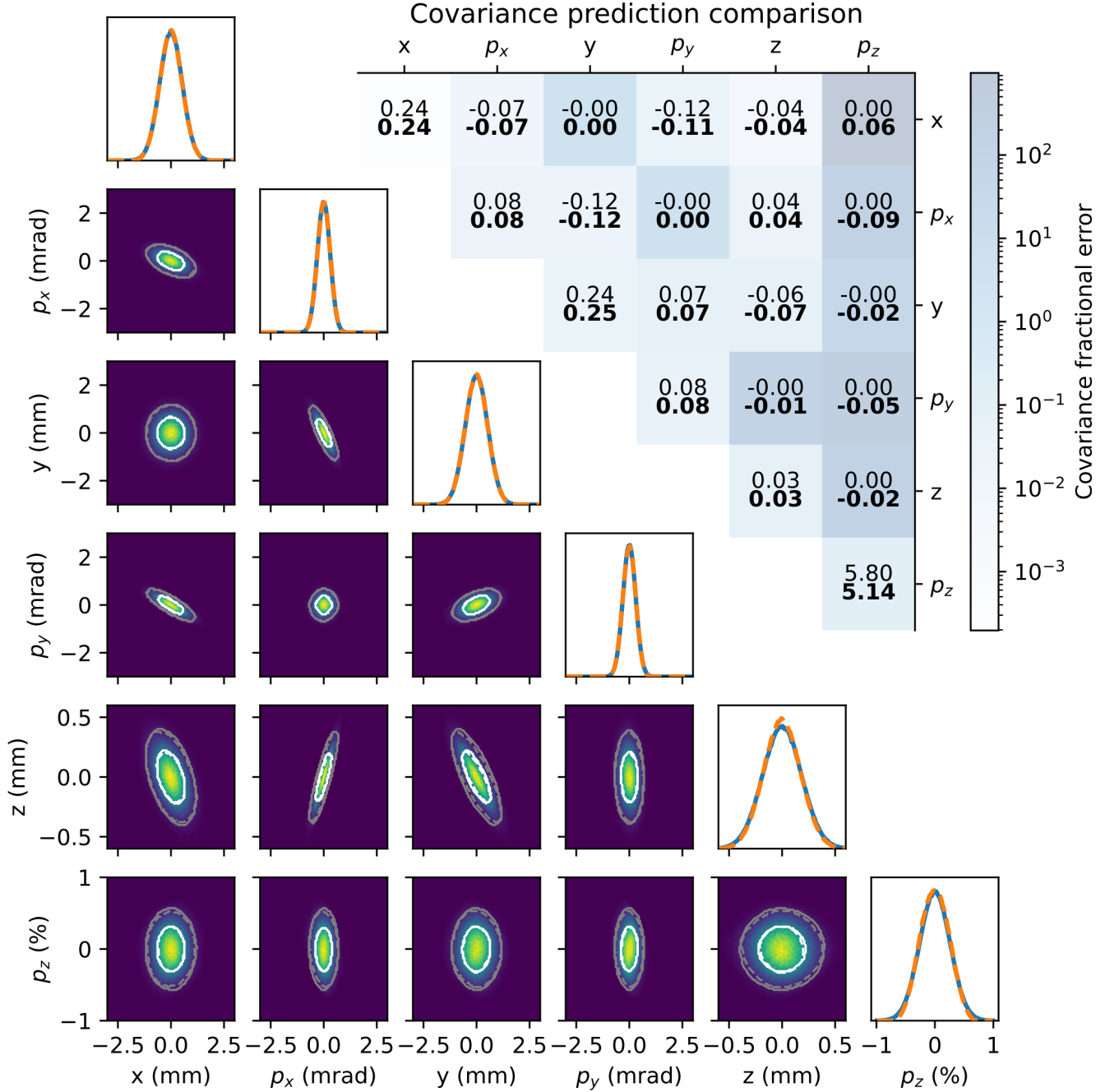


FIG. 3. Reconstruction results from Gaussian synthetic beam distribution. Lower left: Comparison between 2-dimensional projections of the ground truth synthetic beam distribution with the reconstructed beam distribution. Solid lines denote ground truth projections and contours while dashed lines denote reconstruction predictions. White and grey contours denote 50th and 90th percentile intensity levels. Color map intensity denotes reconstructed prediction. Upper right: Comparison between ground truth and reconstructed second order moments of 90th percentile beam particles, where regular text denotes ground truth values and bold text denotes reconstructed predictions.

resulting in a total data set of 20 images.

A. Case 1: Gaussian beam reconstruction

The first case we explore is one where the beam distribution is a Multivariate Normal distribution that contains cross correlations between a number of the 6-

dimensional phase space coordinates. Simulated measurements of the beam on the two diagnostic screens for each of the quadrupole scans is shown in Fig. 2. These images were then used to reconstruct the beam distribution using the GPSR approach with 100k macro particles. To maximize reconstruction accuracy, the reconstruction was trained for 3000 iterations of gradient descent which took roughly 10-15 minutes on an NVIDIA A100 GPU, although its possible that fewer iterations could be used.

The reconstruction algorithm results in a trained neural network that generates a distribution of macro-particles that should approximate the true beam distribution by matching simulated predictions to image measurements. The reconstructed beam distribution, with a comparison to the synthetic ground truth distribution, is shown in Fig. 3. We see from the 50th and 90th percentile contours (calculated by measuring the beam intensities that account for 50/90 percent of the total beam distribution) that the reconstructed beam distribution closely matches the true synthetic distribution. Furthermore, we can also compare measurements of the second order moments of the beam distribution, which also show close agreement between the reconstruction and the ground truth. In most cases the fractional error of the reconstructed second order moments is below the 10% level. In cases where the cross-covariances of the ground truth distribution are much smaller than the principal axis covariance the error of the reconstruction is higher, due to the dominance of the beam size along the principal axis over other beam features.

B. Case 2: Nonlinear beam reconstruction

We now analyze a case where the beam has a variety of non-linearities and correlations throughout the phase space distribution, as shown in Fig. 4 (simulated measurements of this beam distribution can be found in Appendix A). The reconstruction successfully replicates the ground truth distribution generated by performing 4 quadrupole scans with different states of the longitudinal phase space diagnostic elements, including non-linearities contained in cross-correlations of the beam distribution. Looking at the comparison between the predicted and ground truth values of the covariance matrix again shows good agreement within 10% for most elements.

C. Case 3: EEX beam reconstruction

Detailed high-dimensional phase space reconstructions are particularly important when performing complex beam manipulations that impart significant correlations in 6-dimensional phase space distribution and contain precisely shaped features. An example of this is transverse-to-longitudinal emittance exchange (EEX) [42]. This process combines two doglegs and a transverse deflecting cavity to map the horizontal phase space dis-

tribution into a longitudinal distribution, allowing longitudinal profile shaping using transverse masking. For example, the EEX beamline has been used to generate ramped current profiles with a sharp drop off to improve the transformer ratio of dielectric [39] and plasma [43] wakefield acceleration. Precisely characterizing beam distributions created by the EEX beamline enables us to control accelerator and beam parameters, such as the location of leafs in a multi-leaf masking element [44] or focusing magnets before the EEX beamline [42], that lead to improved 6-dimensional beam tailoring for specific accelerator applications like wakefield accelerators.

