Title
High-Performance Accelerator Modeling: Toward Improving Controls and Diagnostics for High-Brightness Beams in Experiment

Author
Cropp, Frederick

Publication Date
2023

Peer reviewed|Thesis/dissertation

FERMILAB-THESIS-2023-20

UNIVERSITY OF CALIFORNIA

Los Angeles

High-Performance Accelerator Modeling: Toward Improving Controls and Diagnostics for High-Brightness Beams in Experiment

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Physics

by

Frederick William Cropp V

2023

ABSTRACT OF THE DISSERTATION

High-Performance Accelerator Modeling: Toward Improving Controls and Diagnostics for High-Brightness Beams in Experiment

by

Frederick William Cropp V
Doctor of Philosophy in Physics
University of California, Los Angeles, 2023
Professor Pietro Musumeci, Chair

One of the main charges of beam physics is to improve the overall beam quality delivered to experiments and applications, which is generally quantified by the beam brightness. High brightness beams have numerous applications, including x-ray free electron lasers and ultrafast electron diffraction (UED). To these ends, this thesis details efforts to use high-performance models — high-fidelity models that execute quickly — for controls and diagnostics. To demonstrate the generality of these techniques, this thesis focuses on three of the most successful photoinjector designs currently used worldwide: the UCLA/SLAC/BNL-type high-gradient S-band gun, the continuous-wave high-repetition rate VHF APEX gun and the L-band DESY-PITZ-type gun. Work will be shown from Pegasus (UCLA), HiRES (LBNL) and FAST (FNAL).

Specifically, data-driven models for online virtual diagnostics are presented, in this case, in the context of UED at HiRES, leading to a temporal resolution improvement. Methods for improving the fidelity of physics-based models are discussed, with examples at HiRES, Pegasus and FAST. Markov-chain Monte Carlo analysis is applied to match simulations, in the context of photocathode studies at HiRES and Pegasus. Lastly, the augmentation of online model-based predictions with model-independent optimization is explored in a fluctuating environment at Pegasus and HiRES.

A central theme of this dissertation is working with parameter fluctuations when modeling an accelerator beamline. Long-term drifts and shot-to-shot jitter exist in every accelerator to a varying degree and therefore play an important role in every chapter of this dissertation. This thesis attempts to address the issues associated with these fluctuations when trying to develop faithful model representations of the system.

The dissertation of Frederick William Cropp V is approved.

James Rosenzweig

Anshul Kogar

Robert N. Candler

Pietro Musumeci, Committee Chair

University of California, Los Angeles

2023

To my family

Table of Contents

List of Figures

Acknowledgments

As physicists, we are taught to ask "why?" and to dig deeper and deeper until we find the answer. First and foremost, I must thank my family, particularly my mother, for fostering this curiosity from infancy. Thank you to my teachers and to my professors throughout school, many of whom had to put up with a lot of trouble thanks to my curiosity.

Thank you to my mentors in graduate school for great advice and help, including Dr. Daniele Filippetto, Dr. Alex Scheinker, Dr. Auralee Edelen and Professor Pietro Musumeci, my advisor. In particular, thank you, Pietro, for providing such great opportunities and assistance from the first day I met you on Skype as an undergraduate summer (REU) student in 2016 to now.

It was a pleasure working with my friends and colleagues in the Particle Beam Physics Laboratory at UCLA. Starting with those who mentored me as an REU student, thank you to Emma Snively, Gerard Andonian, Eylene Pirez, Ryan Roussel and in particular, Jared Maxson and David Cesar. My time with the next generation of graduate students was equally enjoyable; thank you to those with whom I had the pleasure of running Pegasus: Paul Denham, Sophie Crisp, Krish Kabra, Youna Park, Andrew Fisher, and Max Lenz. I cherish the memories of foosball and conversation in the grad student offices with those listed above, along with Gerard Lawler, Walter Lynn, Oliver Williams, David Schiller, Nathan Majernik, Josh Mann, Pratik Manwani, Victor Yu, River Robles, and others.

To my friends who matriculated with me, thank you for the good times in our first year. In particular, thank you to those with whom I studied for the comprehensive exam for the exchange of knowledge (and for our tradition of trying new restaurants and cuisines each time we went out): Eddie Chang, Lukas Lindwasser, Pratik Sathe, Kyle Ferguson and Tyler Lam.

At Lawrence Berkeley National Laboratory, I am grateful to have spent long days and nights at the lab and running HiRES with great scientists, mentors and friends: Khalid Siddiqui, Chris Pierce, Dan Durham and Daniele Filippetto. Daniele, thank you for taking

me under your wing letting me learn from you.

It was a pleasure working with colleagues at SLAC, UC Berkeley, Fermilab, LANL, Cornell, along with CBB and STROBE members and industry partners from Radiasoft, Radiabeam and Special Circumstances. In particular, in addition to all the names listed above, it was a pleasure to work with Daniel Ratner, Siqi Li, Jinhao Ruan, Jon Edelen, and Lauren Moos.

Finally, thank you to my family, my partner, Ana Esmeralda Gomez Martinez, my friends and my colleagues for helping me strike the right balance between work fun and other fun during graduate school.

And no acknowledgements section would be complete without thanking the funding agencies. This material is based partly upon work supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-SC0014664. This work was partially supported by STROBE National Science Foundation Science & Technology Center, Grant No. DMR-1548924 and NSF PHY-1549132, Center for Bright Beams.

Vita

2013 – 2017	B.S. in Physics, University of Missouri
2018	M.S. in Physics, University of California, Los Angeles (UCLA)
2021	Department of Energy, Office of Science, Graduate Student Researcher (SCGSR) Award
2017 – 2023	Ph.D. student in Physics, UCLA

Publications

K. Siddiqui, D. Durham, F. Cropp, F. Ji, S. Paiagua, C. Ophus, N. Andresen, L. Jin, J. Wu, S. Wang et al., "Relativistic ultrafast electron diffraction at high repetition rates," **Submitted to** *Structural Dynamics*, arXiv preprint arXiv:2306.04900, 2023.

F. Cropp, L. Moos, A. Scheinker, A. Gilardi, D. Wang, S. Paiagua, C. Serrano, P. Musumeci, and D. Filippetto, "Virtual-diagnostic-based time stamping for ultra- fast electron diffraction," *Physical Review Accelerators and Beams*, vol. 26, no. 5, p. 052801, 2023.

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*, vol. 107, p. 045302, Apr 2023. [Online].

F. Cropp, D. Filippetto, A. Gilardi, P. Musumeci, S. Paiagua, A. Scheinker, D. Wang et al., "Toward machine learning-based adaptive control and global feedback for compact accelerators," in *Proceedings of International Particle Accelerator Conference 2022*, Bangkok, Thailand, 2022

A. Scheinker, D. Filippetto, F. Cropp, "6D Phase space diagnostics based on adaptively tuned physics-informed generative convolutional neural networks," in Proceedings of *International Particle Accelerator Conference 2022*, Bangkok, Thailand, 2022

A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, "An adaptive approach to machine learning for compact particle accelerators," *Scientific Reports*, vol. 11, no. 1, p. 19187, Dec. 2021. [Online].

K. M. Siddiqui, D. B. Durham, F. Cropp, C. Ophus, S. Rajpurohit, Y. Zhu, J. D. Carlström, C. Stavrakas, Z. Mao, A. Raja et al., "Ultrafast optical melting of trimer superstructure in layered 1T'-TaTe2," *Communications Physics*, vol. 4, no. 1, p. 152, 2021

K. M. Siddiqui, D. Durham, F. Cropp, S. Rajpurohit, C. Ophus, Y. Zhu, J. Carlström, C. Stavrakas, Z. Mao, A. Raja et al., "Ultrafast photo-induced melting of the trimer superstructure in TaTe2," in *CLEO: QELS Fundamental Science*. Optica Publishing Group, 2021

K. M. Siddiqui, D. B. Durham, F. Cropp, A. Schmid, P. Musumeci, A. M. Minor, R. A. Kaindl, and D. Filippetto, "Ultrafast structural dynamics of materials captured by relativistic electron bunches," in *Ultrafast Nonlinear Imaging and Spectroscopy VIII*, vol. 11497. SPIE, 2020

K. M. Siddiqui, D. Durham, F. Cropp, S. Rajpurohit, C. Ophus, Y. Zhu, J. Carlström, C. Stavrakas, Z. Mao, A. Raja et al. "Visualizing the melting of periodic lattice distortions in a complex 2D charge density wave material via MeV-scale ultrafast electron diffraction," in Proceedings of *Ultrafast Phenomena and Nanophotonics XXV*, 2021

K. Kabra, S. Li, F. Cropp, T. J. Lane, P. Musumeci, and D. Ratner, "Mapping photocathode

quantum efficiency with ghost imaging," *Physical Review Accelerators and Beams*, vol. 23, no. 2, p. 022803, 2020.

F. Cropp, N. Burger, P. Denham, J. G. Navarro, E. Liu, P. Musumeci, L. Phillips, A. Edelen, and C. Emma, "Maximizing 2-D beam brightness using the round to flat beam transformation in the ultralow charge regime," in Proceedings of *North American Particle Accelerator Conference*, Lansing, MI, 2019.

S. Li, F. Cropp, K. Kabra, T. Lane, G. Wetzstein, P. Musumeci, and D. Ratner, "Electron ghost imaging," *Physical Review Letters*, vol. 121, no. 11, p. 114801, 2018

CHAPTER 1

Introduction

From their inception, the development of particle accelerators has been motivated by and has resulted in new scientific discoveries. The first accelerators, such as the cathode ray tube, which would pale in comparison to modern accelerators — in terms of precision, energy and capability — bequeathed fundamental discoveries such as Thomson's discovery of the electron [2]. Notably, these kinds of fundamental discoveries also motivated improvements to the mechanism for discovery itself: the particle accelerator. In 1930, the first modern particle accelerator, the cyclotron, was born [3, 4], which shepherded in a new era of high energy beams, and led to numerous discoveries such as nuclear fission research and new elements (e.g. [5]). Newer discoveries, such as the recent discovery of the Higgs Boson [6], were enabled by beams with even higher energies.

However, it is not just higher beam energies that have enabled historical and current experiments and discoveries. Having a high quality beam is also important. In general, beam quality can be thought of as how "ordered" the beam is, or how small the intrinsic angles are. This can be visualized by the divergence of a small, collimated beam; a high quality beam will remain small in this case, while a lower quality beam will diverge due to the internal disorder — or angles — in the beam. In addition, holding other properties constant, a small spot size is desirable. The angles and spot size are generally put together into a quantity called emittance, which will be introduced formally below. Finally, holding other properties constant, having a high charge is also desirable. As such, the ratio of charge to emittance is defined as the beam brightness often cited as the definitive metric of beam quality (and will also be more formally defined below).

Beam brightness has also improved over the years in part because of the technology that allows for higher beam energies, which allow for relativistic suppression of electron Coulomb repulsion. More recent improvements include the introduction of the RF photoinjector [7, 8, 9] and associated improvements in the generation and transport of beams [10]. The resulting high brightness beams have allowed for the development of new means of scientific discovery, including x-ray free-electron lasers [11, 12] and ultrafast electron diffraction (UED) (e.g. [13, 14, 15, 16, 17, 18, 1]), the latter of which will feature prominently in this thesis, in particular in Chapter 2 and some in Chapter 4. Not only did high brightness beams allow for the development of these techniques, these techniques are also applications of the state-of-the-art high brightness beams.

As described above, progress toward making higher brightness beams has been made by leveraging new technologies, ultimately toward better control of beams. Usually, better control of beams requires an improvement of the diagnostics that inform the control policy. In other words, the capability to produce high brightness beams is due to scientists and engineers' use of new technologies to improve controls and diagnostics for particle accelerators. This dissertation concerns the application of novel contemporary tools to improve models of particle accelerators in order to improve controls and diagnostics for particle accelerators and ultimately, to improve the brightness of electron beams.

1.1 High Brightness Beams

1.1.1 Definitions

Beam brightness is generally defined as proportional to the amount of charge that can fit into a given phase space area. As beam emittance is a measure of the area in phase space occupied by the beam, this definition can be given as:

$$B_{\{2D,4D,6D\}} = \frac{Q}{\epsilon_{\{2D,4D,6D\}}} \tag{1.1}$$

where Q is the charge of the beam, while ϵ is the beam emittance. A discussion of the dimensions (2D, 4D or 6D) is below. The emittance can be defined in a geometric sense as

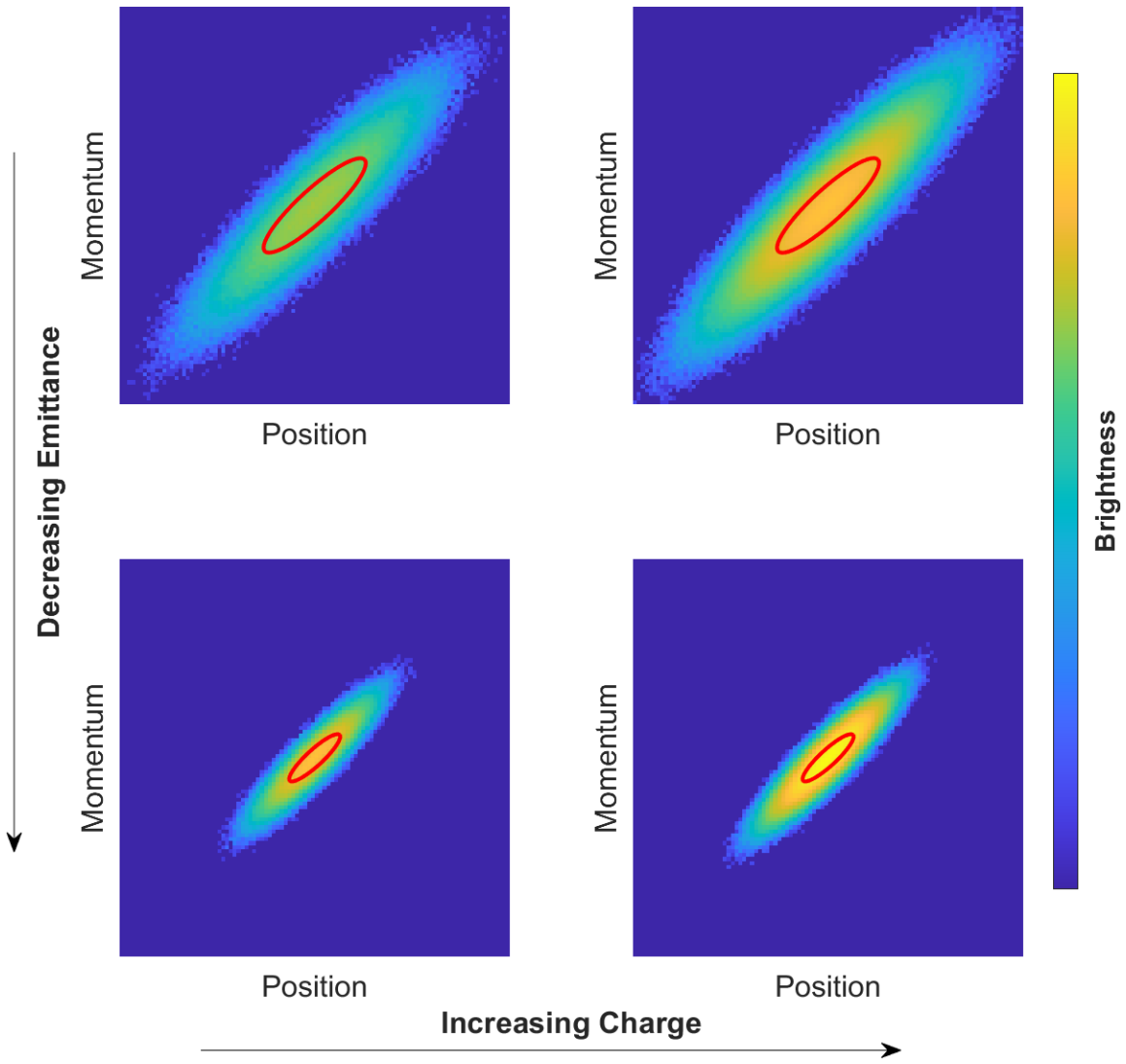


Figure 1.1: Beam brightness increases from top to bottom and left to right. The red ellipse encloses one standard deviation of the distribution and the area inside therefore corresponds to the RMS emittance.

the area enclosed by the red in Fig. 1.1. However, as can be seen, some particles are outside the red ellipse. This is because the a particle distribution can also be treated in a statistical sense. An electron bunch can be defined by a beam distribution function $f(x, x', y, y', z, \delta)$, which is a probability density function in six dimensions. These are the position coordinates

(x,y,z) and the normalized momenta, where

$$x' = \frac{P_x}{P_z} = \frac{\gamma\beta_x mc^2}{\gamma\beta_z mc^2} = \frac{\beta_x}{\beta_z} \tag{1.2}$$

and likewise

$$y' = \frac{\beta_y}{\beta_z} \tag{1.3}$$

where $\vec{\boldsymbol{\beta}}$ is the normalized velocity and $(\beta_x,\beta_y,\beta_z)$ are its cartesian decomposition. The normalized momentum in the longitudinal coordinate is given as

$$\delta = \frac{P_z - \langle P_z \rangle}{P_z} \tag{1.4}$$

where P_z is the longitudinal momentum and $\langle P_z \rangle$ is the average longitudinal momentum of the beam.

As a statistical distribution, the true emittance can be difficult to define, as many distributions, such as a gaussian beam distribution function, will have tails that go to infinity. As such, the RMS emittance is an important concept, where this is the area that encloses all particles within one standard deviation of the mean particle in all dimensions. The RMS emittance is shown by the red ellipse in Fig. 1.1 and differs from the true emittance, which is the smallest ellipse that would enclose all of the particles — or in other words, the total amount of phase space that the beam occupies. In general, it is therefore more practical to use the RMS emittance than the true emittance. Other fractions of the amount of the beam enclosed can be used; metrics such as the 95% emittance exist, but this thesis will use the RMS emittance for convenience. The convenience of this notation will be shown in the remainder of this section. A further discussion of the conservation of the RMS emittance and the true emittance is discussed in Section 1.1.2.

The RMS emittance can be thought of conveniently using the covariance matrix of the beam particle coordinates, hereafter referred to as the "beam matrix."

$$\Sigma^{6D} = \begin{pmatrix} \langle xx \rangle & \langle xx' \rangle & \langle xy \rangle & \langle xy' \rangle & \langle xz \rangle & \langle x\delta \rangle \\ \langle x'x \rangle & \langle x'x' \rangle & \langle x'y \rangle & \langle x'y' \rangle & \langle x'z \rangle & \langle x'\delta \rangle \\ \langle xy \rangle & \langle x'y \rangle & \langle yy \rangle & \langle yy' \rangle & \langle yz \rangle & \langle y\delta \rangle \\ \langle xy' \rangle & \langle x'y' \rangle & \langle yy' \rangle & \langle y'y' \rangle & \langle y'z \rangle & \langle y'\delta \rangle \\ \langle xz \rangle & \langle x'z \rangle & \langle yz \rangle & \langle y'z \rangle & \langle zz \rangle & \langle z\delta \rangle \\ \langle x\delta \rangle & \langle x'\delta \rangle & \langle yz\delta \rangle & \langle y'\delta \rangle & \langle z\delta \rangle & \langle \delta\delta \rangle \end{pmatrix} \tag{1.5}$$

The beam matrix is composed of the second order moments of the beam distribution function, for example:

$$\langle xy \rangle = \int xyf(x, x', y, y', z, \delta) dx dx' dy dy' dz d\delta \tag{1.6}$$

In general, the RMS geometric emittance is defined as the square root of the determinant of the beam matrix:

$$\epsilon_{6D} = \sqrt{\det\left(\Sigma^{6D}\right)} \tag{1.7}$$

An astute reader would note that in Fig. 1.1, the emittance is defined by the area in phase space, i.e. position and momentum coordinates, while in Eq. 1.7, the emittance appears to be defined in terms of position and some angle, rather than momentum. Indeed, this is related to the fact that Eq. 1.7 is defined in trace space, which does indeed use angles rather than momenta. Note that the **geometric** emittance is defined in Eq. 1.7, whereas the **normalized** emittance is defined as:

$$\epsilon_{N,6D} = (\gamma\beta_z)^3 \sqrt{\det\left(\Sigma^{6D}\right)} \tag{1.8}$$

where $\gamma\beta_z$ is the average normalized longitudinal momentum. This can be shown to be equivalent in cases of low spread in P_z to an analogous calculation of emittance using a modified beam matrix in phase space, rather than trace space, where the momenta are not all divided by P_z.

In addition, in Fig. 1.1, the emittance appears to be defined in two dimensions, rather than six, as shown in 1.7. This is related to the dimensionality of the problem. There are equivalent definitions of emittance in 2D and 4D. The beam matrix is defined as the required

blocks of the 6D beam matrix. For example, the 2D emittance in the x-coordinate plane is the upper left 2x2 block of the 6D beam matrix:

$$\Sigma_x^{2D} = \begin{pmatrix} \langle xx \rangle & \langle xx' \rangle \\ \langle x'x \rangle & \langle x'x' \rangle \end{pmatrix} \tag{1.9}$$

and the normalized emittance is:

$$\epsilon_{N,2D} = \gamma\beta_z \sqrt{\det\left(\Sigma^{2D}\right)} \tag{1.10}$$

Likewise, the 2D emittance can be defined in the y- and z-coordinate planes.

The 4D emittance is usually defined in the transverse coordinate plane, such that

$$\Sigma^{4D} = \begin{pmatrix} \langle xx \rangle & \langle xx' \rangle & \langle xy \rangle & \langle xy' \rangle \\ \langle x'x \rangle & \langle x'x' \rangle & \langle x'y \rangle & \langle x'y' \rangle \\ \langle xy \rangle & \langle x'y \rangle & \langle yy \rangle & \langle yy' \rangle \\ \langle xy' \rangle & \langle x'y' \rangle & \langle yy' \rangle & \langle y'y' \rangle \end{pmatrix} \tag{1.11}$$

and the normalized emittance is:

$$\epsilon_{N,4D} = (\gamma\beta_z)^2 \sqrt{\det\left(\Sigma^{4D}\right)} \tag{1.12}$$

This leads to an important point: it is often assumed that the planes are separable, or the off-diagonal blocks of Eq. 1.5 are composed of zeros. This means that the correlation between the planes is zero. In general, for an idealized beam, this is true, but in this thesis, two examples will be clearly indicated (in Sections 3.3 and 4.2) where this assumption does not hold. Where the assumption of separability does hold,

$$\epsilon_{4D} = \epsilon_{x,2D}\epsilon_{y,2D} \tag{1.13}$$

and

$$\epsilon_{6D} = \epsilon_{x,2D}\epsilon_{y,2D}\epsilon_{z,2D} \tag{1.14}$$

Hereafter, the normalized 2D emittances will be written as ϵ_x, ϵ_y and ϵ_z for brevity. As such, there are different definitions of beam brightness, usually clearly defined:

- 2D beam brightness: $B_{2D}=\frac{Q}{\epsilon_\xi}$, where $\xi \in \{x, y, z\}$
- 4D beam brightness: $B_{4D}=\frac{Q}{\epsilon_{4D}}$. If separable, $B_{4D}=\frac{Q}{\epsilon_x,\epsilon_y}$
- 5D beam brightness: $B_{5D} = \frac{I}{\epsilon_{4D}}$. If separable, $B_{5D}=\frac{I}{\epsilon_x,\epsilon_y}$
- 6D beam brightness: $B_{6D}=\frac{Q}{\epsilon_{6D}}$. If separable, $B_{6D}=\frac{Q}{\epsilon_x,\epsilon_y\epsilon_z}$

where the normalized emittances are used in these equations, and Q is the charge and I is the current. The 5-D beam brightness is an often-cited quantity in beam physics, for example in the determination of the efficiency, gain length and saturation power of free electron lasers [19, 20]. Of course, there is the consideration, particularly for the 5-D brightness, of peak versus average brightness. Since the charge in a beam passing a certain point is not uniform in time, a different current can be obtained based on the time window used to calculate it. The average current is the total charge that passes a point divided by the total beam bunch length (in seconds), while the peak current indicates the current in a small time window with the most charge passing a given point. Various normalization factors can be applied to the definitions given above, in order to agree with various conventions in different fields. For the remainder of this work, the above definitions — with the above normalization factors — will apply. In general, accelerator improvements are measured by how much they improve the beam brightness, but it is worth carefully considering which definition of beam brightness is improved for a given problem.

1.1.2 Beam Production and Control

Improving beam brightness can be thought of in two broad arcs:

1. improving source brightness
2. improving beam transport in order to maximize the beam brightness at a point downstream, where a diagnostic or experiment is positioned.

This thesis pertains almost exclusively to the latter, but it is worth stating why the former is critical to the work presented herein.

The initial source quality is generally quantified by two very related metrics: the thermal emittance and the mean transverse energy (MTE). This thesis mostly refers to the latter, but it is worth showing the equivalence. Both seek to quantify the angles of the beam as it is produced. The thermal emittance is simply the emittance of the beam before any deleterious effects such as space charge, RF curvature or nonlinear magnet aberrations can increase the emittance. The normalized thermal emittance can be given as [21, 22]:

$$\epsilon_x = \sqrt{\langle xx \rangle \frac{MTE}{mc^2}} \tag{1.15}$$

Working in the x transverse plane without loss of generality, the spot size term, $\langle xx \rangle$ is easy to determine using a virtual cathode camera (a diagnostic described below). In a beam emitted by near-threshold photoemission, the MTE is given by:

$$MTE = \frac{E_{excess}}{3} \tag{1.16}$$

where E_{excess} is the excess energy beyond the work function:

$$E_{excess} = h\nu - \phi_{eff} \tag{1.17}$$

where ϕ_{eff} is the work function of the material after accounting for the tunneling probability (Schottky effect), which is described in detail in Section 3.3.

From the photocathode, the beam is transported downstream, a process that is more heavily emphasized in this thesis. To a linear approximation, each particle is transported downstream as follows:

$$\begin{pmatrix} x_f \\ x'_f \\ y_f \\ y'_f \\ z_f \\ \delta_f \end{pmatrix} = R \cdot \begin{pmatrix} x_i \\ x'_i \\ y_i \\ y'_i \\ z_i \\ \delta_i \end{pmatrix} \tag{1.18}$$

where the linear transport matrix R evolves the particle coordinates from the initial state to the final state. These transport maps are symplectic [23] and have determinant unity. Notably, the R_{21}, R_{43} and R_{65} elements control the focusing of the element in the x, y and z planes. The R_{12}, R_{34} and R_{56} elements are critical to the evolution of the particle, as it links the angles to the downstream coordinates.

Although single-particle dynamics are critical, the transport matrices can evolve the second order moments of the beam matrix as shown in Eq. 1.5 as it goes through various beam elements. The beam matrix Σ_f can be computed from the initial beam matrix Σ_i as follows:

$$\Sigma_f = R \cdot \Sigma_i \cdot R^T \tag{1.19}$$

These transport matrices can likewise be defined in 2-D and 4-D, as required.

Again, this transport map is shown to work on the RMS quantities as described in Section 1.1.1. But when transported downstream, there is a fundamental difference between the RMS emittance and the true emittance, in terms of conservation. Liouville's theorem in Hamiltonian mechanics states that so long as a system can be described by a Hamiltonian, then the phase space area is conserved. Thus, the true emittance (total phase space area) is always conserved when the beam dynamics can be described by a Hamiltonian. Under most circumstances in a linear accelerator, this is the case, but examples of cases where this assumption is untrue include intrabeam scattering [24, 25] and coherent synchrotron radiation (e.g. [26, 27].

The RMS emittance is conserved under linear transformations. This can be seen easily by taking the determinant of Eq. 1.19 and recalling that the determinant of the transport matrix is unity:

$$\det\left(\Sigma_f\right) = \det\left(R \cdot \Sigma_i \cdot R^T\right) = \det\left(R\right) \cdot \det\left(\Sigma_i\right) \cdot \det\left(R^T\right) = \det\left(\Sigma_i\right) \tag{1.20}$$

RMS emittance is not in general conserved, however. Even in the presence of nonlinear space charge forces, as are experienced in any accelerator, there is RMS emittance growth

(e.g. [28]). The best that can be expected is that the emittance is conserved during transport. Thus the source brightness sets the upper limit on beam brightness in general, so the work presented herein is given more meaning with better source brightness.

While the 6-D true emittance is a constant of motion in Hamiltonian systems, and the 6-D RMS emittance is a constant of motion for linear systems, the 4-D emittance (and longitudinal emittance) are only constants if the transport is separable (see Section 1.1.1) and thus both are independently constant. Similarly, if the transport is uncoupled between the transverse planes, the 2-D RMS emittance is constant in a linear system. This thesis includes work on the round-to-flat beam transform (FBT) in Chapter 4, which is a technique to increase beam brightness in one transverse plane at the expense of another, with a clear coupling between the two transverse planes.

In addition to "borrowing" emittance from one plane at the expense of another, this thesis also contains work toward increasing peak beam brightness without regard to average brightness. However, these "loans" of brightness allow for certain applications, but in general, the limitation on the beam brightness set by the source remains.

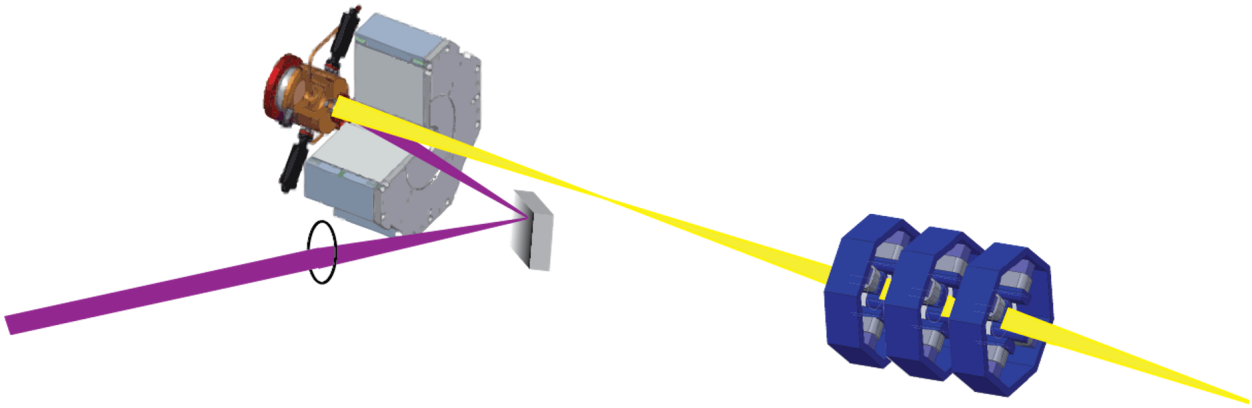


Figure 1.2: A cartoon of an RF photoinjector: a laser pulse is imaged onto a photocathode. The resultant electron beam is focused by an emittance compensation solenoid and additional magnets downstream, exemplified herein by three quadrupole magnets.

1.1.2.1 RF Photoinjectors

The transport of high brightness beams to minimize the emittance growth described above is not an easy feat. Downstream of the production of high brightness beams, numerous techniques and strategies are implemented in order to preserve the beam brightness & emittance. For example, generally, high charge beams are rapidly accelerated to high energies in order to minimize the effect of the nonlinear space-charge forces. The desire to do so in part motivated the development of the RF photoinjector [7, 9, 8], which is the state-of-the-art in high brightness beam production. The photoinjector is the type of electron source that this thesis will focus. A cartoon of which is shown in Fig. 1.2.

In an RF photoinjector, one of the main sources of high-brightness electron beams, on which this thesis will focus, a laser is shone onto the surface of a photocathode, which forms the part of the back-plane of the RF gun. The RF gun accelerates the beam to high energies by exposing the photocathode to a sinusoidally-varying longitudinal electric field. The incident laser pulse must arrive during the time the electric field will accelerate the emitted electrons downstream, rather than back into the photocathode. In general, the laser pulse must be short enough in duration to not sample the curvature of the sinusoidal RF-fields in the gun [29]. An emittance-compensation solenoid then follows the gun, which can focus the beam, as well as line up the different slices in phase space, in order to reduce the overall emittance of the beam [30]. Additionally, a bucking solenoid can be used to minimize the initial correlations in the beam (see description in previous section) (e.g. [31]). Downstream of the gun, optics are used to minimize dispersion (e.g. [1]), focus the beam to the proper size, and in some cases remove correlations in the beam matrix (e.g. [32, 33, 34, 35, 36, 37, 31, 38, 39]) before being used for an experiment, impacting destructive diagnostics, and/or going to a beam dump.

1.2 Modeling for Diagnostics & Control

As can be seen from the description of an RF photoinjector above, there are numerous parameters to optimize, and running an accelerator quickly becomes a complex control problem. However, any control policy is only as good as the diagnostics that inform it. Diagnostics allow users and operators to have knowledge about properties of the beam. In the following paragraphs, diagnostics that are relavant to this thesis are listed.

The simplest example is a scintillating screen. The electron beam impacts the screen, and where it does so, photons are emitted and imaged by a camera. This allows the user and/or operator to see the transverse properties of a beam. A screen naturally integrates over the angle coordinates; it takes only a snapshot in time of the particle positions. In the longitudinal direction, the screen and camera combination are generally considerably slower than the bunch length of the beam, so an integration in z is also performed. As such, the integrals in the form of Eq. 1.6 become much easier:

$$\langle x^2 \rangle = \iint x^2 f(x,y) dx dy \tag{1.21}$$

$$\langle y^2 \rangle = \iint y^2 f(x,y) dx dy \tag{1.22}$$

$$\langle xy \rangle = \iint xy f(x,y) dx dy \tag{1.23}$$

The screen-camera system resolution is limited by the point-spread function (PSF) of the screen, the resolution of the lens system, the pixel size on the camera, among other contributions to the overall PSF.

Other diagnostics include the transverse deflecting cavity (TCAV), which gives a time-dependent transverse kick based on the time of arrival of the beam, so one can measure the time of arrival of the beam by observing the beam centroid position and transverse size on the screen. This is accomplished by introducing a sinusoidally alternating electric field orthogonal to the beam travel direction. By selecting the zero-crossing phase, such that the central particle receives no kick, and by ensuring that the beam is short enough relative to the RF wavelength that the kick received is approximately linear, one can see a linear

spread of the beam splayed transversely on a screen. Limiting factors on resolution include the initial beam size, the RF frequency, along with the screen resolution, and higher order effects [40, 41, 42]. Similarly, one can measure the imprint of the beam energy spread using a dipole magnet and a screen. A streak camera [43] uses a similar principle to streak the image after photon emission from the screen through the use of an additional photocathode and electron optics.

Other nondestructive diagnostics include beam position monitors (BPM) and integrating current transformers (ICT). These measure the current induced from the passing electrons and deduce the transverse position and charge of the bunch, respectively. Another nondestructive diagnostic is the so-called virtual cathode camera (VCC), which allows the user to look at a facsimile of the laser that is shone onto the photocathode in a photoinjector. This is accomplished using a beam splitter and a transport line that is exactly the same length as the transport line to the photocathode.

As can be seen, many of the diagnostics above are destructive diagnostics, meaning that the beam can either be measured, or used for an experiment, but not both. Although nondestructive diagnostics such as the ICT or BPM are often used, they require a specific device that may not be implemented on a given accelerator. Thus, a unique opportunity for an accelerator model exists: the virtual diagnostic. Virtual diagnostics use a model to infer beam properties — that would otherwise need to be measured — based available information, such as the accelerator settings and nondestructive measurements of other beam properties. Virtual diagnostics have become commonplace in accelerator physics (e.g. [44, 45, 46, 47, 48, 49]). These are particularly useful when predicting a time-consuming measurement, or measurements are not available due to other requirements for the beam (i.e. needed for another experiment). However, it is obvious that the model predictions of the virtual diagnostic must be of high fidelity to the real beam properties, or the diagnostic is not useful. Further, the execution time must be short, or the latency of the predictions make the tool less useful in an online setting, where numerous control policies may rely on knowing the current state of the beam.

This is a more general point about models for control. In the simplest case, a beam with certain properties is requested by a user and an operator must make control inputs to comply. With numerous variables, this quickly becomes an optimization problem for a model. Of course, determining the optimal accelerator inputs requires a model of high fidelity to the data, but the execution speed is also important. While a slower model can be helpful to guide beam optimization, a fast model that can be optimized in a short time can significantly decrease the time required for manual or model-independent beam optimization and — in practice — increase the reliability of the procedure, as there is no human intervention.

Clearly, for diagnostics and control, there are numerous benefits to having a model that is of high fidelity to the data and that can be executed quickly. In this thesis, such a model is referred to as a high-performance model.

1.3 Modern Modeling Tools & Machine Learning

While simple matrix-based calculations for particle accelerators, as described in Section 1.1.2, have their place in low charge, simple, linear environments, these models, as expected, tend to diverge from reality in the presence of nonlinear effects. For example, in a space charge dominated regime, the standard transfer matrices will not predict the proper beam evolution, as space charge effects are neglected. Similarly, higher order nonlinearities, such as aberrations in magnetic electron lenses are neglected. Higher order transport tensors can be defined [50], or a linearized space charge matrix can be defined [51], but calculation of higher order effects can quickly become a computational problem.

Over the past decades, as predicted by Moore's law, computational power has doubled nearly every two years. Through sheer computational power and the software improvements that it has spurred, the fields of accelerator physics and engineering now have access to numerous physics-based simulations and mathematical models. Modeling tools for particle accelerators are more accurate, more general and faster to execute than ever before. These tools often use of the following techniques (or combinations thereof):

- Advanced matrix-based codes, which use an approach similar to that described above, but with higher order calculations included
- Particle tracking codes, which calculate the trajectories of the various particles in the electric and magnetic fields in an accelerator at each step
- Particle-in-cell codes, which can model the interaction of particles with electromagnetic fields, which are self-consistently calculated by solving Maxwell's equations for any geometry and medium.

Example codes include ELEGANT [52], OPAL [53], MAD-X [54], ASTRA [55] and General Particle Tracer [56].

In general, mathematical models that make simplifications, such as linear transfer matrix models, are fast to execute, given modern computing power. The downside, as alluded to above, is that with this quick execution time, the fidelity of the model to real beamline behavior may degrade in the presence of nonlinear effects. Computational models, on the other hand, while quite accurate, are generally slow enough that they are incompatible with online operation, in particular when optimization is involved. Even offline, optimization problems can quickly become a task for a supercomputer, particularly as the number of optimization parameters becomes large.

Meanwhile, the same increases in computational power that have spurred higher fidelity computational models have spurred an increase in the capability, availability, and utility of machine learning models. Regardless of the simulation or mathematical modeling approach listed above, the core of the modelling problem remains the same: certain inputs to the accelerator are specified and a prediction of of the beam properties given those settings is given. Accelerators have a clear set of input and output data, along with plentiful data sets. As such, a machine learning approach can be quite useful in bridging the above gap, and learning the relationship between inputs and outputs in simulation. Then the ML model can be a fast so-called "surrogate model" where the ML model can replace the simulation model, or real beamline in the optimization problem or online situation. The ML model can

run at the millisecond scale on a desktop computer, similar to a transfer matrix model, but includes the nonlinear effects of simulation. The downside of using such a model is the range in which it can operate, as discussed below.

1.3.1 Models & Fluctuations

The benefits of models for controls and diagnostics are incontrovertible, but modeling accelerators and beams can be a difficult problem. While they have many inputs and outputs and large quantities of data, short- and long-term parameter fluctuations, particularly when unmeasured, can make modeling difficult.

In each of the parts of an RF photoinjector (shown in the previous section and in Fig. 1.2) and downstream beam manipulations, there are measured and unmeasured fluctuations in accelerator parameter space, which affect the particle trajectories and fields to which the beam is exposed. For example, laser pointing jitter can cause a time-dependent jitter of position of the beam on the photocathode, which can lead to numerous deleterious effects for the production and transport of the beam. The beam could be produced off-axis in the gun, leading to a sampling of RF fields that introduce coupling between longitudinal and transverse planes and increase the transverse projected emittance, which can require further correction downstream [57]. In addition, the laser jitter can cause the laser to sample different parts of the photocathode, which if it has a nonuniform quantum efficiency (QE) map, can cause a charge difference with the jitter [58]. If instead, the operator decides to image an iris on the photocathode using an optical lens or lens system, the pointing jitter becomes an intensity jitter, as on each shot, a different amount of the beam will be aligned with the aperture. Intrinsic fluctuations in the laser can also cause similar intensity changes in the beam produced. To make matters worse, not all beamlines have passive virtual cathode camera data acquisition, so these fluctuations can go unnoticed. In other cases, at high repetition rates, the fluctuations can be averaged, leading to an increase in apparent beam size due to jitter alone.

Other examples of fluctuations in a photoinjector are the intrinsic fluctuations in the RF,

which vary from system to system, and can be measured non-destructively using an antenna, but not all systems have a measurement that is independent of the feedback system. Magnet fluctuations also can play a role in beam stability, but are not generally measured at this level or precision synchronously with the beam. Further, external factors, such as ambient temperatures, time-varying magnetic fields from other experiments, or even the presence of industrial activities in the proximity of the beamline can cause accelerator parameter drift and/or beam movement that is unrelated to accelerator parameter drift.

In general, large fluctuations in laser and RF parameters are mitigated by PID-type feedback, but some fluctuations remain. From a modeling perspective, the main impediment to building models is a change to the beam that cannot be explained by the model inputs. The solution can be as simple as incorporating more inputs into the model, but the real problem is unmeasured parameters, which can be catastrophic to a modeling effort.

Another problem that exists because of (measured) jitter and drift in parameter space is that of so-called "distribution shift" in machine learning. As stated in the previous section, ML models can strike a balance between model fidelity and execution time, but a major downside is that ML models do not extrapolate well outside the training set. Numerous efforts have been made to make models more robust to distribution shift, including a "soft landing" outside of the parameter space, so the models do not immediately deviate grossly from the previous predictions as soon as the system drifts outside of the span of the training set [59, 60]. However, the fact remains that online modeling can be impeded by a lack of training data of all parts of the parameter space. Jitter and drift can take the accelerator to previously-unexplored parts of the parameter space, causing problems for the models, including making physically-impossible predictions.

Modeling in the presence of jitter and drift is a theme in this thesis. In addition, working under the paradigm of "measurement is easier than control," model-informed techniques to use jitter and drift to minimize acquisition times and increase measurement accuracy will also feature prominently.

1.4 Dissertation Overview

In this thesis, various methods to improve diagnostics and controls by improving beamline modeling will be explored. In Chapter 2, virtual diagnostic models to predict beam time of arrival and energy will be presented in the context of ultrafast electron diffraction (UED). In Chapter 3, the building of high-fidelity beamline models is discussed, including using tools such as Markov chain Monte Carlo (MCMC) to infer unmeasured or difficult to measure input parameters for the model. Finally, in Chapter 4, accelerator modeling in the presence of distribution shift and jitter is discussed, in particular focusing on the optimization of the round-to-flat-beam transform (FBT) at the Pegasus laboratory at UCLA, and on the use of extremum seeking (ES) to compensate for drift and jitter. Further studies of ES at HiRES will also be briefly described.

CHAPTER 2

Virtual-Diagnostic-Based Time Stamping for Ultrafast Electron Diffraction

2.1 Motivation and Background

2.1.1 Virtual Diagnostics & UED Temporal Resolution

As outlined in Chapter 1, virtual diagnostics have been shown to be particularly useful for accelerators (e.g. [44, 45, 46, 47, 48, 49]). These tools use fast, high-fidelity accelerator models — models that accurately predict the beam behavior from accelerator parameters — to make predictions where destructive measurements would otherwise be required. This leaves the beam to be used for other experiments, measurements and/or applications. There are numerous situations where this can be advantageous, but it is particularly useful when previous measurements of the beam properties do not provide the required level of confidence in the beam properties of future shots. Two examples of situations when this might occur are:

1. When the desired beam properties change. For example, accelerators often run for multiple experiments in a given time period. Different users might request different beams. After the upstream accelerator properties are changed, the beam properties will change, indicating low confidence in the previous measurements of beam properties. In addition, day-to-day changes in experimental conditions can make previous days' measurements from a similar set point unreliable. For example, hysteresis and/or changes in laboratory thermal or humidity conditions can lead to equipment perfor-

mance differences, which change the beam properties given similar accelerator settings. If upstream measurements of the true magnetic field and/or temperature and humidity exist, then a virtual diagnostic could help.

2. In the presence of large fluctuations on the short- and long-term timescales (hereafter referred to as jitter and drift, respectively), no single shot is the same as any previous shot, leading to a low level of confidence in the future beam properties given the previous measurements.

In the first case, the beam might be able to be measured immediately after the change in accelerator settings, but this is not always the case. For example, in a setup with multiple beamlines, steering from one to the other could change the beam properties and also steer it away from certain diagnostics. In addition, in systems laden with fluctuations — the second case above — virtual diagnostics become particularly useful for two reasons:

1. Traditional techniques of measuring a beam parameter, then assuming that it stays constant for a short period and using a future shot to perform an experiment can be unreliable. The slightly more advanced, but similar approach of characterizing the noise by measuring multiple shots, then factoring in an uncertainty due to this measurement can lead to large uncertainty, and assume a stationary distribution (i.e. that there is jitter and no drift).

2. The "measurement is easier than control" paradigm: if the beam application requires intentionally varying a parameter that is known to fluctuate, a virtual diagnostic can make use of these natural fluctuations, instead of the fighting them. Simply by measuring — or inferring from a virtual diagnostic — the fluctuating parameter, one can avoid or minimize varying the parameter intentionally. This is a general approach that has advantages such as reducing acquisition time, and/or improving measurement precision in applications. There are numerous examples in accelerator physics where a scan of a fluctuating parameter is required, including beam energy scans for x-ray

free electron laser spectrum tuning or, as featured below, time of arrival scans for pump-probe ultrafast electron diffraction (UED).

2.1.2 Ultrafast Electron Diffraction

UED is a staple of ultrafast techniques and has broad application to understanding various phenomena on a fundamental timescale in chemistry, biology and materials science [61]. The technique is as follows: an electron beam (probe) is directed toward a sample, and shortly before it arrives, a laser pulse (pump) excites the sample to a non-equilibrium state. The resulting diffraction pattern is recorded and thus a snapshot is measured and tagged with the time delay between the pump and probe arrival times. UED is typically a multi-shot technique in which the pump-probe time delay is varied and diffraction patterns are recorded at each delay setting. The measurement is complete when a series of diffraction patterns at different time delays are stitched together to show the temporal dynamics.

For UED to be able to have high temporal and spatial resolution there are strong constraints on the electron beam:

1. Due to spatial resolution considerations, a low transverse emittance beam is required.
2. A short bunch is required for high temporal resolution.
3. Due to experimental considerations, such as image signal and acquisition time, a high-beam charge and/or a high repetition rate beam is also helpful [62].

Note that these conditions are equivalent to stating that a high 5-D beam brightness is required (see discussion in Section 1.1.1), for each individual beam and averaged over many beams at high repetition rate. Each of these constraints will be explored in greater detail below.