As a proof-of-concept demonstration, we generated a synthetic beam distribution which mimics those created by the EEX beamline for high transformer ratio wakefield acceleration applications. In this case, EEX aims to create a drive beam with a triangular current profile and a uniform witness beam to sample the wakefield. This is achieved using a laser cut mask to shape the horizontal beam profile before EEX, which is then mapped into a current distribution during the exchange process. However, this results in a correlation between the vertical beam size and longitudinal position within the bunch, since the vertical distribution is not exchanged like the horizontal distribution inside the EEX beamline. This can be detrimental to achieving optimal matching into a wakefield structure due to time-of-flight degradation of the longitudinal profile inside strong final-focus magnets before the wakefield device.

Reconstruction results from simulated diagnostic measurements (seen in Appendix A) of this beam distribution are shown in Fig. 5. We see that the reconstruction algorithm can resolve important features of the beam distribution, including the longitudinal profile and correlations between transverse and longitudinal phase spaces. However, the reconstruction algorithm has difficulty identifying large areas of uniform density within the beam, most notably in the $z-y$ phase space where the triangular head and the lower rectangular regions should both have near uniform density profiles. This leads to inaccurate predictions of the longitudinal beam profile, which is critical for high transformer ratio applications. Identifying how to improve the accuracy of the reconstruction in this case, either through algorithm modifications or changes in the diagnostic beamline, is a topic of future study.

V. EXPERIMENTAL DEMONSTRATION

In addition to simulation studies, we conducted an experimental study of using the 6-dimensional GPSR algorithm to characterize electron beams generated at AWA. We configured the AWA drive beamline to produce 1 nC electron bunches at a 2 Hz repetition rate with an energy of 43 MeV and transported them to the diagnostic configuration shown in Fig. 1 at the end of the AWA beamline. The beam charge was selected to provide a strong enough signal-to-noise ratio of the beam imaging

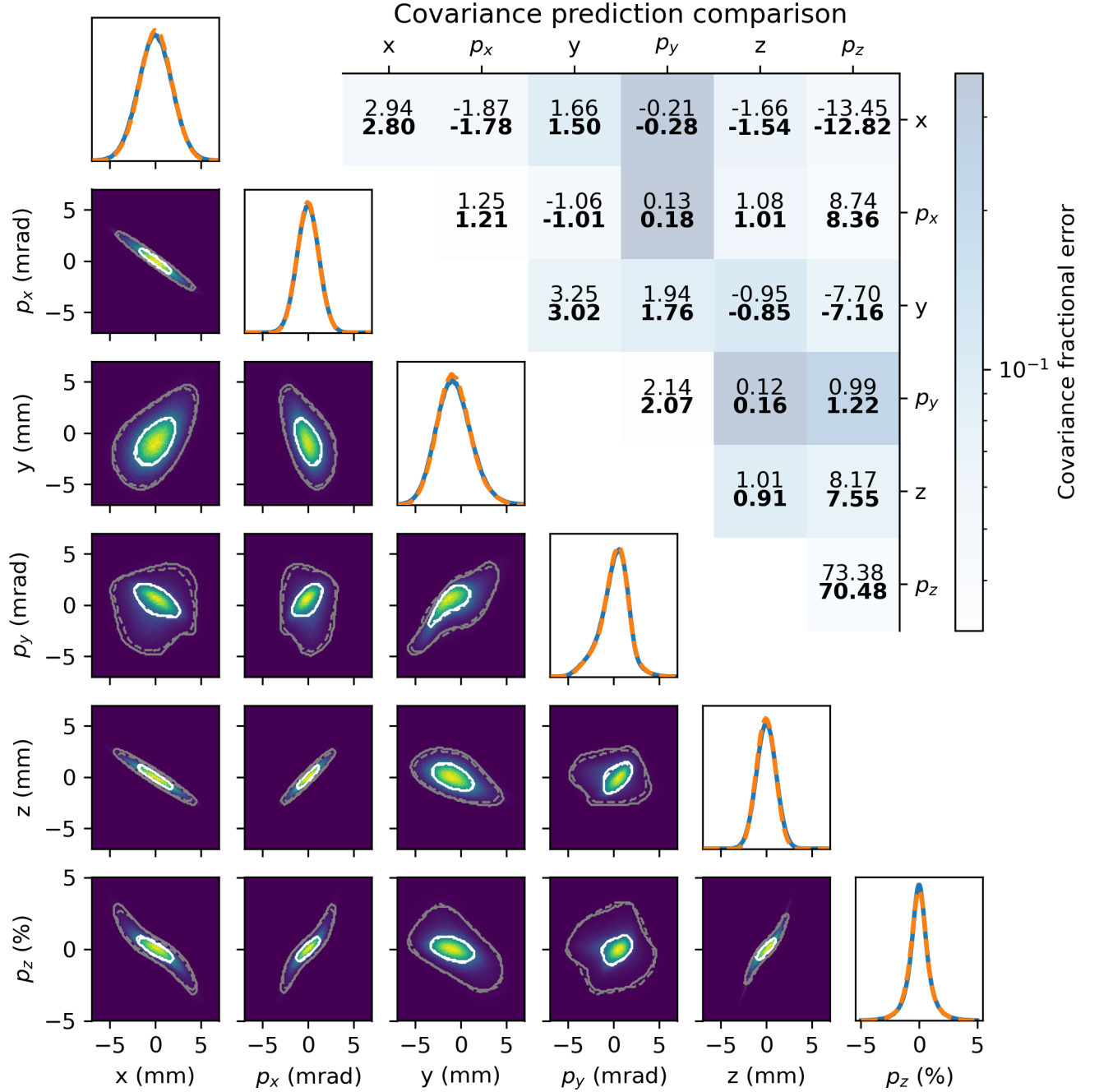


FIG. 4. Reconstruction results from non-linear synthetic beam distribution. Lower left: Comparison between 2-dimensional projections of the ground truth synthetic beam distribution with the reconstructed beam distribution. Solid lines denote ground truth projections and contours while dashed lines denote reconstruction predictions. White and grey contours denote 50th and 90th percentile intensity levels. Color map intensity denotes reconstructed prediction. Upper right: Comparison between ground truth and reconstructed second order moments of 90th percentile beam particles, where regular text denotes ground truth values and bold text denotes reconstructed predictions.

diagnostics while mitigating coherent synchrotron radiation effects in the dipole spectrometer.

Quadrupoles Q1-3 were then used to focus the beam onto YAG1 using Bayesian optimization algorithms while the scanning quadrupole Q4 was turned off. We then re-

peated 4 quadrupole scans (-2.9 T/m to 2.9 T/m, 9 steps) with different TDC and dipole settings as was done in the simulated examples. A set of 5 beam image shots were taken for each parameter setting with a charge window of 0.1 nC. The four parameter scans took approximately