The spatial resolution of a UED experiment is determined by the lateral coherence length:

$$L_c = \frac{\lambda_c}{2\pi\epsilon_{2D}}\sigma \tag{2.1}$$

where $\lambda_c = \frac{h}{mc} \approx 2.42pm$ is the Compton wavelength σ is the transverse beam size in one dimension. Thus, UED requires a low-emittance beam. Particularly with short electron bunches, one of the main impediments to a low emittance beam delivered to the sample is the emittance dilution due to space charge.

At higher beam energies, space charge forces are suppressed. As such, holding everything else constant, a higher energy beam has less emittance dilution due to non-linear space charge (as discussed in Chapter 1). One of the main improvements in UED in recent years is the use of higher energy beams [61]. The use RF photoinjectors (discussed in Chapter 1), have brought MeV-scale energy beams to UED [63, 14], which have a higher energy than those produced by DC guns, which brought the energy range for UED to the tens of keV [64]. Even these relatively low-energy DC guns had a high energy relative to the seminal works in UED from the 1980s [65, 66, 67].

However, a low transverse emittance beam is only part of the story. Having a high beam brightness is critical. Thus, the ability to go to a higher charge is also important, in its own right. In a diffraction experiment, only the scattered electrons are measured, while the "zeroth-order" beam that is not scattered by the sample is discarded. UED is a transmission technique. As such, most samples must be thin, as the mean free path for a metal, for example, is on the order of 100 nm. For a thin sample, this can lead to low-signal environment, which requires a large number of incident electrons. A complimentary approach for the low signal problem is to integrate for many shots to boost SNR. If hardware setups allow for it, a high repetition rate is beneficial in this setup. However, this makes the same assumptions during the integration time that are described in the previous section.

The work shown herein makes use of HiRES [62, 1], which starts with the APEX gun [68], a high repetition rate source of semi-relativistic (nominally 750 keV) beams at up to 250 kHz. This allows for high brightness beams at a high charge, but also allows for many shots to integrated at high repetition rates.

The longitudinal properties of the beam are also critical for UED. In a UED experiment,

temporal resolution is defined as:

$$\tau = \sqrt{\Delta t_{e^-}^2 + \Delta t_{laser}^2 + \Delta t_{fluct}^2 + \Delta t_{VM}^2}, \tag{2.2}$$

where Δt_{e^-} is the electron bunch length, and Δt_{laser} is the laser pulse length. These quantities can be reduced using a bunching cavity and laser compressor, respectively. As can be seen, the development of chirped pulse amplification (and therefore ultrafast laser systems) [69] played a critical role in providing the fast laser pulses and also the short electron pulses needed for UED, as most modern UED sources use a probe beam made from photoelectrons. Δt_{VM} is the velocity mismatch term, which can be neglected for ultra-relativistic beams and thin samples. That leaves the limiting factor of Δt_{fluct}, the uncertainty in time-of-arrival due to the fluctuations between the laser pulse and electron bunch.

In general, how well the time of arrival (TOA) uncertainty between the laser and electron beam has been minimized is the main consideration for the fluctuation term. This work seeks to move beyond the hardware control limitations to improve the TOA uncertainty, and therefore the overall temporal resolution, which is particularly important in systems where long-term drift in the TOA is dominant. In such a system, traditional PID-type feedback can be insensitive to slow changes. Characterization measurements are impractical and the lack of data can lead to a better estimate of the temporal resolution of the system than actually exists. The time stamping approach in this work is shown to compensate well for long term drifts.

In order to measure the time of arrival and tag each shot in a UED scan, nondestructive time stamping techniques have been proposed or implemented. For example, the use of electro-optical sampling [70] and THz-based streaking [71] can be used for UED time stamping. However, these techniques strongly constrain the machine setup (charge, crystal proximity, THz deflector, depending on the setup), which might not be fully compatible with high-quality diffraction patterns. The approach shown in this work is more general and can be used in any setup in which upstream accelerator parameters can be tracked.

2.1.3 Chapter Outline

This chapter details the development of a virtual diagnostic at the HiRES beamline (LBNL) to predict the electron beam time of arrival and energy in a drifting environment, with application to time-tagging shots in a stroboscopic UED measurement, and is based heavily on [72]. The virtual diagnostic serves to reduce the uncertainty, thus improving temporal resolution, along with decreasing acquisition time, as no online corrections will be required.

In the next section, a correspondence between the energy and time of arrival virtual diagnostics will derived to highlight the connection between time of arrival stability and energy stability (and therefore timing and energy measurements). Thereafter, a description of the HiRES beam line, along with the synchronization and measurement schema will be given. Following that, a description of the prediction techniques will be given. Finally, results in time stamping, energy stamping and future directions are discussed.

2.2 Energy & Time of Arrival

Particularly at sub-MeV energies, one of the largest contributions to the time-of-arrival of the beam at the sample is the beam energy. In a drift, higher energy particles arrive sooner. With more complex arrangements of beamline elements, such as the drift-buncher-drift lattice described in this section, the relationship can become more complex. Using the linear transport theory described in Eq. 1.18, the connection is represented by the matrix element R_{56}. Note that in this section, the transport matrix is defined in terms of time of arrival (relative to the reference particle) and energy deviation (from the reference particle), as shown below.

$$\begin{pmatrix} x_f \\ x'_f \\ y_f \\ y'_f \\ t_f \\ \frac{\Delta E}{E}_f \end{pmatrix} = R \cdot \begin{pmatrix} x_i \\ x'_i \\ y_i \\ y'_i \\ t_i \\ \frac{\Delta E}{E}_i \end{pmatrix} \tag{2.3}$$

where $\frac{\Delta E}{E}$ is the relative energy deviation relative to the reference particle.

2.2.1 Longitudinal Dispersion in a Drift

In t-E space, the R_{56} term for a drift can be defined as follows:

$$\Delta t = t - t_0 = \frac{L}{c\beta} - \frac{L}{c\beta_0} = \frac{L}{c}\left(\frac{1}{\beta} - \frac{1}{\beta_0}\right) \tag{2.4}$$

where β_0 is the normalized velocity of the reference particle and

$$\gamma^2 = \frac{1}{1-\beta^2}. \tag{2.5}$$

Thus, Eq. 2.4 becomes

$$t = \frac{L}{c}\left(\frac{\gamma}{\sqrt{\gamma^2-1}} - \frac{\gamma_0}{\sqrt{\gamma_0^2-1}}\right) \tag{2.6}$$

Taylor expand to first order around the central beam energy, γ_0

$$\Delta t = \frac{-L}{c}\frac{\gamma-\gamma_0}{(\gamma_0^2-1)^{\frac{3}{2}}} \tag{2.7}$$

Defining $\Delta\gamma = \gamma - \gamma_0$ and noting that Eq. 2.5 can be rearranged as

$$\gamma\beta = \sqrt{\gamma^2-1}, \tag{2.8}$$

then Eq. 2.23 becomes

$$\Delta t = \frac{-L\gamma}{c\gamma_0^3\beta_0^3}\frac{\Delta\gamma}{\gamma} \approx \frac{-L}{c\gamma_0^2\beta_0^3}\frac{\Delta\gamma}{\gamma} \tag{2.9}$$

Since $\frac{\Delta E}{E} = \frac{\Delta\gamma}{\gamma}$, the R_{56} of the drift is:

$$R_{56} = \frac{-L}{c\gamma_0^2\beta_0^3} \tag{2.10}$$

2.2.2 Longitudinal Dispersion with a Radiofrequency Buncher Cavity

From Eq. 2.2, it is clear that a short electron bunch at the sample is critical. One of the main methods for bunching beams of energy of a few MeV or lower is the so-called ballistic bunching scheme. In this scheme, an RF buncher cavity chirps the beam such that the back of the bunch receives an accelerating kick, while the front of a bunch is decelerated. In a drift, (in this energy range) this allows the back of the bunch to "catch up" with the center; the front of the bunch is slowed to arrive at the same time as the center. To first order, all the electrons to arrive at the same time (of course, other higher order effects make for a short, but nonzero pulse length). The ballistic bunching scheme has been employed for UED numerous times (e.g. [73, 74, 62]) and has been shown to produce beams shorter than 10 fs [75]. At higher energies, a chicane-based bunching scheme can be employed (see Section 3.3).

To a linear approximation, one can use transfer matrices to model this system. The longitudinal transfer matrix for the buncher in time-energy coordinates is:

$$R_b = \begin{pmatrix} 1 & 0 \\ h & \frac{E_0}{E_1} \end{pmatrix} \tag{2.11}$$

where h can be calculated analytically as follows:

$$h = \frac{\alpha\omega_b}{E_1} cos(\phi_b) \tag{2.12}$$

E_0 and E_1 are the central energy of the beam before and after the buncher cavity, respectively. If the buncher cavity is run at the zero-crossing phase, then $E_0/E_1 = 1$.

This assumes that the buncher is a thin lens, focusing longitudinally. This can be seen to be a fair approximation, as the phase slippage of a travelling wave cavity is given as [76]:

$$\frac{d\phi}{dt} = \frac{2\pi c}{\lambda}(1-\beta) \tag{2.13}$$

given the time needed to traverse the approximately 20cm two-cell 1.3 GHz buncher cavity at HiRES at $\beta \approx 0.91$, this comes to only approximately 29 degrees, and thus the effect of the phase change along the buncher length can be neglected.

In the case of an accelerator lattice such as that in the time-stamping section below, composed gun, drift, buncher and drift, the following longitudinal transfer matrix applies for the system (gun to screen):

$$R_{gs} = \begin{pmatrix} (1 + h \cdot R_{56,\mathrm{bs}}) & (R_{56,\mathrm{gb}} + R_{56,\mathrm{bs}} \cdot (h \cdot R_{56,\mathrm{gb}} + \frac{E_0}{E_1}) \\ h & (h \cdot R_{56,\mathrm{gb}} + 1) \end{pmatrix} \tag{2.14}$$

where $R_{56,\mathrm{gb}}$ and $R_{56,\mathrm{bs}}$ refer to the R_{56} terms for the first (gun to buncher) and second (buncher to screen) drifts (see Eq. 2.10), respectively.

Neglecting the second order effect from the changing buncher strength h near the zero crossing, the overall R_{56} of this beamline is:

$$R_{56,\mathrm{gs}} = R_{56,\mathrm{gb}} + hR_{56,\mathrm{bs}}R_{56,\mathrm{gb}} + R_{56,\mathrm{bs}} \tag{2.15}$$

2.2.3 Comparison with Published Works

This is shown to agree with a similar analysis for the design of HiRES [1], where the following equation is given:

$$\begin{aligned} \Delta T_{\mathrm{eTOF}} = & \frac{\partial T_{\mathrm{eTOF}}}{\partial t_{\mathrm{cLaser}}}\Delta t_{\mathrm{cLaser}} + \frac{\partial T_{\mathrm{eTOF}}}{\partial \phi_g}\Delta\phi_g \\ & + \frac{\partial T_{\mathrm{eTOF}}}{\partial \phi_b}\Delta\phi_b + \frac{\partial T_{\mathrm{eTOF}}}{\partial V_{\mathrm{g}}}\Delta V_{\mathrm{g}} + \frac{\partial T_{\mathrm{eTOF}}}{\partial V_{\mathrm{b}}}\Delta V_{\mathrm{b}} \end{aligned} \tag{2.16}$$

where ϕ_{g} and ϕ_{b} are the phases of the gun and buncher, respectively, V_{g} and V_{b} are the gun and buncher voltages and Δt_{cLaser} is the delay of the cathode laser. The fourth term, the term that relates the initial energy to the final TOA, is the R_{56}. Using equations from the appendix of [1],

$$\begin{aligned}\frac{\partial T_{\mathrm{eTOF}}}{\partial V_{\mathrm{g}}}\Delta V_{\mathrm{g}} &= R_{56,\mathrm{gs}}\cdot\frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} \\ &= \frac{\gamma_{\mathrm{g}}-1}{\gamma_{\mathrm{g}}}R_{56,\mathrm{gb}}\left(C_{\mathrm{bs}}+\frac{R_{56,\mathrm{bs}}\gamma_{\mathrm{g}}\beta_{\mathrm{g}}}{R_{56,\mathrm{gb}}\gamma_{\mathrm{s}}\beta_{\mathrm{s}}}\right)\Delta V_{\mathrm{g}}\end{aligned} \tag{2.17}$$

where $\frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}}$ is the relative energy deviation at the exit of the gun. Note that the R_{56} definition in [1] has been modified to meet the definition of using the t-E space from Eq. 2.10. As the buncher is at the so-called zero-crossing phase,

$$R_{56,\mathrm{gs}}\cdot\frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} = \frac{\gamma_{\mathrm{g}}-1}{\gamma_{\mathrm{g}}}R_{56,\mathrm{gb}}(C_{\mathrm{bs}}+\frac{R_{56,\mathrm{bs}}}{R_{56,\mathrm{gb}}})\Delta V_{\mathrm{g}} \tag{2.18}$$

where C_{bs} is the compression factor:

$$C_{\mathrm{bs}} = \frac{\sigma_f}{\sigma_i} = 1 + h\cdot R_{56,\mathrm{bs}} \tag{2.19}$$

Thus, Eq. 2.18 becomes

$$R_{56,\mathrm{gs}}\cdot\frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} = \frac{\gamma-1}{\gamma}R_{56,\mathrm{gb}}(h\cdot R_{56,\mathrm{bs}}+1+\frac{R_{56,\mathrm{bs}}}{R_{56,\mathrm{gb}}})\Delta V_{\mathrm{g}} \tag{2.20}$$

Note that

$$\frac{\gamma-1}{\gamma}\Delta V_{\mathrm{g}} = \frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} \tag{2.21}$$

Thus,

$$R_{56,\mathrm{gs}}\cdot\frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} = (R_{56,\mathrm{gb}}+hR_{56,\mathrm{bs}}R_{56,\mathrm{gb}}+R_{56,\mathrm{bs}})\frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} \tag{2.22}$$

which matches with Eq. 2.15.

2.2.4 Derivation of Time-Energy Comparison at the Screen

Ultimately, the goal is not to derive an R_{56} term for the whole beamline, but to derive a term to correlate measurements of time of arrival and energy deviation. This can be accomplished as as shown below.

Eq. 2.14 leads to two equations:

$$\frac{\Delta E_{\mathrm{s}}}{E_{\mathrm{s}}} = h \cdot \Delta t_{\mathrm{g}} + (h \cdot R_{56,\mathrm{gb}} + 1) \cdot \frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} \tag{2.23}$$

$$\Delta t_{\mathrm{s}} = (1 + h \cdot R_{56,\mathrm{bs}}) \cdot \Delta t_{\mathrm{g}} + (R_{56,\mathrm{gb}} + R_{56,\mathrm{bs}} \cdot (h \cdot R_{56,\mathrm{gb}} + 1)) \cdot \frac{\Delta E_{\mathrm{g}}}{E_{\mathrm{g}}} \tag{2.24}$$

where Δt_s and $\frac{\Delta E_{\mathrm{s}}}{E_{\mathrm{s}}}$ are the time of arrival relative to the reference particle and relative energy deviation at the screen. Since it is assumed that all beams start at the same time from the gun, Δt_{g} is defined as zero, this means that the first term in Eq. 2.23 and Eq. 2.24 go to zero.

As such, by substitution, one can find δ_{s}, the energy deviation at the screen as:

$$\frac{\Delta E_{\mathrm{s}}}{E_{\mathrm{s}}} = \Delta t_{\mathrm{s}} \cdot \frac{h \cdot R_{56,\mathrm{gb}} + 1}{R_{56,\mathrm{gb}} + R_{56,\mathrm{bs}} \cdot (h \cdot R_{56,\mathrm{gb}} + 1)} \tag{2.25}$$

where Δt_{s} is the time of arrival at the screen. Thus, rewriting this in terms of the time of arrival at the screen, the total conversion factor between energy and time for the drift-buncher-drift-TCAV case is:

$$\Delta t_{\mathrm{s}} = \frac{\Delta E_{\mathrm{s}}}{E_{\mathrm{s}}} \cdot \left(\frac{R_{56,\mathrm{gb}}}{h \cdot R_{56,\mathrm{gb}} + 1} + R_{56,\mathrm{bs}} \right) \tag{2.26}$$

which connects the final time-of-arrival with the relative energy deviation $\Delta E/E$ from the reference particle. Thus, the first-order contribution of beam energy is often the dominant contribution to the particle time-of-arrival at a given plane in the beamline and a virtual diagnostic that looks at the kinetic energy of the beam could be useful to refine the predictions of the relative time-of-arrival fluctuations in a UED setup [77].

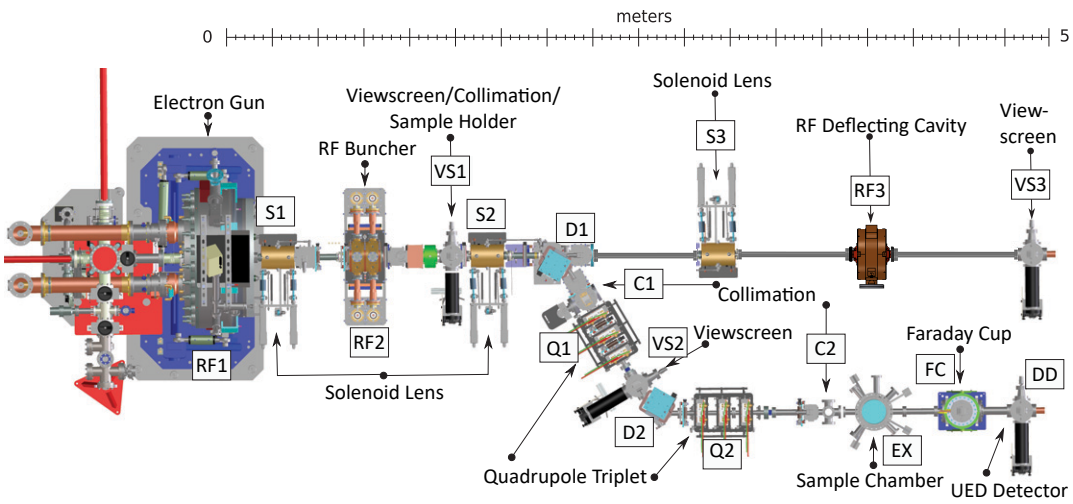


Figure 2.1: The HiRES beamline. The UED beamline starts at D1 and goes through the dogleg to DD, while the diagnostic beamline goes straight from D1 to VS3. Adapted from [1].

2.3 Synchronization & Measurement

2.3.1 The HiRES Beamline

The HiRES accelerator begins with the APEX gun, which is a continuous-wave-class, normal-conducting electron photogun with a nominal resonant frequency of 185.7 MHz [68] (RF1 in Fig. 2.1). The beam is then focused transversely through an emittance compensation solenoid (S1) and focused longitudinally through a bunching cavity (RF2) operating at 1.3 GHz, which is the 7th harmonic of the gun. As a CW-class gun, the maximum electron beam reptition rate is fixed by the photocathode laser to 250 kHz. An acousto-optic deflector at the end of the laser amplification chain can select user-defined patterns and/or lower the repetition rate. The nominal beam energy is 750 keV and all measurements in the chapter were taken with an approximate beam charge of 15 fC.

Referring to Fig. 2.1, a dipole magnet (D1) downstream of a second solenoid (S2) is used to select between two beamlines, each providing access to a series of diagnostic tools. The two beamlines are hereafter referred to as:

1. The diagnostic beamline: This line is the straight line, accessible with D1 turned off, and contains a third focsing solenoid (S3), a transverse deflecting cavity (RF3), which provides accurate pulse length and time of arrival information, with calibration of 23.37 fs/pixel. The deflected beam is measured on the downstream imaging screen (VS3).

2. The UED beamline: This is the beamline that leads on which UED experiments are performed. It branches off at D1 at an angle of 60° with respect to the straight line, resulting in high dispersion at the imaging screen VS2, and enabling high resolution energy measurements. The nominal dispersion is approximately 1 m, but can be increased or decreased using the quadrupole triplet just upstream VS2 (Q1). As such, the nominal $\Delta E/E$ at the screen is 2.5×10^{-5}/pixel. In the measurements presented in this section, the calibration for $\Delta E/E$ was 1.7×10^{-5}/pixel. In normal operations, Q1 is run in dispersion compensation mode, with the first and third quadrupole set to minimize x-P_z correlation at and after the exit of the second dipole D2. The second quadrupole in Q1 has the opportunity to focus in y and have a minimal effect on the x beam size, as the beam is near a focus in x in this setup, due to symmetry. Further downstream, the beam undergoes quadrupole focusing through Q2 and enters the sample chamber (EX), where samples can be placed in the path of the beam for UED. Downstream, there is DD, the UED detector, where diffraction patterns are imaged by an intensified Princeton Instruments PIMAX 4 camera. It is also worth mentioning that there is an insertable Faraday Cup (FC) that can measure the charge of the beam. C1, a v-shaped slit, allows for energy collimation, while C2 is a pinhole, designed to select part of the beam, and/or remove remaining dark current. A collimating pinhole at the position of VS1 removes dark current upstream.

The HiRES accelerator has a demonstrated history as a successful UED beamline (e.g. [78, 79]) and R&D beamline, developing new technologies for compact and large-scale user facilities (e.g.[80, 81, 82]). One of the notable features of the HiRES beamline is the CW APEX gun and the requisite RF subsystems. The low-level-RF control electronics (LLRF),

one of the most critical sub-systems for ensuring electron beam energy stability, has been developed at LBNL and then deployed at the LCLS-II accelerator at SLAC [83]. The system allows precision control and measurement of amplitude and phase, with minimal cross-talk (more the 100 dB isolation in the upgraded version) and white noise background below 150 dBc/Hz. Such development is a key component for developing high precision feedback loop controls, and for high fidelity prediction of beam parameters.