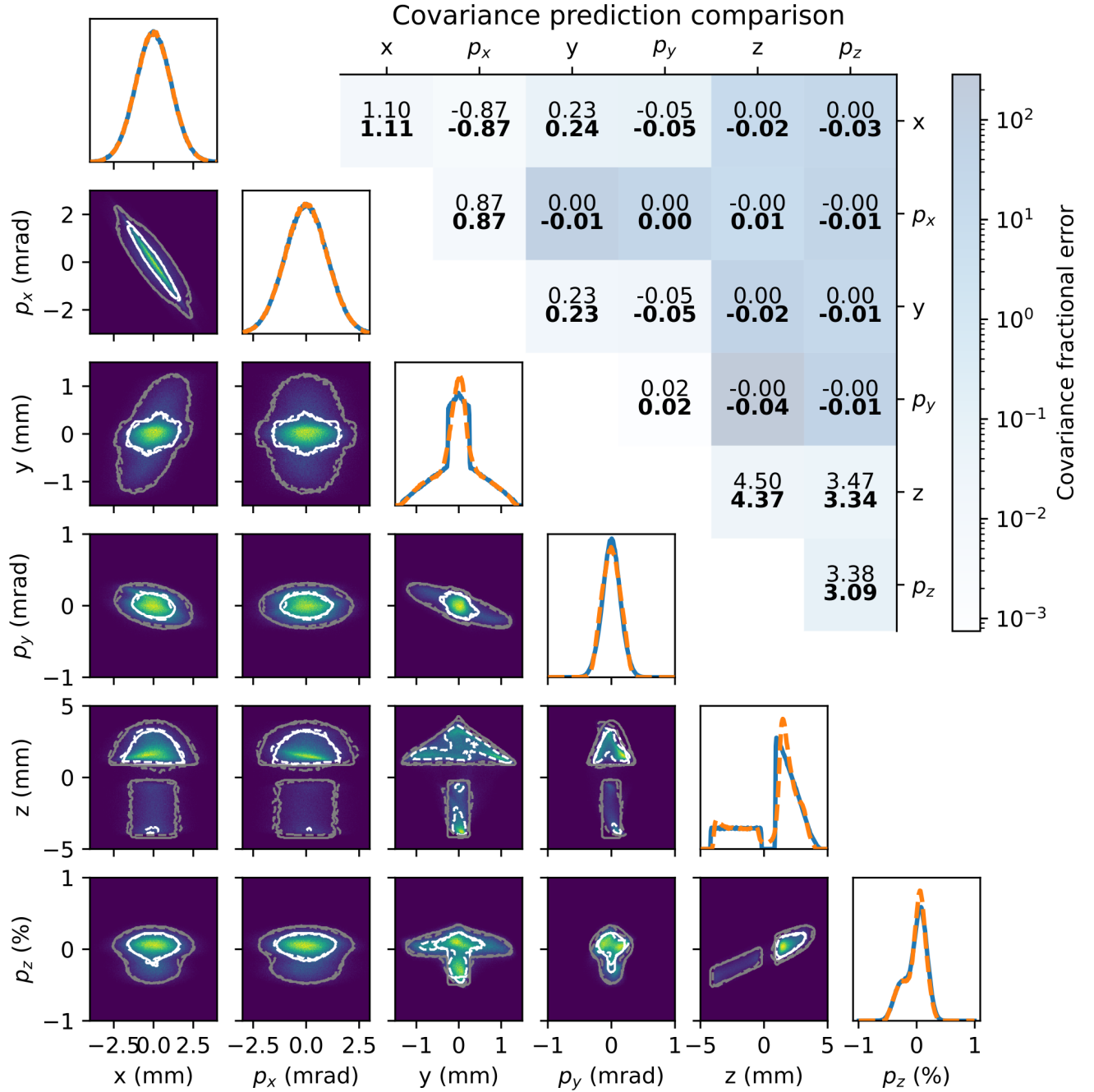


FIG. 5. Reconstruction results from the EEX synthetic beam distribution. Lower left: Comparison between 2-dimensional projections of the ground truth synthetic beam distribution with the reconstructed beam distribution. Solid lines denote ground truth projections and contours while dashed lines denote reconstruction predictions. White contours denote 90th percentile intensity levels. Color map intensity denotes reconstructed prediction. Upper right: Comparison between ground truth and reconstructed second order moments of 90th percentile beam particles, where regular text denotes ground truth values and bold text denotes reconstructed predictions.

8 minutes to perform at AWA, although we estimate that given better charge stability, the scan time could be reduced by a factor of 2.

An additional complexity of the experimental measurement was a slight difference in image resolution between

the two YAG screens used to measure the beam profile. These differences were incorporated into the reconstruction by defining two beamline simulations; one for the case where the dipole was off, and one where the dipole was on, each with different definitions for the screen di-

agnostic. The images were cropped to a size of 300 x 300 pixels, which corresponds to a side length of approximately 13 mm. Finally, the intensity of the image was clipped and set to zero at a lower bound threshold and a Gaussian smoothing filter was applied to the images to remove salt and pepper noise.

To mitigate shot-to-shot jitter of the beam centroid, we shifted the distribution of each image such that the beam centroid was at the center of the region of interest and averaged the intensity profile over the 5 shots. As a result, our reconstruction predicts the structure of the distribution, but not its offset with respect to magnetic element centers, mean energy, or timing relative to the zero-crossing of the TDC cavity. A more robust treatment of the shot-to-shot jitter would allow us to identify these aspects of the beam distribution, as was done in [21], and is a topic of future study.

In order to validate the accuracy of the reconstruction in the absence of a ground truth beam distribution, we compared model predictions to a subset of data that was not included in determining the phase space distribution (referred to here as the test data set). If the generative model can accurately predict measurements inside the test set, we can have confidence that the reconstructed distribution is accurate. We selected every other quadrupole strength in the 4 quadrupole scans to be test data, resulting in a training data set consisting of 20 images (5 quadrupole settings x 4 LPS diagnostic settings) and a test data set consisting of 16 images (4 quadrupole settings x 4 LPS diagnostic settings). The entire data set can be viewed in Appendix A. With this training data set the reconstruction took approximately 17 minutes on an A100 Nvidia GPU.

A comparison between predictions from the reconstruction and the experimentally measured test data is shown in Fig. 6. We see that the reconstruction accurately reproduces the 1D and 2D beam structure (with minor discrepancies in some cases) seen in the test images, including non-linearities and beamlet-like features in some cases. The reconstructed 6-dimensional phase space from experimental measurements is shown in Fig. 7. The reconstruction predicts normalized transverse emittances of $\varepsilon_{x,n} = 19$ mm-mrad, $\varepsilon_{y,n} = 8.5$ mm-mrad which are consistent with non-optimized AWA beamline parameters, the bunch charge of 1 nC, and the observed asymmetric beam size growth as a function of quadrupole strength when the dipole and deflecting cavity are off. One prominent feature of the distribution is the beamlet structure observed in the longitudinal current profile. We believe that this structure results from a set of 5 alpha-BBO crystals used at AWA to produce a longer, flat-top laser pulse profiles from 32 sub-pulses [45], however it is unclear how this leads to variations in beam energy along the bunch.

VI. DISCUSSION

While the examples shown here demonstrate the effectiveness of the GPSR algorithm, the reliability and robustness of the algorithm can be improved for practical applications. The most critical next step for this work is to enable the algorithm to quantify uncertainties in the reconstructed phase space distribution from limited experimental measurements, especially in the context of phase space tomography. Conventional wisdom and empirical evidence [46] suggests that evenly-spaced angular rotations of the phase space over 360 degrees is necessary to accurately reconstruct 2-dimensional phase space distributions from 1-dimensional projections. However, it is not clear how this rule-of-thumb translates to reconstructing beam distributions in 6-dimensional phase space from 2-dimensional images, considering successes demonstrated here in accurately reconstructing different beam distributions using limited quadrupole scans, and considering nonlinear transformations inside the diagnostics used here, such as chromatic focusing effects. Developing an understanding of reconstruction uncertainty would help explain why a given collection of measurements would be sufficient to reconstruct a particular beam distribution. Additionally, this would aid in further reducing the number of measurements needed to reconstruct a particular beam distribution by tracking the reduction in reconstruction uncertainty as measurements are added.