2.3.2 Synchronous Data Acquisition

A virtual diagnostic is a model-based diagnostic, and as such, building a model of the system is the most important part. This process involves correlating measurements of beam parameters with machine settings. In this work, beam parameters are extracted through destructive measurements on a scintillator screen. In the following sections, the beam energy is measured using a dipole spectrometer in the UED line as described above; the beam is intercepted on VS2 in Fig 2.1 after passing through D1. The time of arrival measurement requires intercepting the electron beam at VS3 after receiving a transverse, time-dependent kick from the transverse deflecting cavity (TCAV). In both cases, the figure of merit is derived from analyzing the resulting images for the beam centroid. The beam centroid is the first order moment of the beam distribution, calculated as follows:

$$\langle x \rangle = \iint x f(x,y) dx dy \tag{2.27}$$

using a similar argument to Eqs. 1.23, as the screen integrates over the other dimensions. Note that the choice of x-coordinate is intentional; both the dipole and the TCAV give a transverse kick in the x-direction. Also note that Eq. 2.27 includes an integration in y, also. While a line-out could be used, integrating over the y dimension helps improve the signal-to-noise ratio (SNR). Since the camera pixels are inherently discrete, Eq. 2.27 becomes a summation over the pixel intensities:

$$\langle x \rangle = \frac{\sum_y \sum_x x I(x,y)}{\sum_y \sum_x I(x,y)} \tag{2.28}$$

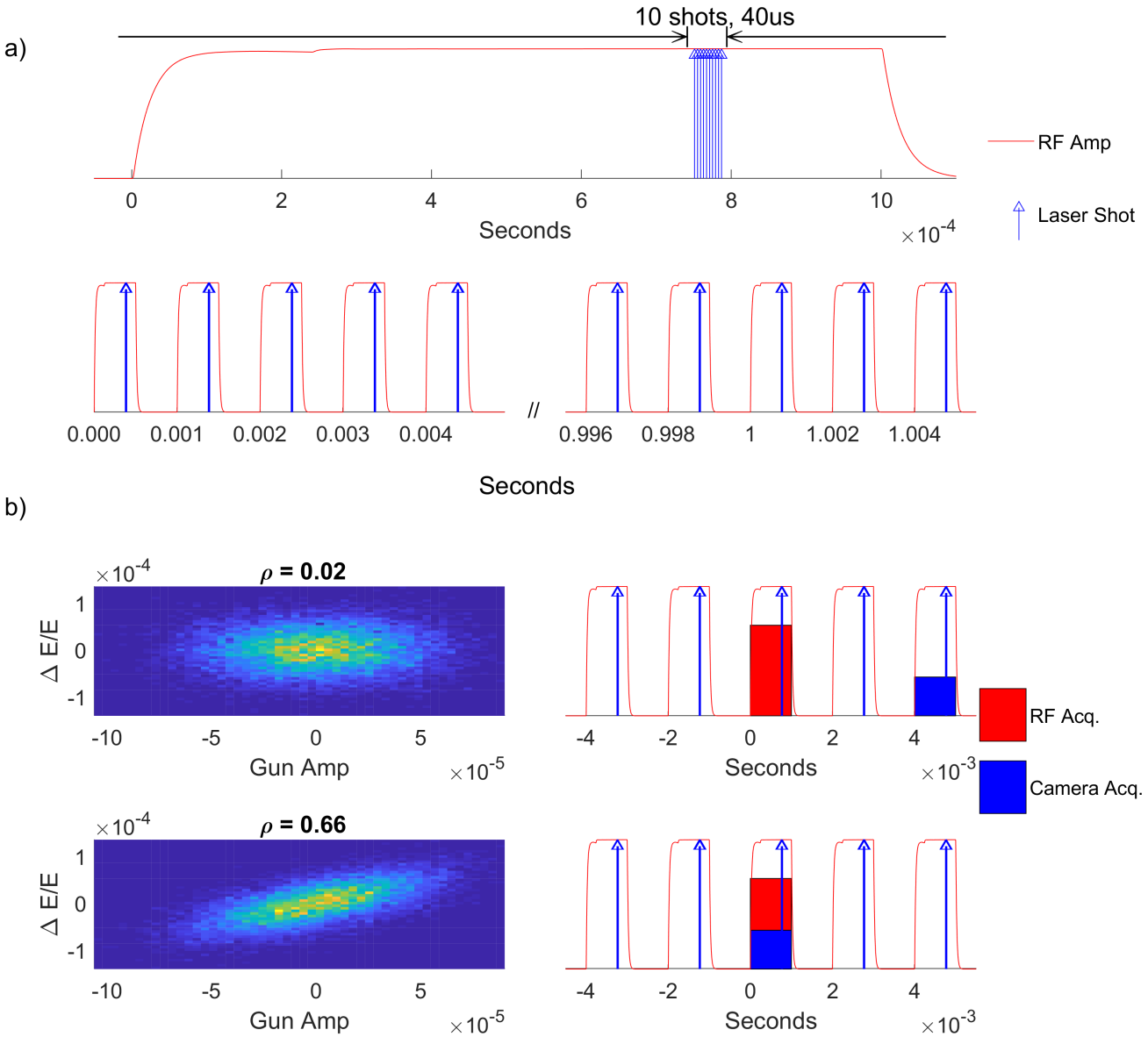


Figure 2.2: a) Synchronization scheme: RF amplifier run at 50% duty cycle with a 2 ms period. Up to 10 laser shots arrive at 4 μs intervals for a short period toward the end of each RF pulse, where the RF is generally most stable. b) Correlation plots of RF gun amplitude and relative energy deviation (measured at the dipole spectrometer) for synchronized and temporally misaligned acquisition schemes.

where $I(x, y)$ is the image pixel values. An alternative to Eq. 2.28 is to use Gaussian fitting, which provides the first and second order moments using an approximation that the beam

is Gaussian.

The SNR of the training datasets is of particular importance as the model will be trained on the processed variables extracted from the images, and any error in the calculation for the parameter corresponds to an effective loss of information. In order to boost image SNR, so that it is easier to see the beam above the background, it is possible to integrate multiple electron beam pulses (because of the high repetition rate of the system, each only 4 μs apart), so long as the timescale of system changes is longer than the averaging period.

At fixed RF power, the rate of change of the phase or amplitude of an electromagnetic field in a resonant cavity is limited by the cavity bandwidth, which acts as a filter for external disturbances, so that every noise component outside its bandwidth is strongly attenuated. In the case of the APEX gun, a 186 MHz CW-RF gun, with a quality factor Q greater than 10^4, we can estimate the time scale of field fluctuations:

$$\tau_{noise} = \frac{1}{\Delta f} = \frac{Q}{f} > 50\mu s \tag{2.29}$$

where f is the resonance frequency of the cavity. Further, there is an intra-pulse PID-type feedback system engaged, which should further reduce jitter. The engagement of the intra-pulse feedback can be seen after approximately 250 μs in the RF traces in Fig. 2.2.

For the RF bunching cavity and deflecting cavity, RMS fluctuations of the amplitude and phase of the RF in both cavities shows only a minimal increase when integrated for 40 μs relative to the case when integrated for 4 μs (the inherent uncertainty in laser shot time of arrival), as shown in Fig. 2.3. The slight increase in fluctuations in each cavity is expected to increase the uncertainty in beam time of arrival at the final screen by less than 10 fs each. Therefore, most of the data produced in this work has been collected by averaging 10 beams per image, therefore sampling the fields over an interval of 40 μs, in order to increase the SNR in the images (see Fig. 2.2).

In order to obtain the most accurate model and predictions, the heterogeneous data acquired (a mix of images and waveforms) requires deterministic time-alignment with a precision equal or better than τ_{noise}. Figure 2.2 describes the timing setup. The electron

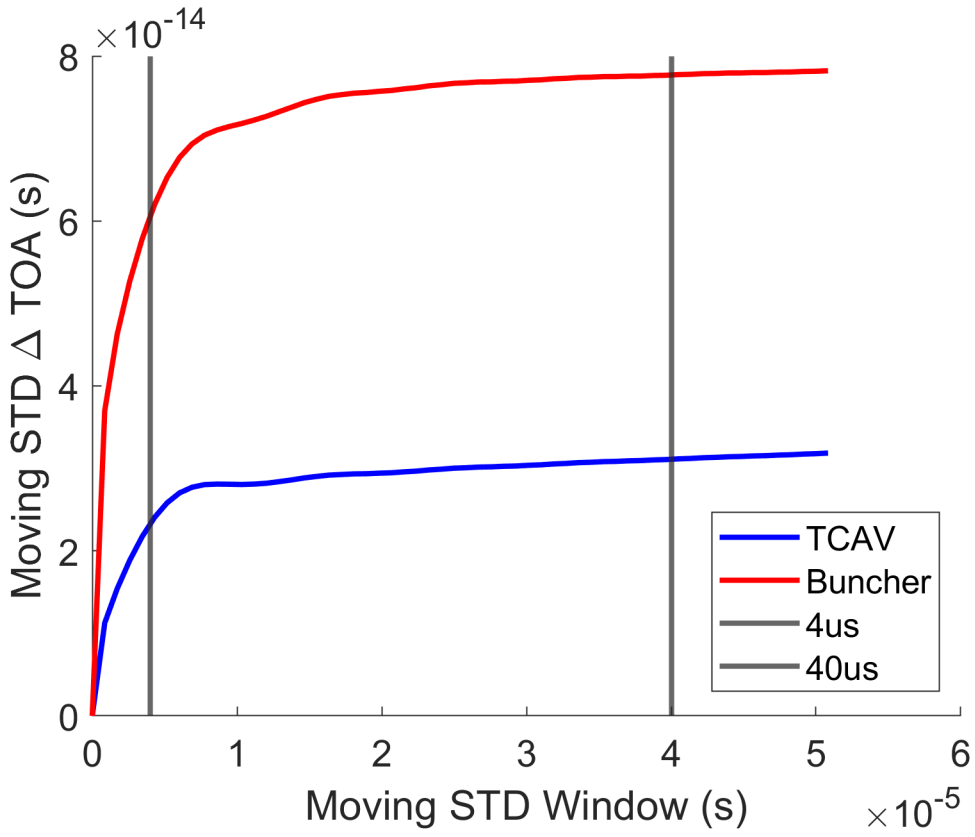


Figure 2.3: Mean moving standard deviation of the intrapulse fluctuations for the TCAV and buncher. These are extrapolated to TOA using the calibrations listed above.

gun is used in pulsed mode for these experiments, with a total duration of the RF pulse of 1 ms and a repetition rate of 500 Hz (corresponding to a duty cycle of 50%). The optical gate sending the burst of 10 consecutive laser pulses can be activated at any time along the RF pulse, with a precision of 4μs. In Fig. 2.2b, the correlation plots are shown of electron beam relative energy deviation and the amplitude of the field in the electron gun, in the simple case where no other cavity is used. The ρ value on top of each plot corresponds to the value Pearson correlation coefficient between the two. The data is shown in the case of synchronized and incorrectly-synchronized acquisition, providing a clear idea of the information lost without precise time alignment.

Since in general, the model learns the relationship between the predictors and a position beam centroid position on a screen, calibrated either to TOA or energy, in general,

readbacks from the measurement apparatus must also be included. In addition, the beam centroid position can change because of a magnet current change, or because of laser pointing fluctuations on the cathode. Therefore a complete model should calculate correlations with all the relevant parameters of the accelerator.

Since the gun is operated in pulsed mode, with approximate macropulse duration of 1 ms, and the cameras integrate over the 10 laser shots, the photocathode laser position and final images must be synchronized to below the ms scale. Up to the limitation on the FPGA-enabled RF chassis, RF parameters are taken synchronously, on the μs scale. Not all data requires the same level of time-alignment. Variations in parameters such as cavity temperatures, water flows and magnet currents mostly contribute to machine drifts, and only require synchrony at the sub-second level, which can be achieved via software. At HiRES, continuous data storing of user-requested machine setting is performed automatically by an online database with 10 Hz periodicity, providing the necessary information to include all machine parameters in the model.

Data for the deflecting cavity deserves a special discussion. As mentioned above, data on this cavity should in principle be taken synchronously, as small short-term fluctuations are expected. In Fig. 2.4 the correlation plots between the electron gun amplitude and the measured beam time-of-arrival before and after the compensation of TCAV short-term fluctuations are shown. Given the small contributions of these jitters to the measured short-term fluctuations and due to the limit in the number of channels that could be acquired at the same time, TCAV data are acquired via the database, and therefore account only for long-term drifts of the field in the cavity.

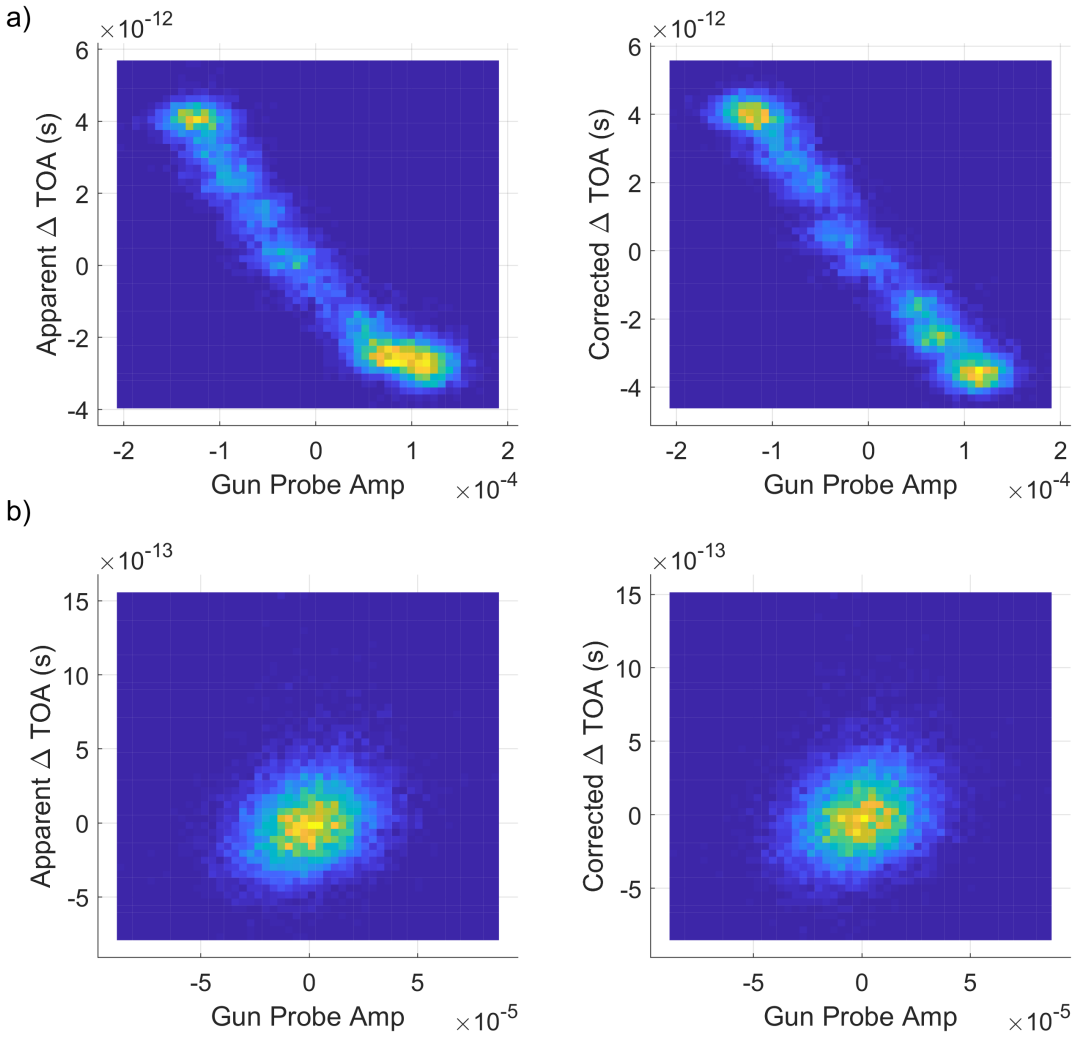


Figure 2.4: Synchronous measurements of beam TOA at TCAV: correlations of TOA with electron gun amplitude (without the use of the RF bunching cavity, similar to data in Fig. 2.2). Left: TCAV RF jitters are not taken into account in post processing. Right: fluctuations in TCAV RF amplitude and phase are used to correct beam time-of-arrival measurements. In a) long term drifts are uncompensated. In b) a moving average is subtracted to show only short time scale jitters. Note the much smaller y-scale in the bottom plots.

2.4 Data Analysis & Prediction

2.4.1 Linear Regression

In the following sections, virtual diagnostics based on multi-variable linear regression are presented. Multiple linear regression is a statistical modeling technique where

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.30}$$

where $\mathbf{Y}$ is a vector the observed quantities, $\mathbf{X}$ is a matrix of dimension number of observations by number of predictors, $\boldsymbol{\beta}$ is a vector of regression coefficients (dimension: number of predictors) and $\boldsymbol{\epsilon}$ is the error term, a vector of individual errors on each observation. Estimates of regression coefficients, $\hat{\boldsymbol{\beta}}$ are learned by minimizing residuals as follows:

$$\mathbf{Y} - \hat{\boldsymbol{Y}} = \mathbf{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{2.31}$$

Rather than producing a black-box model, which produces predictions through an opaque process, multiple linear regression produces an interpretable model. As it will be shown below, the regression coefficients are learned and can be analyzed in order to characterize quantitatively the impact that each predictor has on the overall prediction.

It is also important to observe that linear regression between two variables is agnostic to the time relation between different data points as it is inherently a time-independent method. However, the data sets in this work are all time-series datasets. While linear regression is effective in quantifying the consistent effect that the predictors have on the observation, the method will fail to identify the time-dependent noise processes that perturb the system and affect both the predictors and the observations.

In this case, a slightly adapted procedure was used: 1) the data were not randomized, in order to preserve the time-series ordering and 2) the model is trained on the first part of the data, while the last part of the data is reserved for validation. However, a linear regression model is exactly that: linear. Thus, any presence of small non-linearities in the system will not be captured. Also, the model may change with time, and therefore some sort of adaptive

tuning could be exploited (see [81, 59]). A discussion a forecasting method aimed at reducing each of these contributions follows in the next paragraphs.

2.4.2 Temporal Fusion Transformer

The regression approach relies on the assumption that TOA or energy can be extracted from linear correlations with the predictors, but could be further improved upon by taking into account temporal evolution using more complex models. In this work, an approach is shown that leverages advancements in forecasting to use a Temporal Fusion Tranformer (TFT) architecture to reduce temporal correlation in the residuals and further improve the prediction performance of the model.

The TFT, described in [84], is a state of the art transformer architecture for forecasting. Transformers are a member of the class of auto-regressive models that build upon models such as recurrent neural networks (RNNs) and long-term short-term memory networks (LSTM). Transformers use sub-modules consisting of stacked LSTM neurons as well as novel components such as self-attention in order to train on many datasets or large sequences. Transformers consist of an encoder that maps feature vector $\{x_{1,t}, x_{2,t}....x_{n,t}\}$ to a continuous representation $\{z_{1,t}, z_{2,t}....z_{n,t}\}$, which is then used by a decoder to generate a sequence of m predictions $\{y_{1,t}, y_{2,t}....y_{m,t}\}$. While the transformer architecture was originally used in the field of natural language processing, it has subsequently been expanded to a variety of other modalities including time-series forecasting and visual processing [85, 86].

In a TFT, at each time step t, a context window of length k consisting of past predictors, along with the past values of the ground truth, are sent to the encoder. Known machine parameters at the prediction time steps (or horizons) are also encoded. TFTs contain multiple variable selection networks that reduce network complexity through the modulation of the probability that a given predictor's signal propagates to deeper layers of the network. This reduces the need for data pre-processing, the impact of noisy variables and prevents overfitting. The TFT decoder includes a multi-headed attention module for long-term temporal pattern recognition. Through multi-headed attention, weights are learned that reflect the

degree to which encoded variables attend or correlate with one another [87]. These weights are then passed from the decoder to a dense layer that produces the quantile predictions.

In summary, the TFT offers an easily interpretable model that predicts a distribution of results instead of a point prediction. TFTs extract variable importance using both attention and automated variable selection [84]. Quantile predictions allow confidence of a prediction to be assessed, which is useful for human evaluation and downstream control tasks.

2.5 Time Stamping

2.5.1 Linear Regression

As described above in Eq. 2.2, the temporal resolution in UED experiments is often dominated by relative time of arrival fluctuations between excitation laser and probing electron beam. Therefore an online diagnostic capable of precise non-destructive measurement of TOA would have a profound impact on the overall instrument performance. For the data presented in this section, the accelerator setup matched the beam and machine parameters used during UED experiments, except that the beam was sent to the diagnostic line. As such, the RF bunching cavity (RF2) is set for temporal compression, with nominal field amplitude and zero degree injection phase (the so called *zero-crossing* phase). The fields in both the electron gun and the bunching cavity are stabilized in amplitude and in phase by fast, FPGA-based PID-type feedback loops.

In Fig. 2.5 the standard deviation of the time-of-arrival of the beam at the deflecting cavity (measured converting the centroid variation of the beam on the screen using the pixel to time deflector calibration) is shown as a function of the temporal duration for the data acquisition. This quantity continuously increases due to short and long term drift. In the inset, zooming in on the 1 minute time scale, it is shown how short-term drifts account for less than 200 fs of temporal jitter. On the other hand, with an increase in the temporal width of the acquisition window, the overall stability of the system is observed to degrade at longer timescales. Depending on the duration of the intervals in between re-establishing a

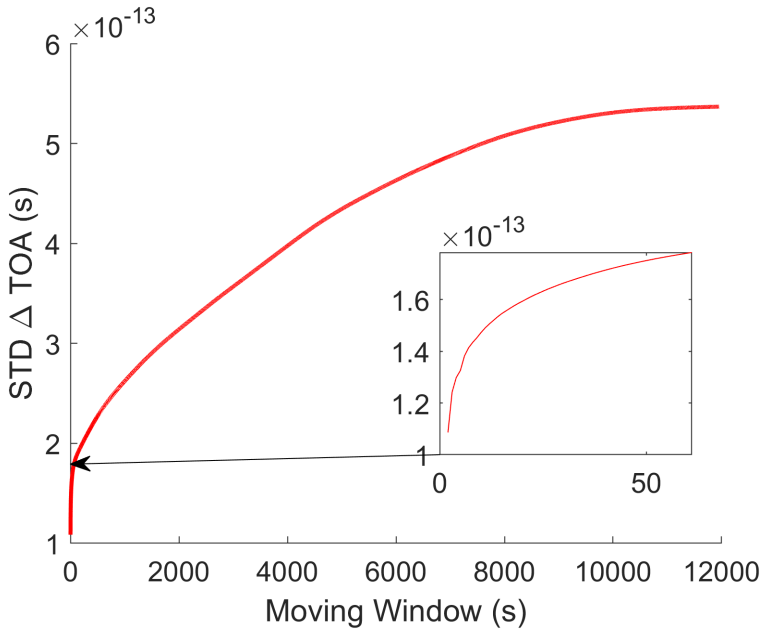


Figure 2.5: Standard deviation of transverse beam centroid calibrated to TOA relative to the reference beam as a function of the width of the acquisition time window. Inset shows a short timescale.

new time-zero position in UED pump probe scans [61], the integrated resolution can become as large as 600 fs.

Figure 2.6 shows the evolution of the beam TOA at the TCAV over about 3 hours. The data are divided in two sections (highlighted by the vertical dashed line), with 75% of the points used for developing a model of the system (the training data set), and the last 25% is used to validate it. A further test set is not required to test generalization, as no hyperparameter tuning was required.

The linear regression model described in Sec. 2.4.1 is then used for prediction. Inputs to the model include RF amplitude and phase of the electron gun, the bunching cavity and the deflecting cavity, the photocathode laser arrival time at the cathode with respect to the RF wave and an image of its transverse shape, intensity and position at the cathode.

Parameters for the model in this section are listed in Table 2.1. Parameters marked as "Async" were taken asynchronously, as described in Section 2.3.2. For the gun and

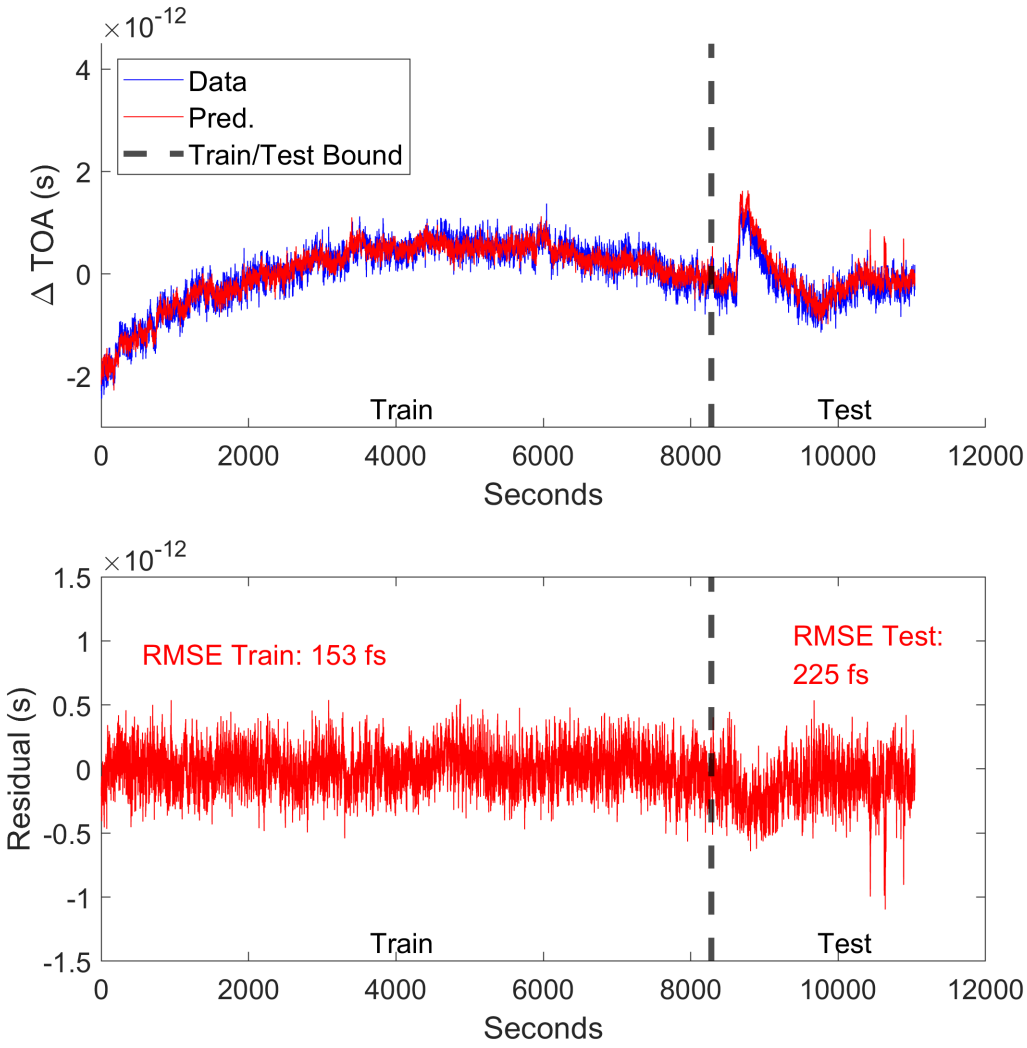


Figure 2.6: Time of arrival fluctuation measured using the TCAV screen while the PID-type feedback was engaged. Residual drifts not corrected by the feedback are present. The linear regression is also shown. The uncertainty due to long-term drifts is reduced to the 200 fs level similar to the shot-to-shot, short-term jitter shown in Fig. 2.5.

buncher, PID-type feedback loops were engaged, based on some of these readings. If PID-type feedback was engaged based on one of the parameters, it is designated as "IL" or "in-loop." If an independent measurement of the parameter exists, it is denoted as "OL" or "out-of-loop."

The red line in Fig. 2.6 shows the result of the regression. The model is able to learn the

Parameter Name	**Description**
Gun Probe Amp (IL)	In-loop gun RF probe amplitude
Gun Probe Phs (IL)	In-loop gun RF probe phase
Gun Probe Amp (OL)	Out-of-loop gun RF probe amplitude
Gun Probe Phs (OL)	Out-of-loop gun RF probe phase
Laser Phs	Phase difference between laser and RF
Laser Crosstalk	"Amplitude" of the above signal – measures channel crosstalk
Gun Rev Amp	Amplitude of reverse power
Gun Rev Phs	Phase of above
Gun Fwd Amp	Amplitude of forward power
Gun Fwd Phs	Phase of above
Buncher Probe Amp (OL)	Out-of-loop buncher RF probe amplitude
Buncher Probe Phs (OL)	Phase of above
Laser Position (x)	x-coordinate of the virtual cathode image centroid
Laser Position (y)	y-coordinate of the virtual cathode image centroid
Buncher Rev Amp 2 (Async)	Amplitude of reverse power at buncher coupler 2
Buncher Rev Phs 2 (Async)	Phase of the above
Buncher Fwd Amp (Async)	Amplitude of buncher forward power
Buncher Fwd Phs (Async)	Phase of above
Buncher Rev Amp (Async)	Amplitude of buncher reverse power
Buncher Rev Phs (Async)	Phase of above
Buncher Probe Amp (IL – Async)	Amplitude of in-loop buncher probe
Buncher Probe Phs (IL – Async)	Phase of above
TCAV Rev Amp (Async)	Amplitude of TCAV reverse power
TCAV Rev Phs (Async)	Phase of above
TCAV Fwd Amp (Async)	Amplitude of TCAV forward power
TCAV Fwd Phs (Async)	Phase of above
TCAV Probe Amp (Async)	Amplitude of TCAV probe
TCAV Probe Phs (Async)	Phase of above

Table 2.1: List of predictors for the time stamping models.

correlations in the data and predict to a high degree of accuracy, decreasing the uncertainty in the long-term data (RMSE) by more than a factor of two in the test set. This represents

a major improvement, one that reduces the uncertainty from the hours-long timescale in Fig. 2.5 to the minutes-long stability.

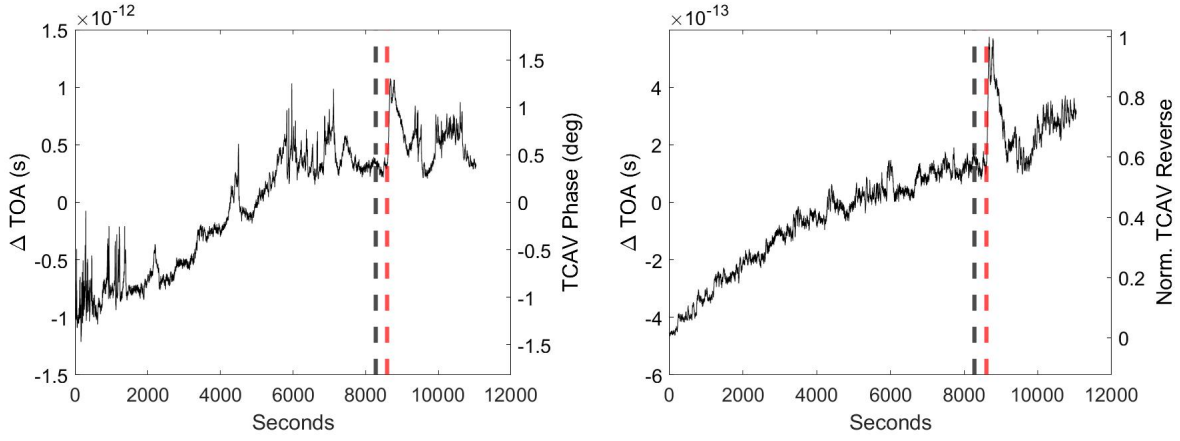


Figure 2.7: Left: TCAV phase calibrated to apparent change in TOA. Right: TCAV reverse power calibrated to apparent change in TOA. In both, the dashed gray line indicates the train/validation boundary, while the red line shows the TCAV parameter spike.

Notably, the model was able to predict accurately the outcome of a sudden large phase shift in the TCAV in the test dataset (visible at around 8600 seconds). Such a jump is mostly due to a sudden variation in the phase and amplitude of the TCAV, and not to an actual change in beam TOA. Using linear regression, this can be referenced to (apparent) TOA fluctuations. See Fig. 2.7 for plots of the TCAV phase and TCAV reverse power as a function of time (interpolated to synchronous timescale, as discussed in Sec. 2.3.2). The phase jump at approximately 8600 seconds (denoted by a red dashed line; the train/validation split is denoted with a black dashed line) has an amplitude of 0.9 degrees in approximately 60 seconds, and it is most severe phase jump seen in the data. This is likely due to a spike in the reverse power of the TCAV, which is also included in the model. By tracking parameters of the measurement system, i.e. TCAV RF amplitude and phase, real temporal shifts can be isolated from simple beam centroid fluctuations due to variations of the fields in the TCAV. Indeed, once the system correlations have been learned from the training dataset, all the coefficients that would contribute to a beam movement on the screen but not necessarily to

a change in TOA can be removed.

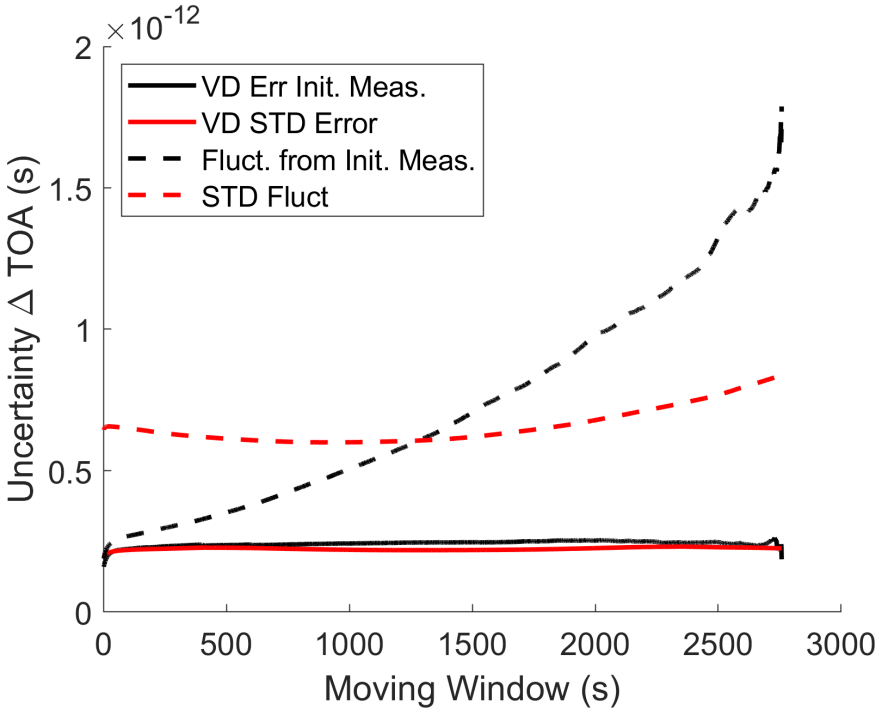


Figure 2.8: Validation set: Uncertainty with and without virtual diagnostic (VD) prediction as a function of time since previous measurement. The lines in black show the RMS fluctuations from a fixed reference at the beginning of the measurement, whereas the STD measures the RMS fluctuations relative to the mean.

Although the improvements made so far are impressive, they do not fully take into account the complexity of a drifting, time-series dataset. The discrepancy arises because of the distinction between the root mean square error (RMSE) of the linear regression prediction from the actual measurement, and the standard deviation (STD), which measures the deviation from the mean value. It is essential to carefully choose the reference point for assessing the virtual diagnostic's best performance. With traditional feedback, the typical approach is to minimize the system's drift until it becomes negligible, and the uncertainty is then determined by the STD. Essentially, when destructive measurements are turned off, the beam is presumed to have the same properties as the last-measured beam. Thus, in a drifting system, the uncertainty in the TOA would increase with time (due to the variation

in the mean value), as is shown in Fig. 2.5.1, where the mean STD fluctuations for a window of a certain size is shown, compared with the mean RMS fluctuations from the shot prior to the window. This comparison is shown for the system fluctuations and the virtual diagnostic error. As can be seen, by utilizing the linear regression virtual diagnostic, it is possible to compensate these drifts, and the uncertainty remains relatively constant over time, with little to no degradation.

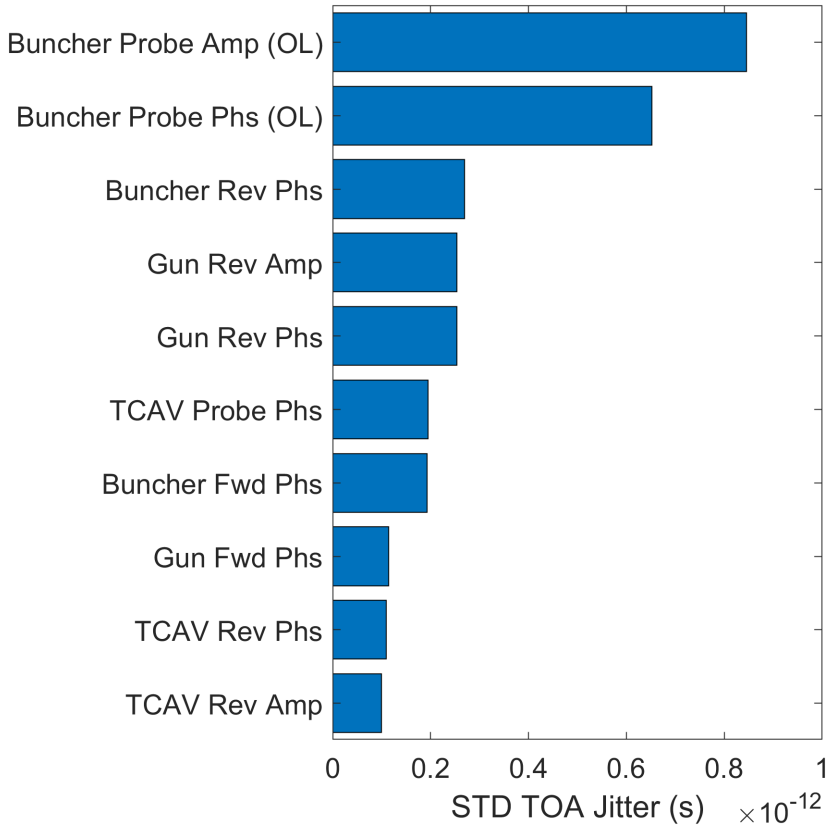


Figure 2.9: Top 10 model predictors and the associated TOA movement with 1 STD movement in the validation set. Details of the predictors can be found in Table 2.1 in Appendix 2.5.1.

As described in the Section 2.4.1, the impact of each predictor on the overall time of arrival estimate can be extracted from the model. In Fig. 2.9 the top predictors' impacts (one standard deviation of variation converted to TOA prediction using the corresponding β from Eq. 2.30) are reported. This can be helpful for several reasons, including 1) to

see if conventional feedback systems can be better tuned and 2) to see if the perceived TOA variation is due to measurement uncertainty (i.e. the TCAV measuring the TOA is jittering) or if the TOA is actually moving. Although the effect of the TCAV is significant, the dominating contribution is that of the buncher, meaning that the TOA is actually moving, despite conventional feedback systems. It also suggests that these conventional feedback systems could be improved for the buncher cavity.

These results show how the combination of a linear-regression-based model and time-aligned data can help enhance the performance of traditional feedback systems. However, as a time-independent method, predicting, any time-dependent correlations are neglected, as are all higher order correlations. In the following section, a technique to address both of these contributions is discussed.

2.5.2 Advanced Prediction

2.5.2.1 Method

In offline forecasting tasks, access to past ground truth data allows for the context window to be populated with observations of the target as shown in Eq. 2.32. A prediction at time step t is given by:

$$\hat{Y}_t = f(\{\mathbf{X}_{t-k}, ..., \mathbf{X}_t\}, \{Y_{t-k}, ..., Y_{t-1}\}) \tag{2.32}$$

where $\mathbf{X}$ at each time step is a vector of predictors, such as machine parameters. Y at any given time step is the ground truth data, in this case the beam TOA, where $\hat{Y}$ is the prediction of this quantity. k is the context window, or history, that the model is given.

In the context of the HiRES virtual diagnostic, the lack of ground beam TOA measurements during deployment about the beam TOA has to be negotiated once destructive measurements cease. In such an online or a multi-horizon task, one approach would be to introduce previous predictions recursively as shown in Eq. 2.33.

$$\hat{Y}_t = f(\{\mathbf{X}_{t-k}, ..., \mathbf{X}_t\}, \{\hat{Y}_{t-k}, ..., \hat{Y}_{t-1}\}) \tag{2.33}$$

In this approach, residual bias introduced in the model's estimates would accumulate with each recursive prediction and over thousands of timesteps would become significant. Even without a significant increase in error, error would time-correlated rather than being normally distributed. In order to avoid this problem, once destructive measurements cease, previous ground truth measurements must be replaced with time-independent predictions as shown in Eq. 2.34.

$$\hat{Y}_t = f(\{\mathbf{X}_{t-k}, ..., \mathbf{X}_t\}, \{g(\mathbf{X}_{t-k}), ..., g(\mathbf{X}_{t-1})\}) \tag{2.34}$$

where $g(x)$ is any time-independent model. In the work presented herein, the linear regression model as shown in Sec. 2.5.1, as described in Eq. 2.30 is used to replace ground truth data, after training with access to the ground-truth data, as shown in Eq. 2.32. During training, the TFT has access to a context window of long-term trend information in $\mathbf{X}$ and $\mathbf{Y}$ in order to learn from a more complete view of the system's dynamics. Following training, during online application, accurate estimates of $\mathbf{Y}$ as provided by the linear regression-based estimates $g(\mathbf{X}_t)$ are utilized and therefore the approach does not suffer from catastrophic degradation of the predictions after destructive measurements are no longer available.

The results in this section make use of a TFT implemented in Pytorch Lightning [88] with the PyTorch forecasting package [89]. A 75/25 training split identical to that of the linear regression model above was employed with one caveat: for both the training and validation sets, the first $\mathbf{k}$ instances (with k being the length of the context window) do not have corresponding predictions. The predictors were the same as those used for the linear regression model in Section 2.5.1, as shown in Table 2.1. Finally, a robust normalizer was applied to the data, which scales and centers it with regard to but without transforming the target.

The TFT architecture is outlined in detail in [84]. The hyperparameters for the TFT model are listed in the Table 2.2.

Hyperparameter	Value
Hidden Size	89
Dropout	0.276
Attention Head Size	2
Learning Rate	0.0012
Max Encoder Length	25

Table 2.2: List of hyperparameters for the model shown in Section 2.5.2.

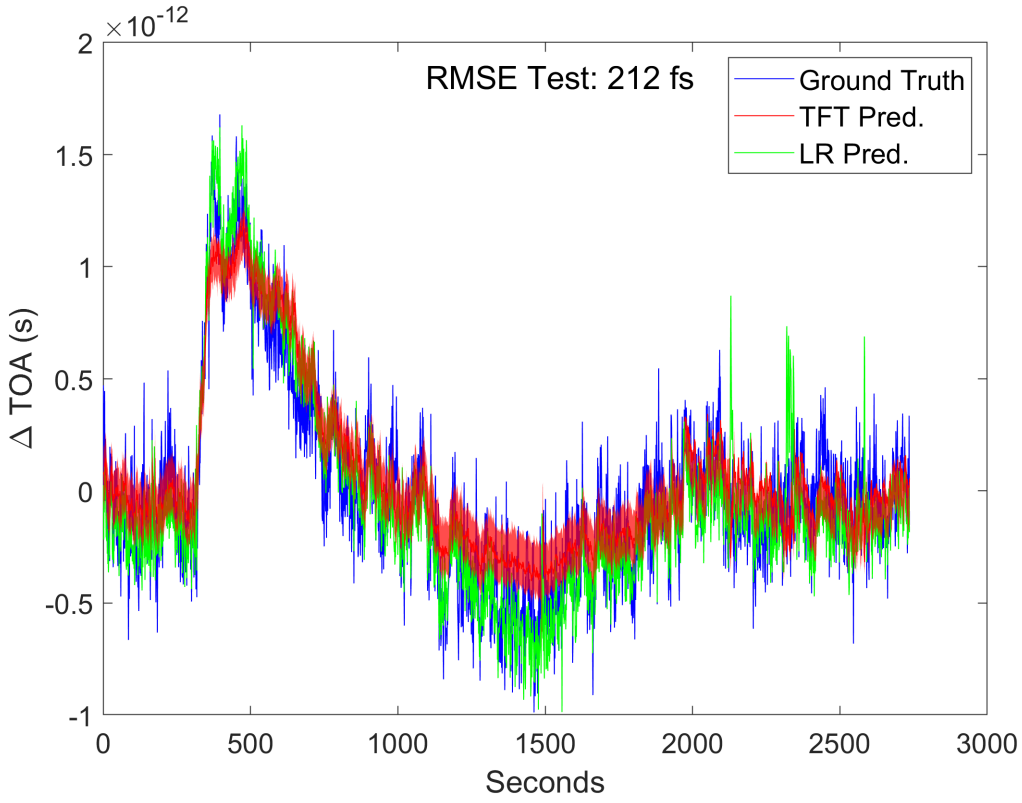


Figure 2.10: TFT TOA median and interquartile range (shaded) predictions in the validation set. Predictions are compared with linear regression (LR) predictions and the ground truth.

2.5.2.2 Results and Discussion

The results shown in Fig. 2.10 are the quantile predictions trained on ground-truth observations of the beam TOA. Note that the RMSE is approximately 6% better than that of linear

regression, as the residuals of the model are in general closer to zero, as shown in Fig. 2.11.

This improvement can be explained by noting that despite state-of-the-art stability at HiRES, Fig. 2.11 demonstrates that the processes causing long-term TOA drift introduce correlations of error over time that could be reduced via the use of a forecasting model. As shown in Fig 2.11, residuals of the TFT show little auto-correlation between time steps relative to the linear regression results. This demonstrates that the approach of incorporating historical accelerator parameter information despite the challenges of online forecasting detailed above is a reasonable one.

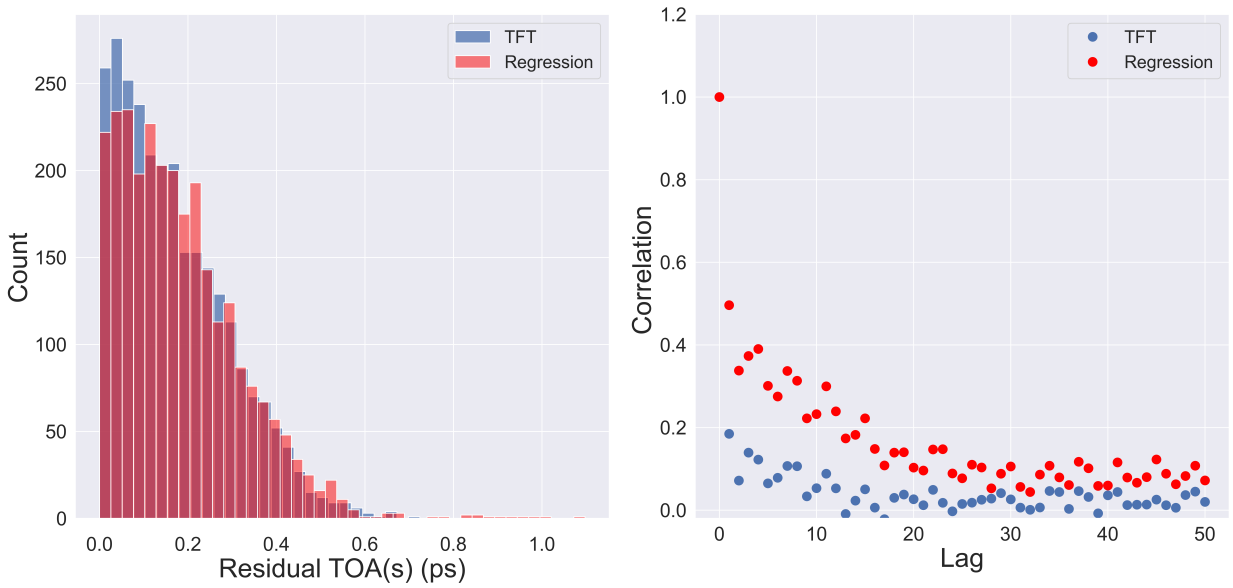


Figure 2.11: Left: Error histogram comparing the absolute value of the residuals from the linear regression and TFT models. Note that the error is bunched closer to zero for the TFT. Right: Residual correlation for linear regression and TFT models. Note that the use of a TFT model reduces the correlation in the residuals between time steps relative to a linear regression model.