The flexibility of the GPSR algorithm can be used to develop novel beam manipulations and diagnostic techniques that were previously impractical to analyze using conventional techniques. Unlike other analysis methods, the GPSR algorithm does not require simplifications or approximations of beam dynamics or measurement signals needed to enable analytical tractability. By utilizing high-performance numerical optimization techniques, GPSR can solve extremely complex optimization problems that can incorporate exact measurement information into the reconstruction process. As a result, GPSR can also easily combine data from multiple, potentially heterogeneous sources of information (as done here, where different diagnostics have different resolutions) about the beam distribution by adding additional terms into the training loss function shown in Fig. 1. For example, screen measurements of the 2-dimensional beam profile can be easily combined with upstream measurements of the beam profile, non-destructive measurements of the beam distribution, such as edge radiation in bends [47], or ML based predictions of the longitudinal phase space [48] into a single, self-consistent description of the beam distribution.

Finally, one advantage of GPSR is that the generative model creates these macro-particles by transforming samples from a simple random distribution, as a result, predicted particle distributions can contain any number of macro-particles regardless of the number of particles used during training the model. For example,

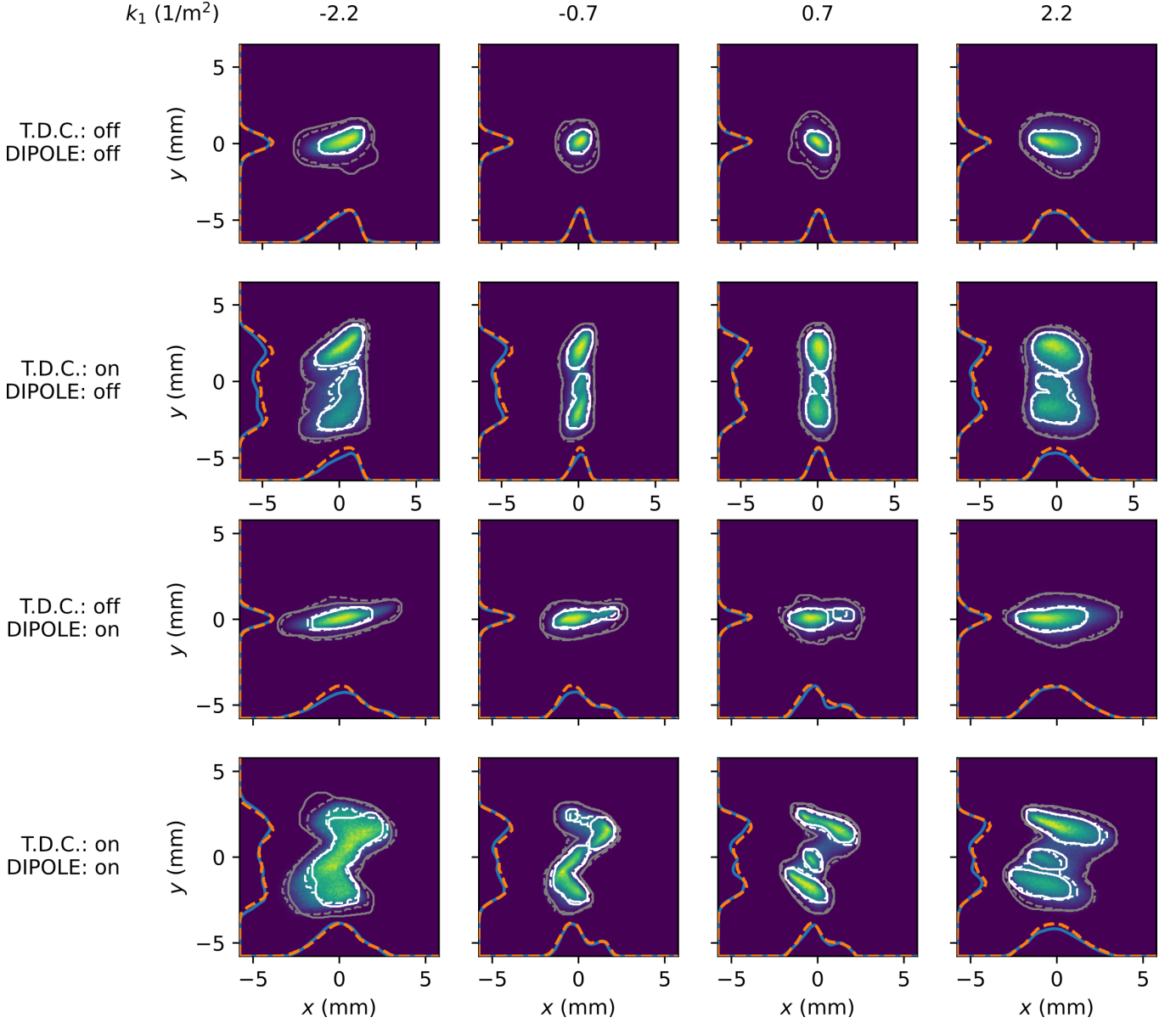


FIG. 6. Comparison between averaged, experimentally measured test images with predictions from propagating the reconstructed beam distribution to the screen diagnostics. White and grey contours denote 50th and 90th percentile intensity levels, with solid lines representing measurements and dashed lines denoting predictions. Color map intensity denotes reconstructed prediction on an arbitrary scale. Blue lines denote measured 1-dimensional projections, while orange dashed lines denote projection predictions.

even though the generative model in the cases demonstrated here was trained with 100k particles, reconstructions shown here predicts distributions containing one million macro-particles. This method for representing beam distributions is advantageous in a number of ways for use in simulations. Saving and transporting high-fidelity beam distributions containing a large number of 6-dimensional macro-particle coordinates is a memory intensive process and is restricted to a single-fidelity. Generative models on the other hand use only a few thousand scalar quantities (the weights and offsets of neural net-

work parameters) and a description of the model structure to represent the beam distribution at any level of fidelity.

VII. CONCLUSION

In this work, we have demonstrated in several simulated and experimental case studies, that detailed characterizations of 6-dimensional beam distributions can be achieved rapidly using the GPSR algorithm such that it