It should be noted that because of the small number of observations, only a validation set was used in this work, and a test set was omitted. Although the validation set was used for hyperparameter optimization and as a metric for training, the generalization to a

validation set is still notable, under these conditions. In the future, pre-training on data from different accelerator runs prior to fine-tuning on related data should be investigated to further reduce residual error and allow for increased observation of anomalous states in the machine parameters and their correlated effect on the target beam parameters. Every state-of-the-art use of the Transformer model used in Natural Language Processing since 2015 has relied on pre-training to increase performance [90] and the use of transformers for forecasting would likely benefit from similar methods. Greater diversity of observations from better exploration of the parameter space of predictors would lead to better generalization and better model performance. The inability for the predictions to capture the full variation of the ground truth beam TOA exhibited in Fig. 2.11 could be explained by the fact that while relationships between machine parameters and beam TOA have been learned in training, anomalous transitions in machine parameters have not been observed before, forcing the model to extrapolate. While anomalous observations are, by definition, rare, a larger number of observations of similar transitions in other experiments would allow for greater predictive power during these periods. Training with more raw data and possibly with lagged data could help capture temporal relationships between changes in the variables.

2.6 Energy Stamping

A similar approach can be used to obtain very accurate predictions of the electron beam energy. To showcase this capability, the work in this section makes use of separate beamline settings. In particular, the electron beam is transported into the UED line and measured at the VS2 screen (Fig. 2.1) after acceleration by the electron gun, while the RF bunching cavity (RF2) is left off for simplicity of interpretation. Three hours of data were acquired for the two different cases of stabilized and unstabilized accelerating field in the gun (using the active LLRF PID stabilization loop mentioned earlier).

The simpler setup leads to fewer predictors, as is shown in Table 2.3. The quantity that the virtual diagnostic model predicts is the beam position after the dipole, calibrated to energy deviation from the reference.

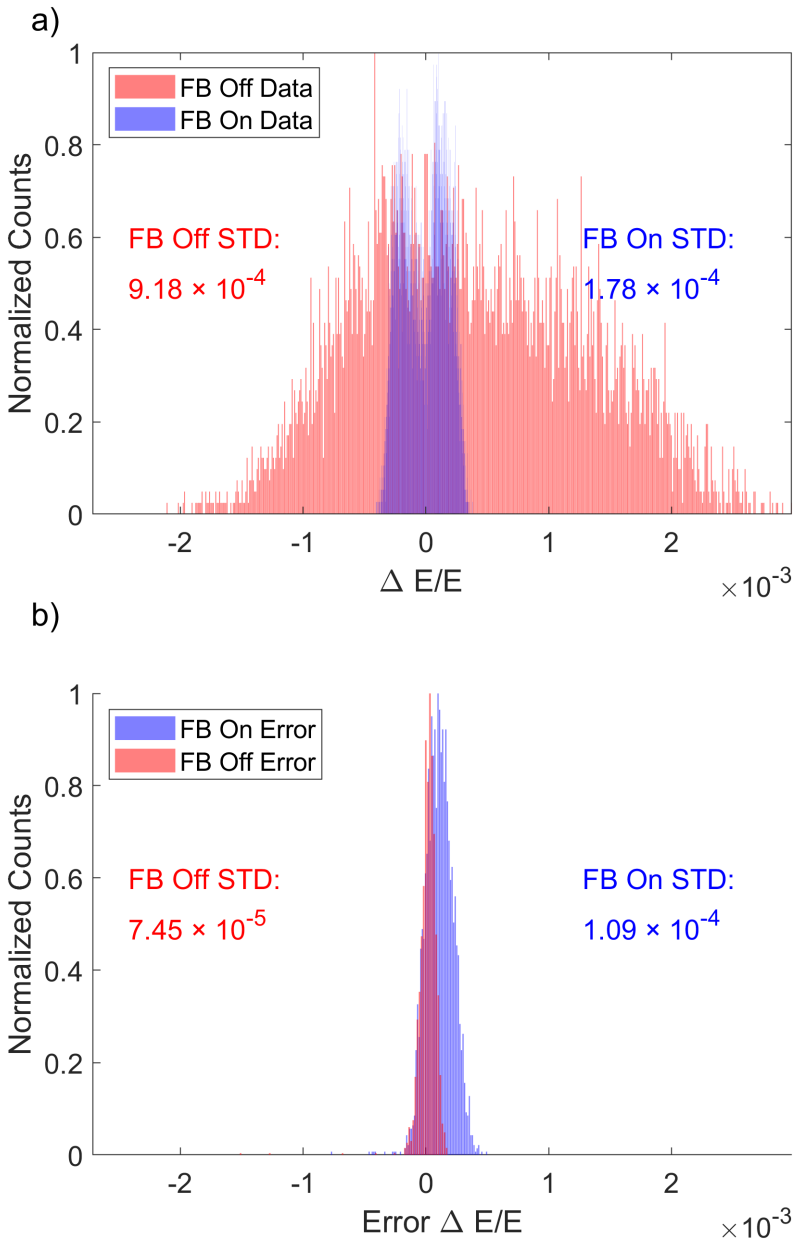


Figure 2.12: Linear regression predictions with and without traditional PID-type feedback (FB) engaged. Top: The variation from the mean of the training data is shown, after conversion to relative energy deviation. Bottom: the validation errors are shown. Note that the FB off case shows greater improvement than the FB on case on both an absolute and a relative scale.

Parameter Name	**Description**
Gun Probe Amp (IL)	In-loop gun RF probe amplitude
Gun Probe Phs (IL)	In-loop gun RF probe phase
Gun Probe Amp (OL)	Out-of-loop gun RF probe amplitude
Gun Probe Phs (OL)	Out-of-loop gun RF probe phase
Laser Phs	Phase difference between laser and RF
Gun Rev Amp	Amplitude of reverse power
Gun Rev Phs	Phase of above
Gun Fwd Amp	Amplitude of forward power
Gun Fwd Phs	Phase of above
Laser Position (x)	x-coordinate of the virtual cathode image centroid
Laser Position (y)	y-coordinate of the virtual cathode image centroid

Table 2.3: List of predictors for the energy stamping models.

Results of energy stability measurements in the two cases are shown in the histogram of Fig. 2.12a. The effect of the fast feedback in stabilizing the energy is evident, with RMS relative energy stability going from approximately 10^{-3} to 2×10^{-4} over a 3-hour run. Nevertheless, a clear structure is evident in the stabilized case, with a double peaked distribution of unknown cause, suggesting even better performance may be achieved.

The application of the linear regression model to both scenarios results in the histograms of Fig. 2.12b. Here, the residual error left after comparing the model predictions with the measured ground truth is shown. Importantly, the application of the model increases the precision with which the energy of each electron beam can be asserted, by a tenfold factor for feedback-off, and by a small factor in the feedback-on case.

The application of the model also results in a distribution of residual error that is much closer to a Gaussian, to be expected when only random noise is left, and nothing else can be learned from the system. This is therefore a first indication that our model is close to optimal.

The final RMSE is similar in the stabilized and unstabilized case. This point is quite interesting, as it shows that virtual diagnostic tools have comparable performances with respect to traditional feedback systems. For applications where time-of-arrival may vary, but in a well-controlled fashion, such as pump-probe experiments, this suggests that in the future, software-based approaches could outperform hardware-based approaches in the stabilization and control of particle beams. In the future, considering likely increase in computing power, network speed and bandwidth, feedback could even be based on virtual diagnostics.

However, it is worth noting the greater improvement in the feedback off case, both relatively and absolutely, which is related to the accuracy in the measurement of key parameters, and the level of noise in the measured ground truth used for training the model. In this work, final predictions with accuracy at the 10^{-5} level are desired, which requires more than 100 dB SNR in measurements of radiofrequency signals. Potentially more dangerous is the requirement on the currents energizing the different magnets in the beamline. Measuring 10^{-5} variations on this currents requires specialized hardware that it is usually not available for each magnet of the accelerator. Therefore, an offline experimental sensitivity study was performed to verify the dipole D1 as the one with the highest impact on the beam position on the screen. The current fluctuations driving the magnetic dipole were then measured with high precision by a specialized setup. A Danisense DS50ID ultra-stable flux-gate current transducer with a 16-bit digitizer was set up to measure the current provided to the dipole from the CAEN A3620 power supply, when set to a nominal value corresponding to a 750 keV electron beam. A 66-hour-long measurement of the current was taken, and showed fluctuations on the high 10^{-5} level, corresponding to apparent relative energy functions on the of 5×10^{-5} level. See Fig. 2.13 for more details. The RMS fluctuations found during this test, although not contextual with beam measurements, are of the same scale of the residual error obtained in both cases of Fig. 2.12, showing that direct synchronous measurement of the current in the dipole magnet could increase the precision of the virtual tool.

As such, when the PID-based feedback is engaged, the energy fluctuations in the training

set are small and on the same order of magnitude as the noise. This has two important consequences: 1) The training set does not explore much of the input parameter space, which makes it difficult for the model to learn the trend and 2) measurements of the change in energy of the beam are dominated by noise, making it even harder to learn the pattern in training.

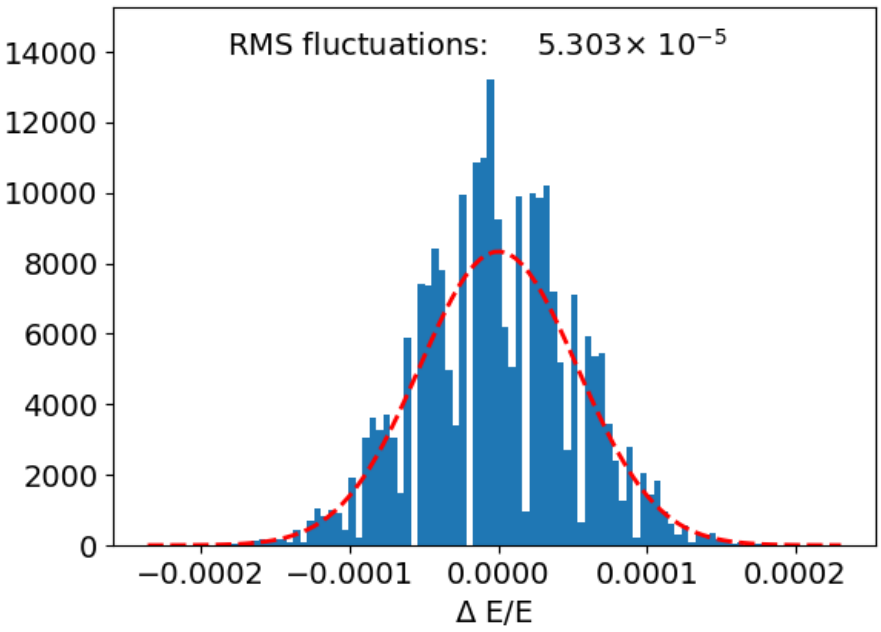


Figure 2.13: Measured current fluctuations on D1 (see Fig. 2.1) to perceived relative energy fluctuations (as would be measured on VS2 in the dogleg) for a 750 keV electron beam, as in experiment.

Using the matrix formalism for the linear transport in longitudinal phase space, developed in Section 2.2 one can compare the virtual diagnostics results for the energy and time of arrival measurements. At the TCAV screen,

$$\Delta t = \left(\frac{R_{56,\mathrm{gb}}}{h \cdot R_{56,\mathrm{gb}} + 1} + R_{56,\mathrm{bs}} \right) \frac{\Delta E}{E} \tag{2.35}$$

where $R_{56,\mathrm{gb}}$ and $R_{56,\mathrm{bs}}$ are related to the drift distances from the gun to the buncher and from the buncher to screen, respectively. These R_{56} elements are given in time-energy co-

ordinates for convenience. For HiRES, these can be calculated to be 0.73 ns and 3.20 ns, respectively. h is the R_{65} term associated with a thin lens description of the buncher cavity. For HiRES, calibration measurements indicate this to be $(-2.02\text{ns})^{-1}$. This can be expressed in engineering units to be approximately:

$$\Delta t \text{ [ns]} = 4.3415 \frac{\Delta E}{E} \tag{2.36}$$

where Δt is given in nanoseconds. Using an uncertainty from the energy stamping virtual diagnostic of about 7.45×10^{-5} (see the feedback off case of Fig. 2.12), the uncertainty is 227 fs, when subtracting a resolution term for the dipole. This result is comparable to the 225 fs uncertainty of the time stamping virtual diagnostic.

2.7 Outlook & Conclusion

In this chapter, a novel application of model-based virtual diagnostics has been explored, toward enhancing UED temporal resolution by predicting electron beam TOA or the main contributor to TOA in this energy regime, beam energy. It is also worth noting here that there are other cases where non-destructive measurements of the beam energy (which otherwise require bending the beam in a dipole spectrometer) would greatly improve accelerator performances. For example, in multi-shot measurements of transverse phase spaces, such as a quadrupole or solenoid scan emittance measurements [91, 92] energy fluctuations change the focusing strength of the magnets, which would be considered to be constant for such a scan; poor energy stability is catastrophic to such a measurement. Even single-shot emittance measurement techniques, such as [93] require knowing the beam energy.

Linear-regression-based models can be used to greatly reduce uncertainty in machine parameters. For energy stamping, linear-regression-based virtual diagnostics were shown to mitigate the long-term drift to a level comparable to what can be accomplished with the PID feedback loops. For time stamping, linear-regression-based virtual diagnostics were shown to work in concert with traditional feedback to mitigate long-term drift and lower the uncertainty to 225 fs, which is on same scale as the shot-to-shot fluctuations and a

significant reduction from the uncompensated standard deviation uncertainty of 600 fs. The linear regression results serve as a practical baseline result that – by itself – shows promise for improving stability, but state-of-the-art forecasting models were applied to mitigate the temporal correlation of residuals of the model predictions, resulting in a nominal reduction in prediction uncertainty to 212 fs in the time stamping case. The tool has been tested for single-shot beam predictions using information collected passively during beam runs and, while in these first tests the acquisition was limited to 1 Hz, minor modifications to the timing system would allow much faster repetition rates, in the kHz range and beyond.

In the future, this model-based virtual diagnostic approach can be applied to reduce the TOA uncertainty in UED measurements. This could be accomplished by informing a control policy (i.e. feedback) or by working under the "measurement is easier than control" paradigm that is a theme of this thesis. Rather than working to control further the natural parameter drift of the machine in an already state-of-the-art stability environment, the remaining drift and jitter can be harnessed to improve measurements. In UED experiments, if for each shot, the virtual diagnostics showcased in this work are applied to retrieve the relative time of arrival within the shot-to-shot error, one would be able to reorder the data using the shot-tag information with a corresponding improvement in the temporal resolution as well as significant reduction of acquisition times.

CHAPTER 3

Physics-Based Model Matching

3.1 Motivation & Background

As discussed in Chapter 1, high-performance models are invaluable to accelerator controls and diagnostics. In general, these models can be anything that predicts beam properties given accelerator settings. Two main categories of these models are the physics-based model and the data-driven model.

Data-driven models are described elsewhere in this thesis and are applied to control (Chapter 4) and diagnostics (Chapter 2). These are powerful tools, in particular if two conditions are met:

1. the model can be trained such that during experiments, the parameters do not drift outside the span of the training set.

2. the input parameters can be measured continuously (i.e. the synchronous measurements that were detailed in the previous chapter). An effort to forecast unmeasured parameter drift was explored in the previous chapter, but is not the ideal solution: tracking input parameter drift is.

Physics-based models are complimentary to data-driven models and are particularly beneficial to controls and diagnostics when the two conditions above are not both met. Meeting the first condition above can be challenging, as acquiring the requisite amount of training data can be an arduous task, despite methods to sample efficiently the space [94, 95].

A general description of physics-based models appears in Chapter 1. In this section, General Particle Tracer [56] (GPT) is used extensively. GPT is a particle tracking code, in which the user defines the input distribution and accelerator lattice. GPT then solves the equations of motion at each time step, as the beam evolves down the beam line. Both external fields (due to lattice elements) and self-fields (space charge) are incorporated. Care must be taken that each of these input parameters is properly specified. Not only does the full initial 6-D phase space need to be specified — at least to an equivalent beam approximation [96, 97, 98, 28] — each of the field maps, including those of magnets and RF cavities must match reality. In some cases, it can be best to approximate these using analytic functions, for faster and more accurate (due to interpolation errors) results. Additionally, while GPT has an automatic time-step determination algorithm, sometimes the user must select a smaller time-step for more accurate results.

In a perfect world, physics-based models would match perfectly the reality of the experiment (and the predictions would match those of a data-driven model), but generally do not, due to factors that fall outside the control of the experiments or uncertain parameters that serve as simulation inputs, such as those listed above. Jitter and drift, discussed throughout this work, also can be a contributor to the lack of agreement between measurements of input parameters and reality.

One of the many benefits of having a physics-based model that matches the true behavior is that one can use the measurements of downstream parameters to extract physical knowledge of about unmeasured parameters. This chapter will employ Markov chain Monte Carlo (MCMC) sampling of inputs to physics-based models in order to gain extra information about unmeasured photoinjector parameters in photocathode studies.

In photocathode studies, in order to ascertain the photocathode MTE, solenoid scans are performed. This measurement generally requires varying the solenoid and measuring the beam distribution at the exit of the screen. There are numerous parameters that affect the beam behavior, aside from the photocathode MTE, including the photocathode recession depth in an RF photoinjector, which can lead to additional focusing or defocusing (depending

on the sign of the recession depth) in cases where the plug is not seated flush with the gun back-plane. Other factors include higher-order magnetic field components in the solenoid, for example quadrupole moments, which affect the transverse focusing of the beam and introduce an asymmetry between measurement axes. As a multi-shot technique, solenoid scans can also be susceptible to fluctuations in the laser, RF and solenoid, where these are supposed to be held constant. By determining these parameters all together, the MTE can be found with more confidence.

This chapter will highlight a method for sampling the posterior distribution of this parameter space, such that error estimation and correlation analysis of these parameters can be performed. This is followed by discussion of a method to improve the accuracy of the predictions even when the second condition above — the acquisition of synchronous data — is satisfied.

Following a discussion of MCMC for photocathode studies, an additional example of physics-based model fitting in the absence of a data-driven model is shown at FAST, where limited data availability precludes an experimentally-trained data-driven model. This model will benefit numerous studies at FAST, including proposed photocathode studies. This chapter borrows heavily from manuscripts that will be submitted on MCMC and on the FAST model matching.

3.2 Injector Characterization Using Markov Chain Monte Carlo Methods

3.2.1 MCMC Background

Markov Chain Monte Carlo (MCMC) methods are extremely widespread in other branches of physics [99, 100, 101], but have been so far less commonly applied in accelerator and beam physics. In this approach, a reliable model can be used determine a posterior probability distribution function of parameters that are known to affect the measurements. Thus, the goal of the MCMC sampling is matching model parameters given data. One could use an

optimizer for this type of problem. Depending on the problem, a local optimizer could be selected, such as the derivative-free Nelder-Mead Simplex [102] or a gradient-descent optimizer. A description of the various types of optimizers is beyond the scope of this dissertation; textbooks such as [103, 104] list many optimizers and their advantages and disadvantages. Although similar in their aim for this problem, MCMC sampling is fundamentally different than an optimizer; an optimizer seeks to "fit" and predict a singular point in parameter space that best matches data. MCMC sampling can be used to predict the high probability region in parameter space that matches data.

The key is to estimate the probability of the model parameters given the data. This can be seen through Bayes' Theorem:

$$P(\theta_M|D) = \frac{P(D|\theta_M)P(\theta_M)}{P(D)}. \tag{3.1}$$

This is a general theorem for any θ_M or D. However, for the problem posed above of fitting model parameters to data, D represents the data and the model parameters are represented by θ_M. The probability of the model parameters given (conditional upon) the data (the LHS of Eq. 3.1) is the quantity that is sought. The denominator on the RHS of Eq. 3.1 is simply a normalization factor. Explicit calculation of this factor requires knowing and integrating over all possible states of D. In general, it is simpler to work with the following:

$$P(\theta_M|D) \propto P(D|\theta_M)P(\theta_M). \tag{3.2}$$

The first term on the RHS of Eq. 3.2 is is referred to as the likelihood function and is the conditional probability of the data given the model parameters. This is in general what one hopes to gain from simulations: information about the result given certain model parameters. The second term is referred to as the prior probability, or the probability of the model parameters. This is where the user can inject prior knowledge, such as limits for the parameters.

In MCMC sampling, the probability distribution is sampled as a function of various θ_M, given some D. An initial starting point is specified by the user. A move is proposed

in parameter space, and the move is accepted with probability related to improvement in the posterior probability computed using Eq. 3.2. As in the Monte Carlo method, this is completed numerous times with pseudorandom numbers determining the outcome of each step. The chain of moves is only related to the current and proposed positions in parameter space, and is therefore a Markov Chain.

A common algorithm used in determining the sequence of random samples is the Random Walk Metropolis-Hastings algorithm [100], where candidate moves are accepted with probability proportional to the ratio of new value of Eq. 3.2 relative to the old value. However, in all of the examples in this thesis, the "Stretch-Move" algorithm, as proposed by Goodman and Weare is used [101]. In this case, the probability of acceptance of the proposed moves is related to the distance of the move and proportional to the ratio of the new and old posterior probabilities given in Eq. 3.2.

If allowed to converge with sufficient number of moves, then the density of the samples in the steady state is approximately the posterior distribution. Knowledge of the a posterior distribution of the fitted parameters can be used to understand the reliability of the retrieved answer. MCMC allows for the computation of an interval of confidence (or even an actual probability density function) for each of the retrieved parameter, gives a high-probability region for each parameter, along with the covariance of various parameters.

3.2.2 Solenoid Scans

Solenoid scans are used to characterize various unseen accelerator and beam parameters that are impossible or impractical to measure directly, such as the mean transverse energy of the electrons emitted from the photocathode (e.g. [21, 105]), or indirectly, the photocathode response time [106]. The general idea is to build a reliable model of the accelerator and fit input parameters to that model based on the beam sizes on a screen as the solenoid is varied.