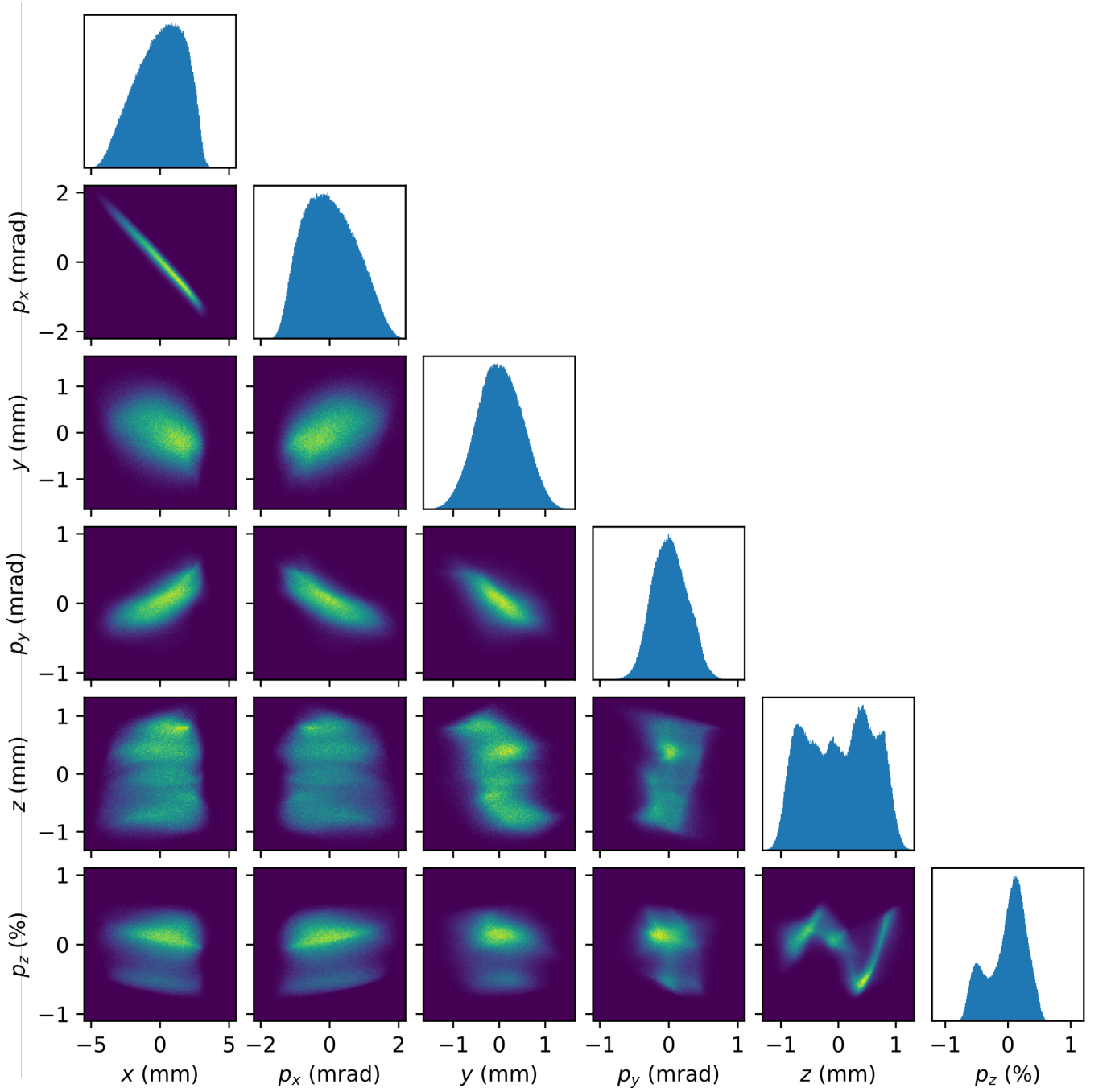


FIG. 7. Projections of the reconstructed 6-dimensional beam distributions where the color map intensity denotes beam density (independent scaling for each projection).

can be used in practical accelerator applications. This analysis method leverages the detail contained in 2-dimensional screen images of the beam distribution and knowledge of beam dynamics in the accelerator to significantly reduce (by up to a factor of 75x) the time needed to produce detailed predictions of the 6-dimensional phase space distribution. Furthermore, the generative ML representation of the beam distribution efficiently captures the complex structure of the beam distribution in the few

thousand parameters of a small neural network, which then can be used to generate macro-particle distributions which can then be used for analysis or beam dynamics simulations. This technique has major implications for allowing 6-dimensional information to be used to inform accelerator control and understand complex physical phenomena.

ACKNOWLEDGMENTS

Conceptualization, R.R., J.P.G.A., and A.L.E.; Data curation, R.R., and J.P.G.A.; Formal analysis, R.R., and J.P.G.A.; Funding acquisition, A.L.E., J.P. and Y.K.K.; Investigation, R.R. and J.P.G.A.; Methodology, R.R., T.B., J.L.M., and A.L.E.; Software, R.R., and J.P.G.A.; Experiment, R.R., J.P.G.A., W.L., E.W., and A.O.; Supervision, A.L.E., R.R., Y.K.K. and J.P.; Validation, R.R., and J.P.G.A.; Visualization, R.R., and J.P.G.A.; Writing—original draft, R.R., and J.P.G.A.; Writing—review and editing, R.R., J.P.G.A., and A.L.E. All authors have read and agreed to the published version of the manuscript.

This work is supported by the U.S. Department of Energy, Office of Science under Contract No. DE-AC02-76SF00515 and the Center for Bright Beams, NSF award PHY-1549132. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award ERCAP0020725.

Appendix A: Synthetic and Experimental Datasets

The following plots contain training data sets used to reconstruct the beam distributions described in Section IV and Section V .

-
- [1] S. Nagaitsev, Z. Huang, J. Power, J.-L. Vay, P. Piot, L. Spentzouris, J. Rosenzweig, Y. Cai, S. Cousineau, M. Conde, and others, Accelerator and beam physics research goals and opportunities, arXiv preprint arXiv:2101.04107 (2021).
 - [2] Y. S. Derbenev and A. Skrinskii, *Magnetization effects in electron cooling*, Tech. Rep. (AN SSSR, 1977).
 - [3] A. Marinelli, D. Ratner, A. A. Lutman, J. Turner, J. Welch, F. J. Decker, H. Loos, C. Behrens, S. Gilevich, A. A. Miahnahri, S. Vetter, T. J. Maxwell, Y. Ding, R. Coffee, S. Wakatsuki, and Z. Huang, High-intensity double-pulse X-ray free-electron laser, *Nature Communications* **6**, 10.1038/ncomms7369 (2015).
 - [4] C. Lindstrøm, J. Garland, S. Schröder, L. Boulton, G. Boyle, J. Chappell, R. D’Arcy, P. Gonzalez, A. Knetsch, V. Libov, G. Loisch, A. Martinez de la Ossa, P. Niknejadi, K. Pöder, L. Schaper, B. Schmidt, B. Sheeran, S. Wesch, J. Wood, and J. Osterhoff, Energy-Spread Preservation and High Efficiency in a Plasma-Wakefield Accelerator, *Physical Review Letters* **126**, 014801 (2021), publisher: American Physical Society.
 - [5] H. H. Braun, R. Corsini, L. Groening, F. Zhou, A. Kabel, T. O. Raubenheimer, R. Li, and T. Limberg, Emittance growth and energy loss due to coherent synchrotron radiation in a bunch compressor, *Physical Review Special Topics - Accelerators and Beams* **3**, 124402 (2000).
 - [6] C. A. Lindstrøm and M. Thévenet, Emittance preservation in advanced accelerators, *Journal of Instrumentation* **17** (05), P05016, publisher: IOP Publishing.
 - [7] V. Yakimenko, M. Babzien, I. Ben-Zvi, R. Malone, and X.-J. Wang, Electron beam phase-space measurement using a high-precision tomography technique, *Physical Review Special Topics - Accelerators and Beams* **6**, 122801 (2003).
 - [8] O. Sander, G. Minerbo, R. Jameson, and D. Chamberlin, *Beam Tomography in Two and Four Dimensions* (1980).
 - [9] B. Hermann, V. A. Guzenko, O. R. Hürzeler, A. Kirchner, G. L. Orlandi, E. Prat, and R. Ischebeck, Electron beam transverse phase space tomography using nanofabricated wire scanners with submicrometer resolution, *Physical Review Accelerators and Beams* **24**, 022802 (2021).
 - [10] C. B. McKee, P. G. O’Shea, and J. M. J. Madey, Phase space tomography of relativistic electron beams, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **358**, 264 (1995).
 - [11] 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, *Physical Review Accelerators and Beams* **21**, 102802 (2018).
 - [12] J. C. Wong, A. Shishlo, A. Aleksandrov, Y. Liu, and C. Long, 4D transverse phase space tomography of an operational hydrogen ion beam via noninvasive 2D measurements using laser wires, *Physical Review Accelerators and Beams* **25**, 042801 (2022).
 - [13] A. Wolski, M. A. Johnson, M. King, B. L. Militsyn, and P. H. Williams, Transverse phase space tomography in an accelerator test facility using image compression and machine learning, *Physical Review Accelerators and Beams* **25**, 122803 (2022), publisher: APS.
 - [14] A. Scheinker, Adaptive Machine Learning for Robust Diagnostics and Control of Time-Varying Particle Accelerator Components and Beams, *Information* **12**, 161 (2021), number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
 - [15] 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, *Physical Review E* **107**, 045302 (2023), publisher: American Physical Society.
 - [16] B. Marchetti, A. Grudiev, P. Craievich, R. Assmann, H. H. Braun, N. Catalan Lasheras, F. Christie, R. D’Arcy, R. Fortunati, R. Ganter, P. González Caminal, M. Hoffmann, M. Huening, S. M. Jaster-Merz, R. Jonas, F. Marcellini, D. Marx, G. McMonagle, J. Osterhoff, M. Pedrozzi, E. Prat Costa, S. Reiche, M. Reukauff, S. Schreiber, G. Tews, M. Vogt, S. Wesch, and W. Wuensch, Experimental demonstration of novel beam characterization using a polarizable X-band transverse deflection structure, *Scientific Reports* **11**, 3560 (2021).

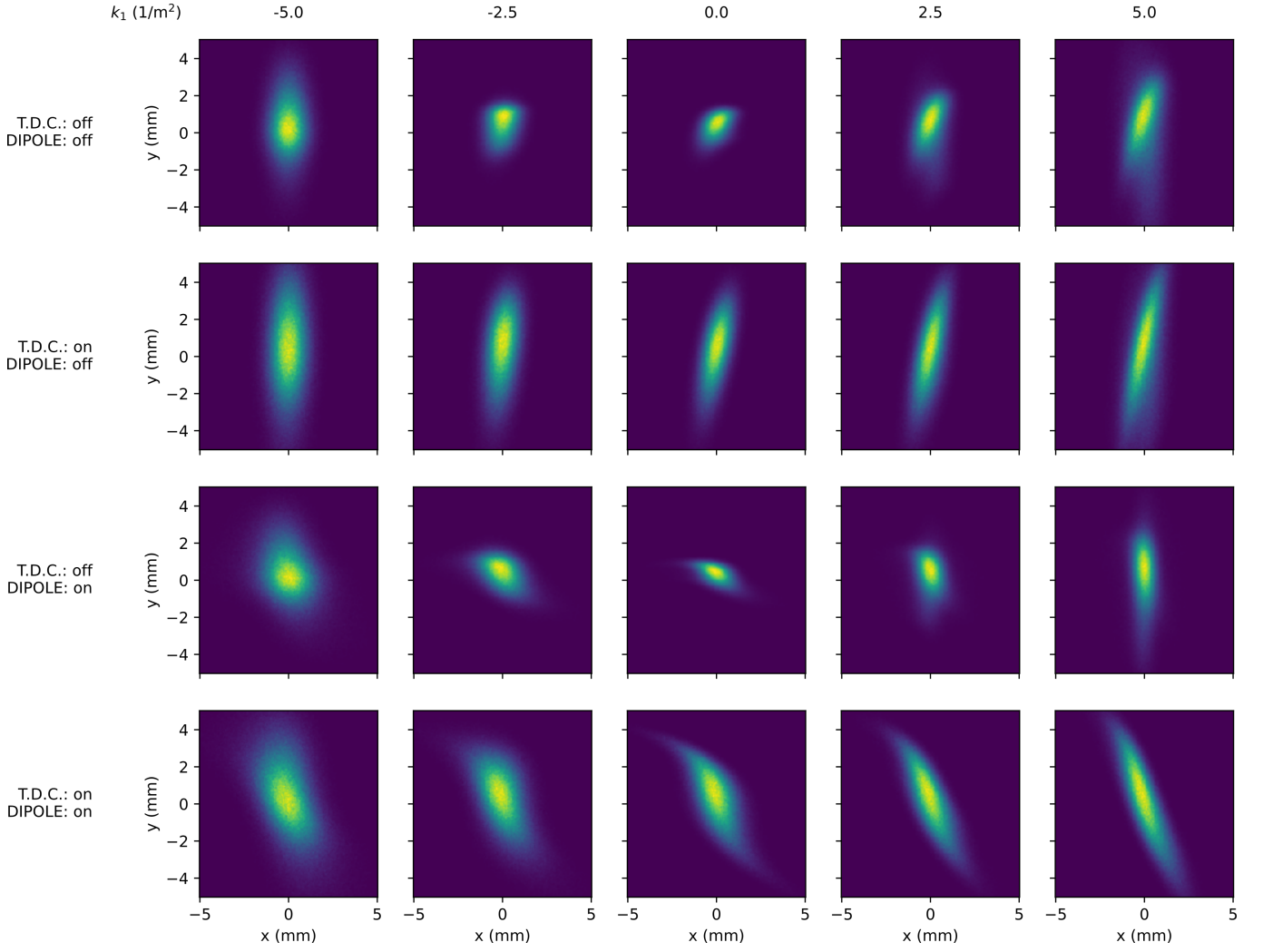


FIG. 8. Simulated screen images of the nonlinear beam distribution during the 6-dimensional reconstruction scan. Brighter colors denote higher beam intensity (arbitrary scale for each image).

- (2021), publisher: Nature Research.
- [17] S. J.-M. e. al, 5D phase-space reconstruction of an electron beam, in *Proc. IPAC'23*, IPAC'23 - 14th International Particle Accelerator Conference (JACoW Publishing, Geneva, Switzerland, 2023) pp. 2605–2608, iSSN: 2673-5490 Issue: 14.
 - [18] A. H. Andersen and A. C. Kak, Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm, *Ultrasonic Imaging* **6**, 81 (1984).
 - [19] A. Hoover, K. Ruisard, A. Aleksandrov, A. Zhukov, and S. Cousineau, Analysis of a hadron beam in five-dimensional phase space, *Physical Review Accelerators and Beams* **26**, 064202 (2023), publisher: APS.
 - [20] B. Cathey, S. Cousineau, A. Aleksandrov, and A. Zhukov, First Six Dimensional Phase Space Measurement of an Accelerator Beam, *Physical Review Letters* **121**, 064804 (2018), publisher: American Physical Society.
 - [21] 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), publisher: American Physical Society.
 - [22] Noting that a quadrupole scan does not provide information about the longitudinal phase space distribution.
 - [23] S. Kim, J. P. Gonzalez-Aguilera, P. Piot, G. Chen, S. Doran, Y.-K. Kim, W. Liu, C. Whiteford, E. Wisniewski, A. Edelen, R. Roussel, and J. Power, Four-Dimensional Phase-Space Reconstruction of Flat and Magnetized Beams Using Neural Networks and Differentiable Simulations (2024), arXiv:2402.18244 [physics].
 - [24] G. Ha, M. Cho, W. Namkung, J. Power, D. Doran, E. Wisniewski, M. Conde, W. Gai, W. Liu, C. Whiteford, Q. Gao, K.-J. Kim, A. Zholents, Y.-E. Sun, C. Jing, and P. Piot, Precision Control of the Electron Longitudinal Bunch Shape Using an Emittance-Exchange Beam Line, *Physical Review Letters* **118**, 104801 (2017).
 - [25] K. M. Hock and M. G. Ibison, A study of the maximum entropy technique for phase space tomography, *Journal of Instrumentation* **8** (02), P02003, publisher: IOP Publishing.