Accurate determination of the beam sizes is critical to the success of the solenoid scan. Due to signal-to-noise ratio considerations, in this work, the projection 1-D projection method is used to determine RMS beam sizes; two of the three transverse spatial second-

order moments, σ_{xx} and σ_{yy}, can be obtained relatively trivially by means of a projection. The third, σ_{xy} can also be calculated by means projections, if a 45-degree rotation of the image is used. It can be shown that the beam's spatial projection on the screen has the following form:

$$f(x,y) = \exp\left(\frac{-1}{2}\begin{bmatrix}X\\Y\end{bmatrix}^T \Sigma^{-1}\begin{bmatrix}X\\Y\end{bmatrix}\right) \tag{3.3}$$

where X, and Y are vectors of he mean-subtracted particle positions and Σ^{-1} is given by:

$$\Sigma^{-1} = \frac{1}{\det(\Sigma)}\begin{bmatrix}\langle yy\rangle & -\langle xy\rangle\\ -\langle xy\rangle & \langle xx\rangle\end{bmatrix} \tag{3.4}$$

Note that $\det(\Sigma)$ is an invariant of rotation. These coordinates, X and Y, can be related related to those in the $u-v$ plane of a rotation of 45 deg. Hence,

$$\begin{bmatrix}X\\Y\end{bmatrix} = \begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}\begin{bmatrix}u\\v\end{bmatrix} = R\begin{bmatrix}u\\v\end{bmatrix} \tag{3.5}$$

Inserting Eq. 3.5 into Eq. 3.3, the following relation can be found:

$$f(u,v) = \exp\left(\frac{-1}{2}\begin{bmatrix}u\\v\end{bmatrix}^T R^T\Sigma^{-1}R\begin{bmatrix}u\\v\end{bmatrix}\right) \tag{3.6}$$

By comparing Eq. 3.6 with Eq. 3.3, the following equations are obtained:

$$\begin{aligned} R^T\Sigma^{-1}R &= \frac{1}{\det(|\Sigma|)}\begin{bmatrix}\langle vv\rangle & -\langle uv\rangle\\ -\langle uv\rangle & \langle uu\rangle\end{bmatrix}\\ &= \frac{1}{2}\frac{1}{\det(|\Sigma|)}\begin{bmatrix}\langle xx\rangle+\langle yy\rangle-2\langle xy\rangle & \langle xx\rangle-\langle yy\rangle\\ \langle xx\rangle-\langle yy\rangle & \langle xx\rangle+\langle yy\rangle+2\langle xy\rangle\end{bmatrix} \end{aligned} \tag{3.7}$$

Hence,

$$\langle xy\rangle = \frac{\langle xx\rangle+\langle yy\rangle-2\langle vv\rangle}{2} = \frac{2\langle uu\rangle-\langle xx\rangle-\langle yy\rangle}{2} \tag{3.8}$$

The determination of the second moments in u and v can be likewise computed using the 1-D projection method after rotating the image by 45 degrees. Thus, it can be seen that all three second order moments of the beam distribution can be determined by four projections

(two rotated). Due to propagation of error considerations and the fact that the $\langle xy \rangle$ term goes to zero when the semi-major axes of the beam ellipse are aligned with the x and y axes, hereafter, solenoid scans will be fit to match the x, y, u and v second order moments, rather than $\langle xx \rangle$, $\langle yy \rangle$ and $\langle xy \rangle$. This is equivalent in an error- and noise-free environment; $\langle xy \rangle$ can be determined from the four projections.

3.2.3 Markov Chain Monte Carlo Methods for Injector Characterization at HiRES

Characterizing the beam is critical to optimizing the beamline for UED experiments, which is one of the primary functions of the HiRES beamline (e.g. [78, 79], and as described in the previous chapter). HiRES is shown in Fig. 2.1 and is described in the previous chapter. These data in this chapter were taken on VS1, and the first solenoid, S1 was varied to create a solenoids scan.

It is worth mentioning that the HiRES gun is equipped with a INFN-style load-lock photocathode plug insertion system, as shown in Fig. 3.1. As such, the photocathode plug is not necessarily flush with the back plane of the gun, which as discussed above, can result in focusing effects. Further, the solenoid has quadrupole moments of unknown strength and angle. In general, a quadrupole moment can be decomposed into its normal and skew moments. As such, the elements are considered separately for the remainder of the chapter. Along with the photocathode recession depth and the quadrupole moments, the peak field in the gun and the MTE were included as free parameters in the model to which this chapter seeks to give confidence. Five parameters were identified as unknown and not possible to measure in situ to a high degree of accuracy: 1) the photocathode MTE, 2) the beam energy (known to within a few percent), 3) the normal element of the quadrupole moment in the solenoid, 4) the skew element of the quadrupole moment in the gun and 5) the recession depth of the photocathode plug relative to the back plane of the gun. As described above, the four second order moments, x, y, u, and v, are measured using the projection method on the removable YAG screen. A model of the beamline was created and simulated using

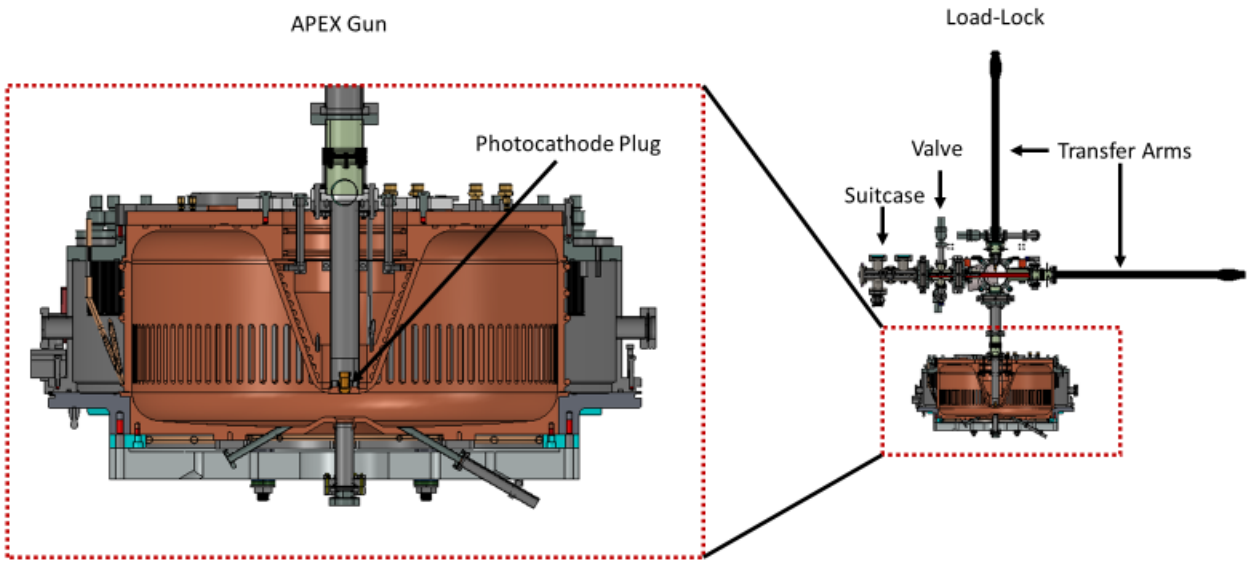


Figure 3.1: Load-lock system at HiRES, where the beam travel direction is oriented downward. On right: full schematic of load lock system at HiRES. Inset on left shows the photocathode plug in the gun. Often, the plug has some longitudinal offset relative to the nominal position, leading to a focusing or defocusing effect in the gun.

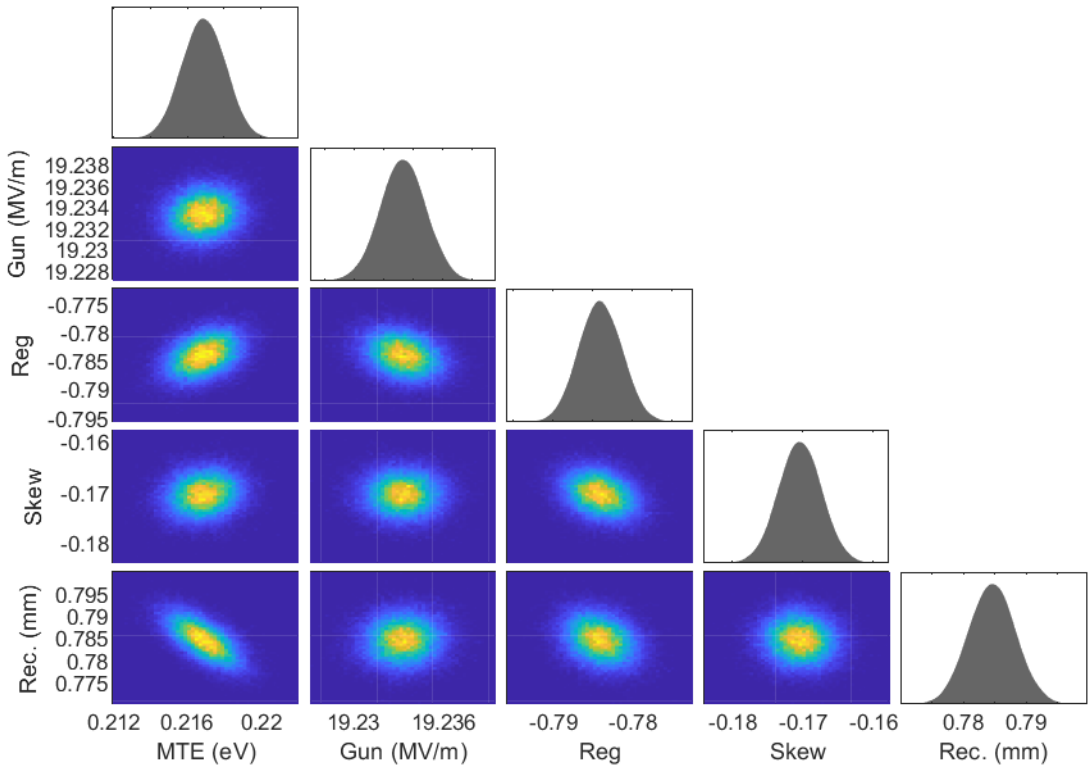


Figure 3.2: The various two-dimensional projections of the posterior distribution predicted by MCMC sampling. These 2-D projections are useful for seeing correlations, such as that of the MTE and recession depth. On the diagonal, the 1-D projections are shown in gray.

General Particle Tracer (GPT) particle tracking simulations [56].

Because of the relatively slow execution of the GPT model, a neural network surrogate model was developed. Approximately 20k iterations of GPT were run using a pseudorandom samples from a uniform distribution of of the five above parameters and used as training data for a feed forward, fully connected neural network with 6 hidden layers of 10 nodes each, with tanh activation function. The output is the x, y, u, and v moments. The model was trained to predict the beam sizes to within 5 μm, a small price to pay for the orders-of-magnitude speed increase. On a standard desktop computer, the model can execute in around 50 μs. A major improvement can be seen with larger numbers of simulation by vectorizing multiple simulations.

Of course, the surrogate model has limitations. As always, if an input parameter is outside

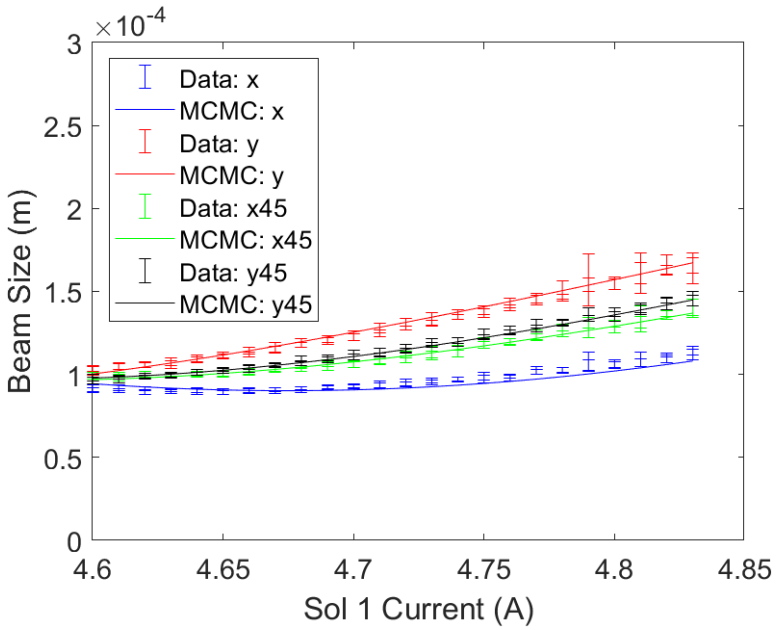


Figure 3.3: Plot of model predictions relative to measured data.

the span of the training set, the result from the NN will not be reliable. Assumptions on the input distribution of the beam were made. The model assumes that the beam is spatially Gaussian with no correlation between any dimension. In angles, the beam was assumed to have angles specified by the MTE of the beam (also Gaussian in x' and y', but a half-gaussian in longitudinal momentum). In this work, these assumptions are generally good, as the charge was seen to have a negligible effect on the transverse dynamics, meaning that the transverse dynamics are generally decoupled from the longitudinal dynamics. The VCC indicated a near-Gaussian beam and using angles given by the MTE is a standard assumption (e.g. [21])

The MCMC sampling was run on the surrogate model. The resulting projections of the posterior distribution are shown in Fig. 3.2. The mean values of the posterior distribution, with one standard deviation shaded error bars are shown plotted against the data in Fig. 3.3. Note that because of the relatively noise-free data, the error bars decrease as a $\frac{1}{\sqrt{N}}$ where N is the number of solenoid scan data points.

It is worth noting the correlation between the MTE and the photocathode recession

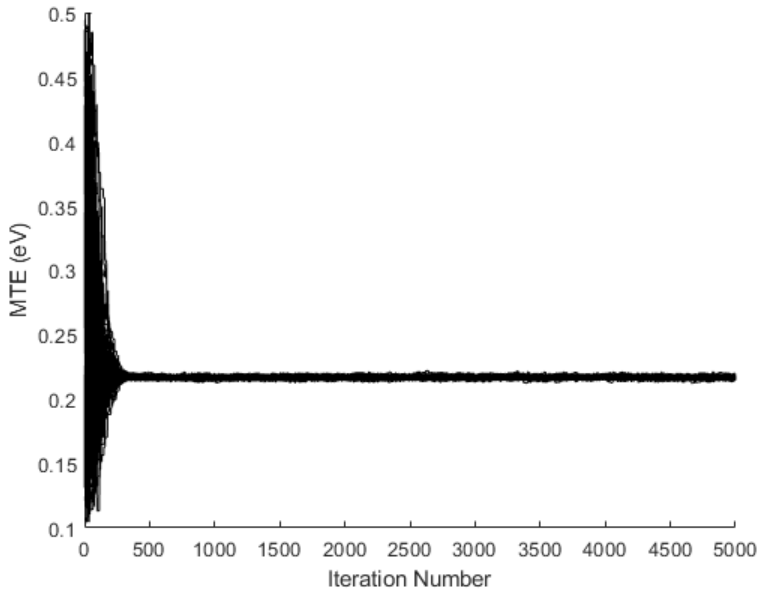


Figure 3.4: Walker convergence for MTE. Note that the walkers start spread out over the entire parameter space, but eventually converge.

depth. At the photocathode, assuming a non-magnetized photocathode (for examples of magnetized cathodes, see [37, 35, 36, 38, 107]), the particle positions and momenta are decoupled. As such, the inital emittance is given by:

$$\epsilon = \sigma \sqrt{\frac{MTE}{mc^2}} \tag{3.9}$$

where σ is the RMS size of the beam at the photocathode. It is apparent that the MTE represents the divergence. A cathode recession also affects how much the beam diverges, due to a focusing effect. Although not the same effect, so the solutions are unique, there is a correlation between the two effects in the posterior distribution, as expected, as can be seen in Fig 3.2 in the lower lefthand plot.

The walker convergence is shown in Fig. 3.4. The walkers are started with uniform probability in the parameter space, but converge to a steady state posterior distribution. The period before convergence is called the burn-in period and is omitted from the analysis in Figs. 3.2 and 3.3. A similar plot can be made in the other dimensions of the parameter space, but the MTE plot is shown as an example.

3.2.4 Measurement Resolution Improvement

As a model-based reconstruction technique, the technique shown above is open to the framework of incorporating measured fluctuations into the reconstruction. At each point in the solenoid scans above, multiple shots were recorded in order to characterize the fluctuations and increase the SNR of stationary noise. The characteristic $\sqrt{N}$ (where N is the number of shots) improvement in the SNR assumes that the measurements are centered around a fixed mean; if this assumption is violated, less or no improvement will be seen by taking multiple shots, depending on the relative scale of the drift (mean fluctuations) to that of the jitter (fluctuations about the mean). Given the state-of-the-art short-term stability of HiRES, which is discussed in the previous section, taking multiple shots to improve the SNR is clearly a good path (in fact, most measurements at HiRES, including the ones shown above inherently average over multiple shots, as the maximum repetition rate is 250 kHz).

However, this is not always the case, as many beamlines have significant jitter and drift components. For example, the jitter and drift of a previous version of Pegasus is discussed in the next chapter. Numerous beamlines suffer from fluctuations, but in a model-based reconstruction technique, if the input parameter fluctuations are tracked synchronously with the measurements, the additional information can be put into the model and used to improve the resolution of the measurement. In the following paragraphs, a simulation study will be undertaken on the Pegasus beamline to show the effect of synchronous data acquisition on the measurement resolution. As discussed in the next section, the Pegasus beamline has a UCLA/SLAC/BNL-style gun, operating at 2.856 GHz, followed by an emittance-compensation solenoid, which is followed by a YAG screen. Notably, this setup differs from the Pegasus setup described in the next chapter, as — prior to the work shown in this chapter — the photocathode back-plane was replaced with an INFN-style plug-compatible back-plane, which was installed with a load-lock system to exchange photocathodes, similar to the one shown in Fig. 3.1 for HiRES. Simulations took this setup difference into account by using a gun field map with a plug recession depth of 750 μm, which is known to be approximately the experimental condition, due to changes in the measured overall operating

Description	**Minimum**	**Maximum**
Solenoid Peak Field (T)	0.1	0.25
Photocathode MTE (eV)	0	2
RMS x Laser Size (μ m)	30	500
y/x Laser Size Ratio	0.7	1.3
Laser Rotation (rad)	0	pi
RMS Laser Pulse Length (fs)	50	200
Gun Peak Field (MV/m)	60	90
Gun Phase (deg)	5	30
Beam Charge (fC)	50	650
Photocathode Recession Depth (mm)	0	1.5

Table 3.1: List of free parameters and their limits for the feed forward neural network surrogate model for GPT for Pegasus

frequency of the gun.

The simulation of the beamline was accomplished on a feed forward neural network surrogate model for the GPT [56] model of Pegasus, with five hidden layers with 10 nodes each, which predicts the beam spot sizes in x, y, u and v to within 5 μm. Training data comprised approximately 10^5 samples. Free parameters are listed in Table 3.1. The model assumes a uniform transverse profile of the beam in particle positions, which is a reasonable assumption given the the small beam sizes at Pegasus, which are created by demagnifying and imaging an aperture that cuts the core a Gaussian laser pulse. The remaining spatial dimension of the initial 6-D phase space (discussed in Chapter 1) is a Gaussian with RMS pulse length left as a free parameter in simulation. The transverse angles of the initial 6-D phase space distribution are Gaussians as determined by the photocathode MTE. The initial forward momentum is a half-Gaussian distribution determined by the photocathode MTE.

The goal of this study is to show the improvement in the accuracy of the reconstruction of the photocathode MTE by incorporating the shot-to-shot measured fluctuations upstream

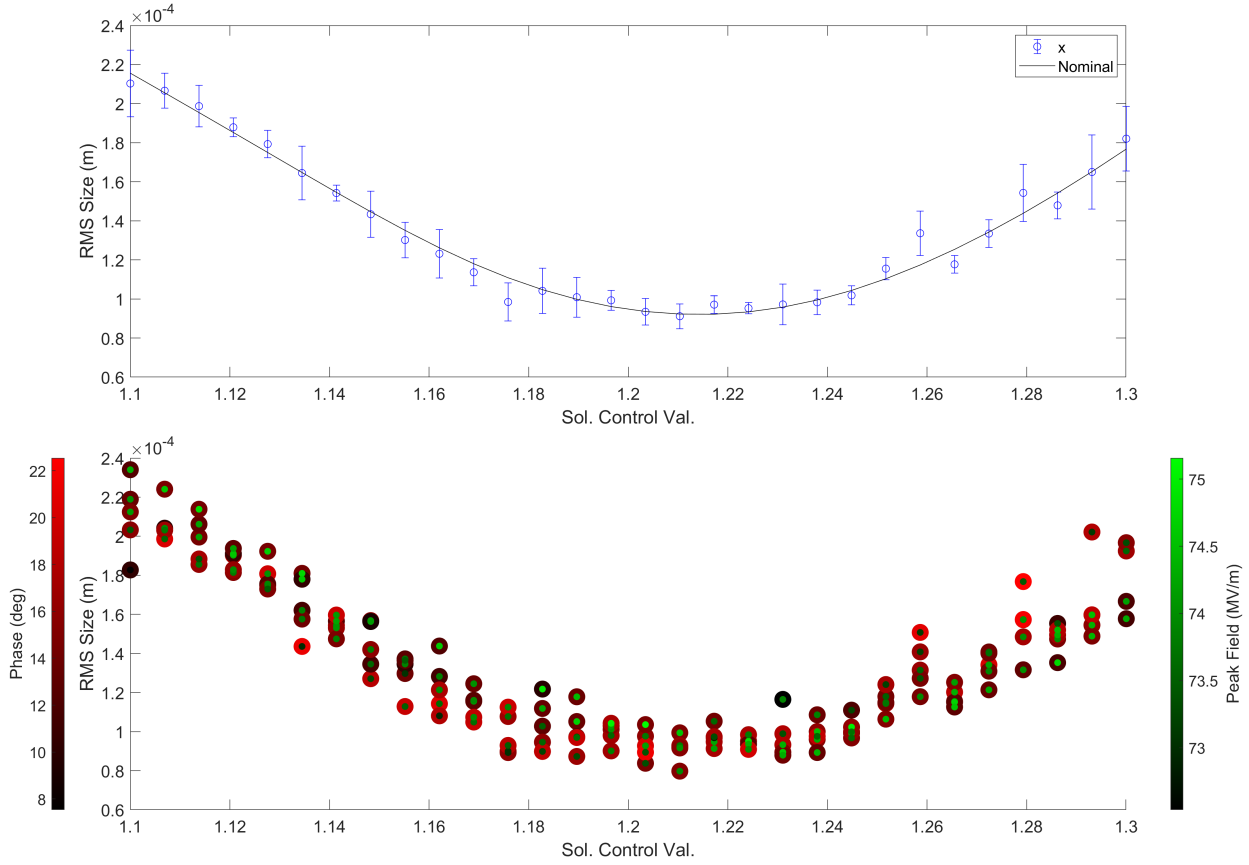


Figure 3.5: Schematic of the approaches. The same simulated data are plotted twice using the traditional approach (top) and the approach in this section (bottom). Top: Plot of RMS beam size in the x-dimension vs. solenoid control value. Amplitude and phase jitter is averaged at each solenoid value. The black line represents a nominal scan — the same scan with no jitter. Bottom: RMS beam size in the x-dimension for individual shots tagged by RF phase and amplitude.

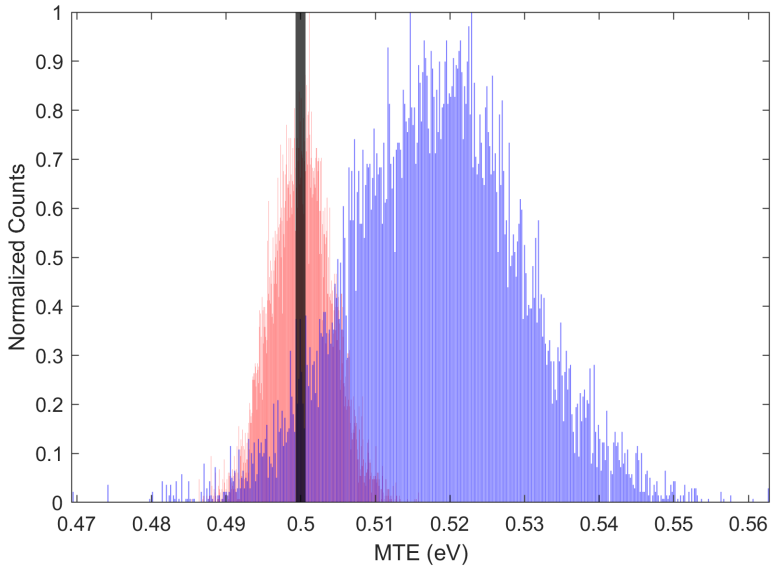


Figure 3.6: Red: Posterior projection while taking into account the shot-to-shot jitter with 5 measurements per solenoid setting. Blue: Posterior projection using the traditional method of averaging. Ground truth MTE is shown in black.

accelerator parameters. The comparison between the case of simply averaging over the fluctuations and taking into account each shot is shown in Fig. 3.5, which was created by sampling two Gaussian distributions of phase and amplitude centered about a nominal value. Each point (amplitude and phase) in parameter space was simulated using the model described above in order to give an RMS beam size.