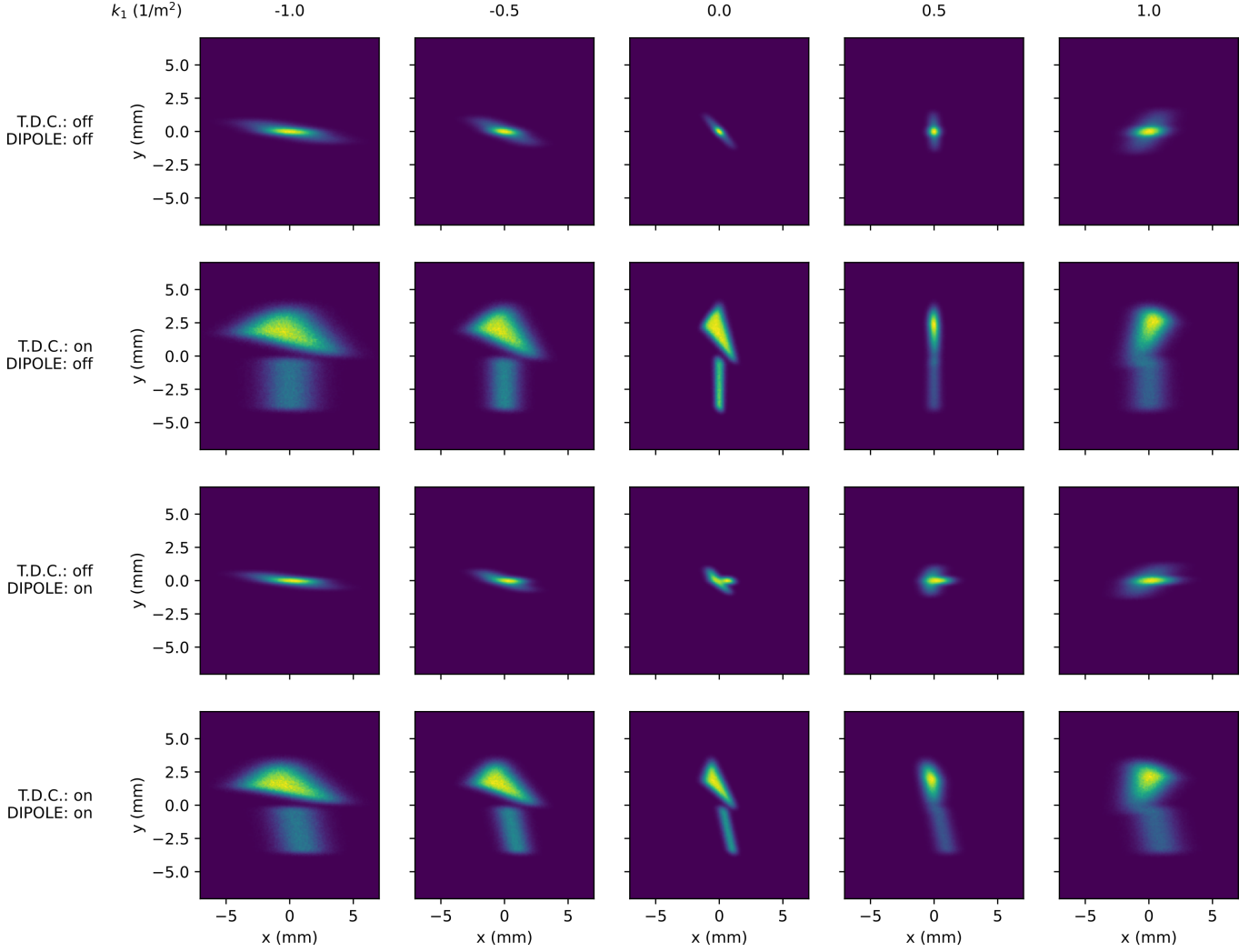


FIG. 9. Simulated screen images of the nonlinear beam distribution during the 6-dimensional reconstruction scan. Brighter colors denote higher beam intensity (arbitrary scale for each image).

- [26] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis (2020), arXiv:2003.08934 [cs].
- [27] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359 (1989).
- [28] R. Roussel, A. L. Edelen, T. Boltz, D. Kennedy, Z. Zhang, F. Ji, X. Huang, D. Ratner, A. S. Garcia, C. Xu, J. Kaiser, A. F. Pousa, A. Eichler, J. O. Lubsen, N. M. Isenberg, Y. Gao, N. Kuklev, J. Martinez, B. Mustapha, V. Kain, W. Lin, S. M. Liuzzo, J. S. John, M. J. V. Streeter, R. Lehe, and W. Neiswanger, Bayesian Optimization Algorithms for Accelerator Physics (2024), arXiv:2312.05667 [physics].
- [29] R. Bellman, Dynamic programming, *science* **153**, 34 (1966), publisher: American Association for the Advancement of Science.
- [30] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, Efficient BackProp, in *Neural Networks: Tricks of the Trade: Second Edition*, Lecture Notes in Computer Science, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer, Berlin, Heidelberg, 2012) pp. 9–48.
- [31] T. Dorigo, A. Giammanco, P. Vischia, M. Aehle, M. Bawaj, A. Boldyrev, P. de Castro Manzano, D. Derkach, J. Donini, A. Edelen, F. Fanzago, N. R. Gauger, C. Glaser, A. G. Baydin, L. Heinrich, R. Keidel, J. Kieseler, C. Krause, M. Lagrange, M. Lamparth, L. Layer, G. Maier, F. Nardi, H. E. S. Pettersen, A. Ramos, F. Ratnikov, D. Röhrich, R. R. de Austri, P. M. R. del Árbol, O. Savchenko, N. Simpson, G. C. Strong, A. Taliercio, M. Tosi, A. Ustyuzhanin, and H. Zaraket, Toward the end-to-end optimization of particle physics instruments with differentiable programming, *Reviews in Physics* **10**, 100085 (2023).
- [32] B. Martin, Differential Algebraic Techniques, in *Advances in Imaging and Electron Physics*, Vol. 108 (Elsevier, 1999) pp. 81–117.
- [33] L. Deniau and C. I. Tomoiagă, Generalised Truncated Power Series Algebra for Fast Particle Accelerator Transport Maps, in *Proc. 6th International Particle Accelerator Conference (IPAC'15), Richmond, VA, USA, May*

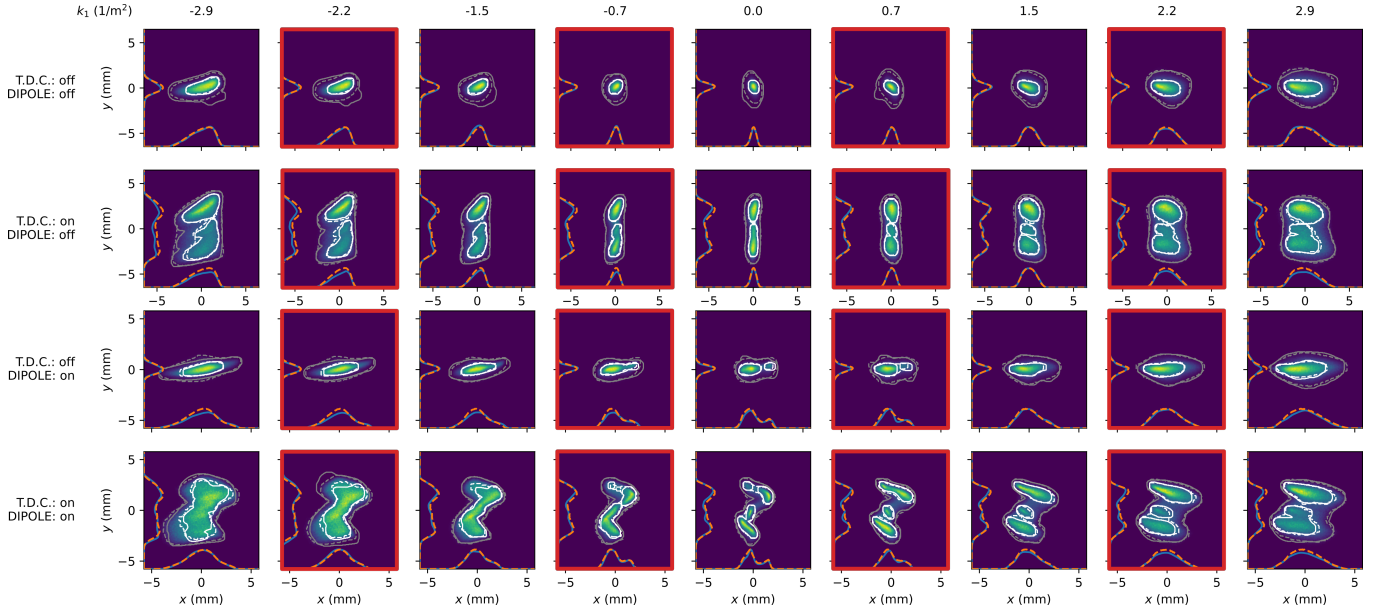


FIG. 10. Comparison between all averaged, experimentally measured images with predictions from propagating the reconstructed beam distribution to the screen diagnostics. White and grey contours denote 50th and 90th percentile intensity levels, with solid lines representing measurements and dashed lines denoting predictions. Color map intensity denotes reconstructed prediction on an arbitrary scale. Blue lines denote measured 1-dimensional projections, while orange dashed lines denote projection predictions. Images with red borders denote test images not included in determining the phase space distribution.

- 3-8, 2015, International Particle Accelerator Conference (JACoW, Geneva, Switzerland, 2015) pp. 374–377, issue: 6.
- [34] J.-P. Gonzalez-Aguilera, Y.-K. Kim, R. Roussel, A. Edelen, and C. Mayes, Towards fully differentiable accelerator modeling, in *Proc. IPAC'23*, IPAC'23 - 14th International Particle Accelerator Conference (JACoW Publishing, Geneva, Switzerland, 2023) pp. 2797–2800, iSSN: 2673-5490 Issue: 14.
- [35] D. Sagan, Bmad: A relativistic charged particle simulation library, Computational accelerator physics. Proceedings, 8th International Conference, ICAP 2004, St. Petersburg, Russia, June 29-July 2, 2004 **A558**, 356 (2006).
- [36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d. Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.
- [37] S. Weglarczyk, Kernel density estimation and its application, in *ITM web of conferences*, Vol. 23 (EDP Sciences, 2018) p. 00037.
- [38] J. Kaiser, C. Xu, A. Eichler, and A. S. Garcia, Cheetah: Bridging the Gap Between Machine Learning and Particle Accelerator Physics with High-Speed, Differentiable Simulations (2024), arXiv:2401.05815 [physics].
- [39] Q. Gao, J. Shi, H. Chen, G. Ha, J. G. Power, M. Conde, and W. Gai, Single-shot wakefield measurement system, *Physical Review Accelerators and Beams* **21**, 062801 (2018).
- [40] M. Conde, H. Chen, W. Gai, C. Jing, R. Konecny, W. Liu, D. Mihalea, P. Piot, J. Power, M. Rihaoui, and others, *Commissioning of a 1.3-GHz deflecting cavity for phase space exchange at the Argonne wakefield accelerator*, Tech. Rep. (Argonne National Lab.(ANL), Argonne, IL (United States); Fermi National . . . , 2012).
- [41] R. Roussel, C. Mayes, A. Edelen, and A. Bartnik, Xopt: A simplified framework for optimization of accelerator problems using advanced algorithms, in *Proc. IPAC'23*, IPAC'23 - 14th International Particle Accelerator Conference (JACoW Publishing, Geneva, Switzerland, 2023) pp. 4796–4799, iSSN: 2673-5490 Issue: 14.
- [42] G. Ha, M. Cho, W. Gai, K.-J. Kim, W. Namkung, and J. Power, Perturbation-minimized triangular bunch for high-transformer ratio using a double dogleg emittance exchange beam line, *Physical Review Accelerators and Beams* **19**, 121301 (2016).
- [43] R. Roussel, G. Andonian, W. Lynn, K. Sanwalka, R. Robles, C. Hansel, A. Deng, G. Lawler, J. Rosenzweig, G. Ha, J. Seok, J. Power, M. Conde, E. Wisniewski, D. Doran, and C. Whiteford, Single Shot Characterization of High Transformer Ratio Wakefields in Nonlinear Plasma Acceleration, *Physical Review Letters* **124**, 044802 (2020).
- [44] N. Majernik, G. Andonian, W. Lynn, S. Kim, C. Lorch, R. Roussel, S. Doran, E. Wisniewski, C. Whiteford, P. Piot, J. Power, and J. Rosenzweig, Beam shaping using an ultrahigh vacuum multileaf collimator and emittance exchange beamline, *Physical Review Accelerators and Beams* **26**, 022801 (2023), publisher: American Physical Society.
- [45] J. G. Power and C. Jing, Temporal Laser Pulse Shaping for RF Photocathode Guns: The Cheap and Easy way using UV Birefringent Crystals, *AIP Conference Proceed-*

- ings **1086**, 689 (2009).
- [46] K. Hock, M. Ibison, D. Holder, A. Wolski, and B. Muratori, Beam tomography in transverse normalised phase space, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **642**, 36 (2011), publisher: Elsevier.
- [47] R. Fiorito, A. Shkvarunets, D. Castronovo, M. Cornacchia, S. Di Mitri, R. Kishek, C. Tschalaer, and M. Veronese, Noninvasive emittance and energy spread monitor using optical synchrotron radiation, *Physical Review Special Topics-Accelerators and Beams* **17**, 122803 (2014), publisher: APS.
- [48] C. Emma, A. Edelen, M. Hogan, B. O'Shea, G. White, and V. Yakimenko, Machine learning-based longitudinal phase space prediction of particle accelerators, *Physical Review Accelerators and Beams* **21**, 112802 (2018).