The results of the projected posterior distribution of MTE is reconstructed using MCMC and is shown in Fig. 3.6. Something to note is that since MCMC now sees more points, the uncertainty goes down, leading to more confident predictions. Note that Fig. 3.5 and Fig. 3.6 use five measurements per solenoid setting. Note that the uncertainty in the posterior distribution goes down accordingly by a factor of $\sqrt{N}$ where N is the number of shots using MCMC. See Fig. 3.7 for a plot of posterior MTE uncertainty as a function of number of samples. By increasing the number of samples per solenoid setting to 50 shots per solenoid setting, note that the uncertainty improves accordingly, while no improvement is seen for the traditional, averaging scan in Fig. 3.8.

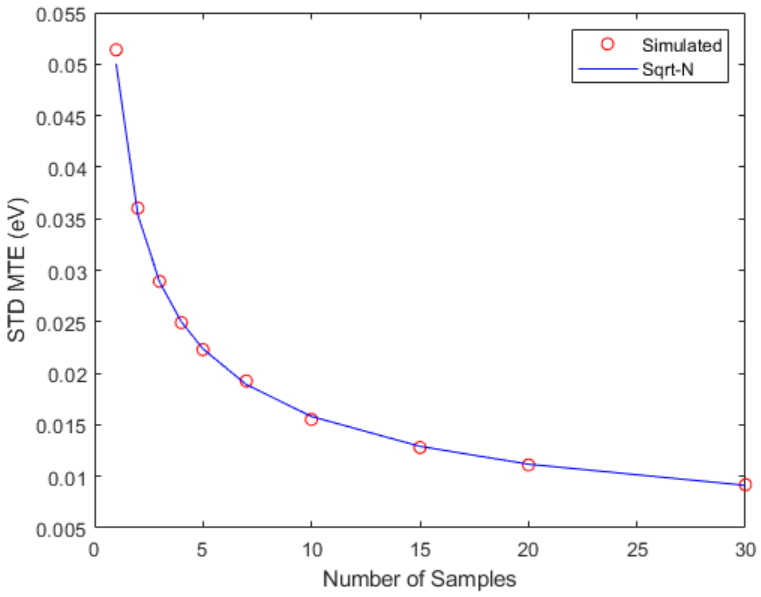


Figure 3.7: $\sqrt{N}$ dependence of the projected posterior uncertainty of the MTE

However, as is seen in Fig. 3.8, by injecting additional information into the fit (in this case RF amplitude and phase), the uncertainty reduction is even greater than $\sqrt{50}$ between the shot-to-shot versus averaging case. This is because even with 50 samples per point, there is some error associated with approximating the sample mean as the population mean. This can result in a different MTE.

Similarly, as the fluctuations become large, the MTE appears to become larger. This is because at different input settings, different solenoid settings will bring the beam to a minimal size at the screen. The effect of having a different beam energy at each point results in different focusing effects, which appear to "wash out" the minimum, resulting in an apparent MTE increase. This effect is shown in Fig. 3.9. While the fluctuations shown to the right of this plot are too high to accurately represent the fluctuations at Pegasus, it shows that by using knowledge of the RF phase and amplitude, one can compensate for this increase.

In summary, if data are available, it helps the accuracy and confidence in the MCMC predictions of the MTE to include upstream fluctuations. This is due to three reasons:

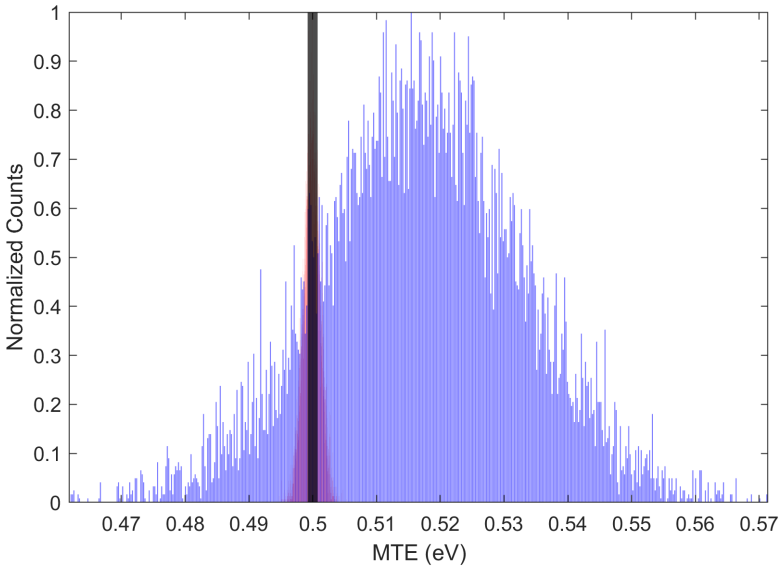


Figure 3.8: Red: Posterior projection while taking into account the shot-to-shot jitter with 50 measurements per solenoid setting. Blue: Posterior projection using the traditional method of averaging. Ground truth MTE is shown in black. Note that the standard deviation of the posterior distribution using the traditional method is 0.0157 eV, whereas the new, shot-to-shot method produces 0.0013 eV using the same metric.

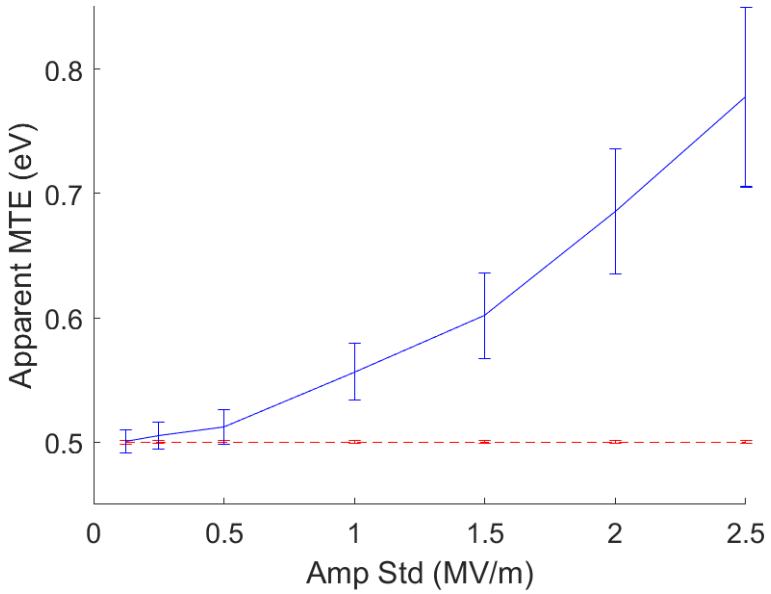


Figure 3.9: Apparent MTE increases with upstream fluctuations.

- Increased number of data points for MCMC
- Increased information injected into the fit
- Reduction in the "washout" of the MTE by having MCMC aware of the different focusing that occurs due to the fluctuations

3.3 FAST

3.3.1 FAST Background

In a similar problem, matching a physical model to experimental results at the Fermilab Accelerator Science and Technology (FAST) will be highly beneficial for several ongoing and upcoming projects. These projects include providing beam for Integrable Optics Test Accelerator (IOTA) [108], and the under-construction FAST Gamma-Ray Electron Enhanced Source (GREENS) [109] and for upcoming photocathode studies in a high-gradient photoinjector proposed by the Center for Bright Beams. FAST is a very versatile machine, and as such, does not satisfy the first condition listed in Section 3.1, as making the experimental training data for all modes of operation would be incredibly costly.

Having a physics-based model for photocathode studies, as mentioned in the previous section will be required, as such studies require a model to invert/optimize to deduce photocathode properties from solenoid scans (e.g. [21]). For FAST-GREENS, the requirements on peak 5-D beam brightness from the injector are particularly demanding. To these ends, FAST-GREENS will require the FAST injector to run at near the boundary of capability on the Pareto front of beam charge vs. transverse emittance. Indeed, the nominal requirements of 3 mm mrad emittance in each dimension, along with a peak current of 600 A will require optimization of the upstream injector parameters on a model. This section reports on the development of a high-fidelity model of FAST for optimization toward high peak 5-D beam brightness for FAST-GREENS. IOTA experiments could also benefit from such a model, if further optimization is needed for beam properties.

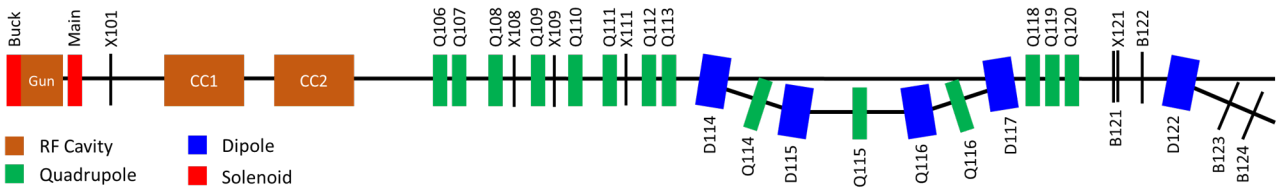


Figure 3.10: The FAST injector, with selected diagnostics shown. Q106, Q107 and Q111 are 45 degree skew quadrupoles. Diagnostics with X delineation are screens (YAG or OTR), while those delineated with a B are noninvasive beam position monitors (BPM).

The FAST beam line (shown in Fig. 4.3) starts with the gun, of similar type to the DESY-PITZ gun [110]. The 1.3 GHz, 1.5 cell cavity is followed by an emittance compensation solenoid, with adjustable position. The bucking solenoid directly upstream of the gun is generally set to produce a magnetic field in the opposite direction of the main solenoid at a magnitude such that the field at the photocathode is zero, so that the beam is born with zero angular momentum. The gun as an INFN-style load lock system that allows for the exchange of photocathodes, similar to the one shown in Fig. 3.1. Two TESLA cavities [111] (CC1 & CC2 in Fig. 4.3) are used to accelerate the beam to 35-50 MeV in nominal cases. A series of quadrupoles are used to manipulate the beam transversely. The quadrupoles are followed by a chicane that is used to compress the beam. The chicane can also be bypassed and the beam can be directed from Q113 to Q118, the start of a quadrupole triplet. Following B122, the beam energy can be measured on the beam dump line.

As mentioned above, one of the primary goals of this study is to provide a model for optimization of compression in the chicane. In a chicane, beam emittance will be diluted by coherent synchrotron radiation (CSR) effects, which must be modeled properly. To these ends, General Particle Tracer (GPT) [56] has been shown to be a good choice to model CSR effects [112]. As such, GPT was chosen as the physics model to use in this effort.

There are numerous free parameters for the model. Table 3.2 lists the free parameters and the data taken to fit them.

Parameter	Data
Main Solenoid Position	Main Solenoid Scan & Downstream Images
Bucking Solenoid Position	Bucking Solenoid Scan
X101 Position	Main Solenoid Scan & Downstream Images
Gun Peak Field	CC1 Energy Scan
Gun Phase	Gun Phase Scan
Initial Pulse Length	Gun Phase Scan & Streak Camera Images
Beam Initial Transverse Size	Virtual Cathode Image
Solenoid Strengths	Solenoid Magnetic Measurements
CC1 Peak Field	CC1 Energy Scan & Downstream Images
CC1 Phase	CC1 Energy Scan

Table 3.2: Free parameters in the model and the tests completed to fit these parameters.

3.3.2 Longitudinal Dynamics

In order to verify the longitudinal dynamics of the beam and match a model, two tests were performed: a gun phase scan and a streak camera measurements at X121 (see Fig. 4.3), approximately 17.4 m downstream of the photocathode. Both will be described below. In addition, a scan to verify the energy of the beam before and after CC1 (see Fig. 4.3) is shown.

For the first test, the gun phase was changed in 0.2 degree intervals and extracted charge was measured at the exit of the gun. The time it takes for the beam to be extracted from the cathode can be measured from this scan. Neglecting the Schottky effect, under purely classical conditions, where the laser induces above-threshold photoemission, the measured charge should start as soon as the RF phase is such that the field accelerates the head of the bunch downstream. As the phase is increased, more of the bunch experiences an accelerating field, until the full bunch does, at which point the full beam charge is measured. Based on the degrees of phase sampled between first electrons and full beam, the length of the pulse from the gun can be determined. For a uniform distribution, the rise would be linear, but

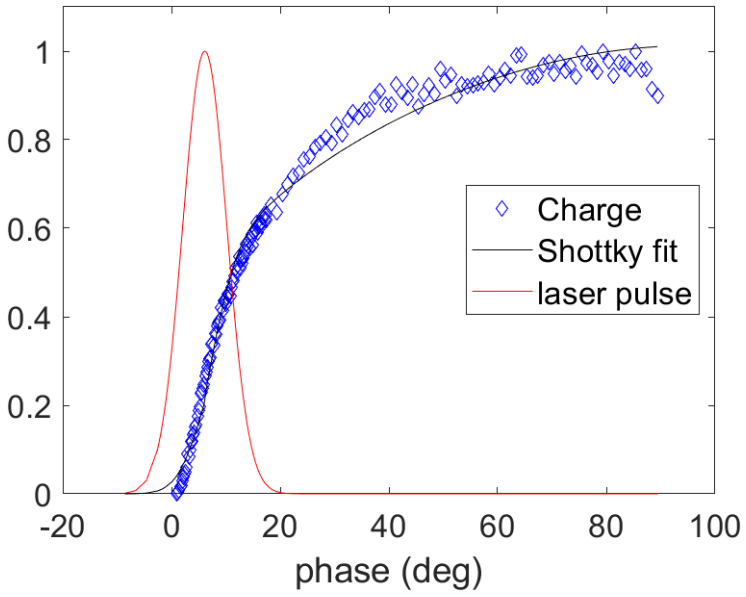


Figure 3.11: RMS laser pulse length fit with an adjustment for the Schottky effect due to quantum tunneling probabilities.

for a Gaussian beam, the rise can be modeled as an error function.

However, this is a simplified picture. In fact, as the RF phase changes, the magnitude of the field at the cathode also changes, which changes the tunneling probability of the electrons. As such, the Schottky enhancement on the quantum efficiency can be modeled as proportional to:

$$QE \propto (E_{excess})^2 \tag{3.10}$$

where E_{excess} is given in Eq. 1.17, but is repeated here with more detail on the contributions to the barrier to photoemission:

$$E_{excess} = h\nu - \phi_w + \phi_{Schottky} \tag{3.11}$$

and

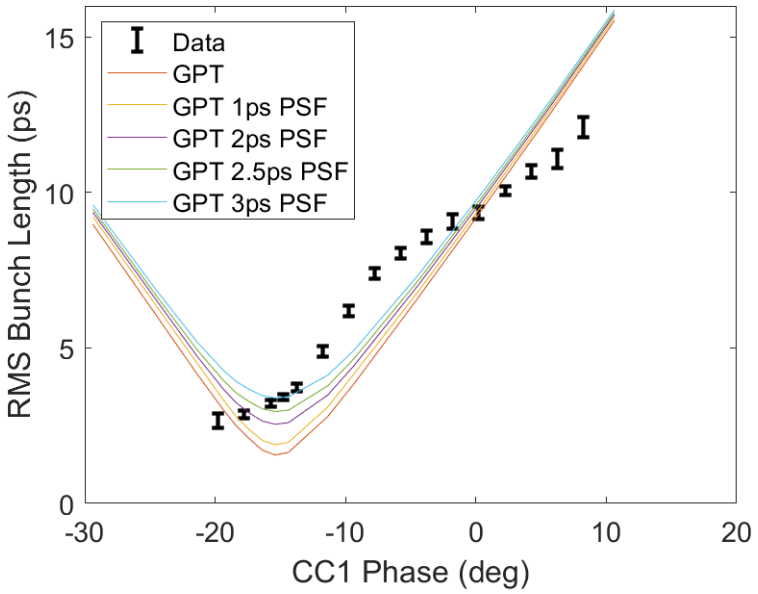


Figure 3.12: Measured compression using CC1 and the chicane for compression. Streak camera measurements are overlaid with GPT predictions convolved with a resolution term (PSF).

$$\phi_{Schottky} = \sqrt{\frac{e}{4\pi\epsilon_0}}\sqrt{\beta E_0 \sin(\theta)} \tag{3.12}$$

where $h\nu$ is the photon energy and ϕ_w is the nominal work function of the photocathode. E_0 is the magnitude of the electric field, while θ is the RF phase. In practice, the constants, $\sqrt{\frac{e}{4\pi\epsilon_0}}$ can be lumped into β, the Schottky enhancement factor, in a fit. As such, the error function can be combined with the Schottky fit to fit the initial pulse length. At FAST, this yielded an 8.6ps RMS initial pulse length, as shown in Fig. 3.11.

Additionally, the electron beam bunch length was measured downstream using a streak camera setup at X121. This measurement used an OTR screen and the used the streak camera method established in numerous studies (e.g. in [43]) to characterize the pulse length. These yielded comparable results to the GPT simulations, when factoring in a resolution term due to the width of the collimating slit width for the streak camera measurement. See Fig. 3.12 for details.

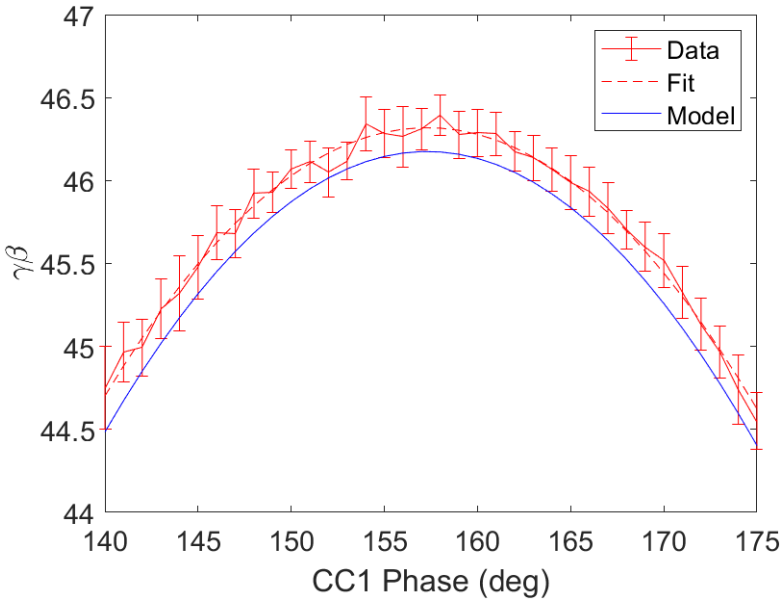


Figure 3.13: Kinetic energy measurements as a function of CC1 phase. Measured using D122 in Fig. 4.3.

The beam energy was measured using a dipole spectrometer. Using the geometry of the system and the magnetic field in D122 (as shown in the Fig. 4.3), the beam's kinetic energy can be calculated using the relationship:

$$B\rho = pc/e \tag{3.13}$$

where B is the magnetic field, ρ is the bending radius of the magnet and p is the beam momentum. ρ is determined by the geometry of the system. By varying the phase in CC1, the beam energy can be controlled directly. Given that the a sinusoidal energy kick imparted by CC1, the kinetic energy of the beam at the exit of the gun can also be deduced.

$$E_{kin} = E_{gun} + E_{CC1}\cos(\theta) \tag{3.14}$$

The geometry of the system is determined by four non-destructive BPMs, B121, B122, B123, and B124 in Fig. 4.3. The latter two of these BPMs are in the dogleg beamline. As such, both the beam position and angle of the beam can be determined before and after the

dipole. While the NMR probe inside the dipole gives a readback on the magnetic field inside the dipole, this measurement has systematic (mostly due to calibration errors and physical misalignment) and random error associated with it. As such, the energy in the preliminary GPT model is adjusted slightly to match the transverse behavior described below. See Fig. 3.13.

3.3.3 Transverse Dynamics

On the transverse side, a virtual cathode camera image was recorded on each run day. This was used to start a distribution of particles in simulation.

Solenoid scans were measured on screen X101, which has a position that is not recorded in the survey data of the beamline. See Fig. 3.14 for more details. The position of the screen is a free parameter within a range of the estimated position of 1.062 meters downstream of the photocathode. The screen was moved approximately 5cm upstream in simulation relative to its estimated position. Note that in simulation, the positions of the solenoids are also free parameters, as there is no reliable survey data on the location of these. The main solenoid was positioned 32 cm downstream of the photocathode, nearly 10cm downstream of its estimated position.

Importantly, the transverse beam sizes were measured at X108 (when available), X109 and X111 and compared to simulation for various charges. See qualitative agreement with preliminary model in Fig. 3.15.

The transverse emittance was measured using the two quadrupole scan method [92]. Generally, the emittance is higher than predicted by the preliminary model. For example, the emittance of an approximately 700pC beam was approximately 20 mm mrad in each dimension, while the simulated RMS normalized projected emittance was approximately 9 mm mrad in x and 17 mm mrad in y. There are several possible reasons for this discrepancy, which are discussed, along with sources of error, in the following paragraphs.

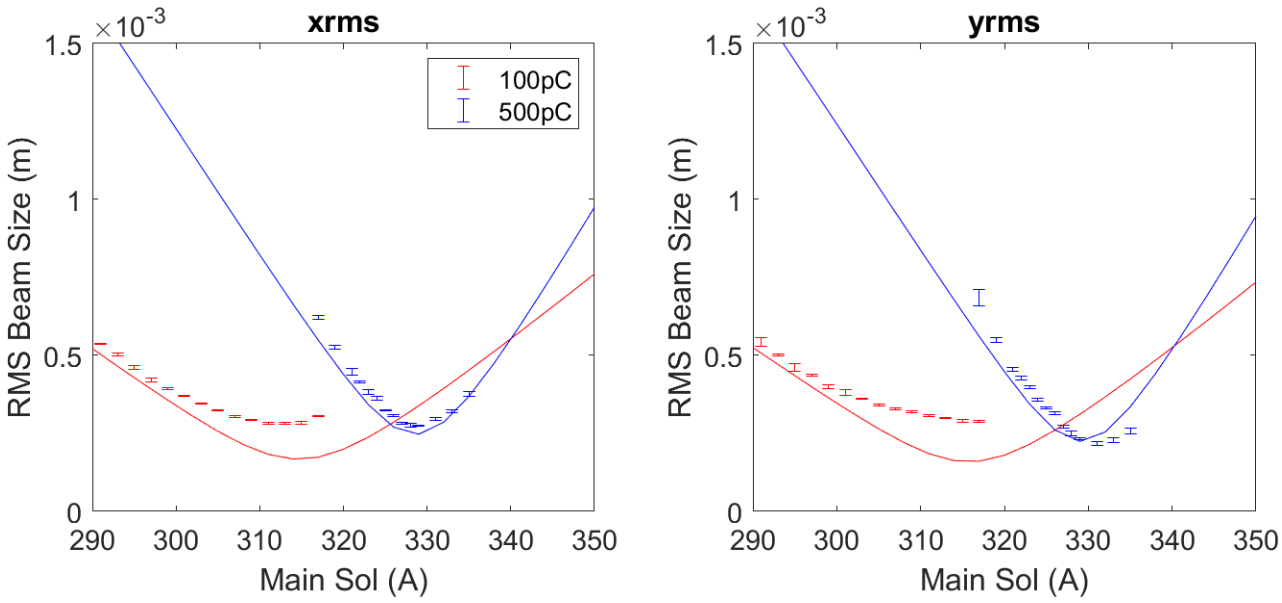


Figure 3.14: Example solenoid scans at multiple charges with GPT predictions overlaid. Measured on X101 (see Fig. 4.3).

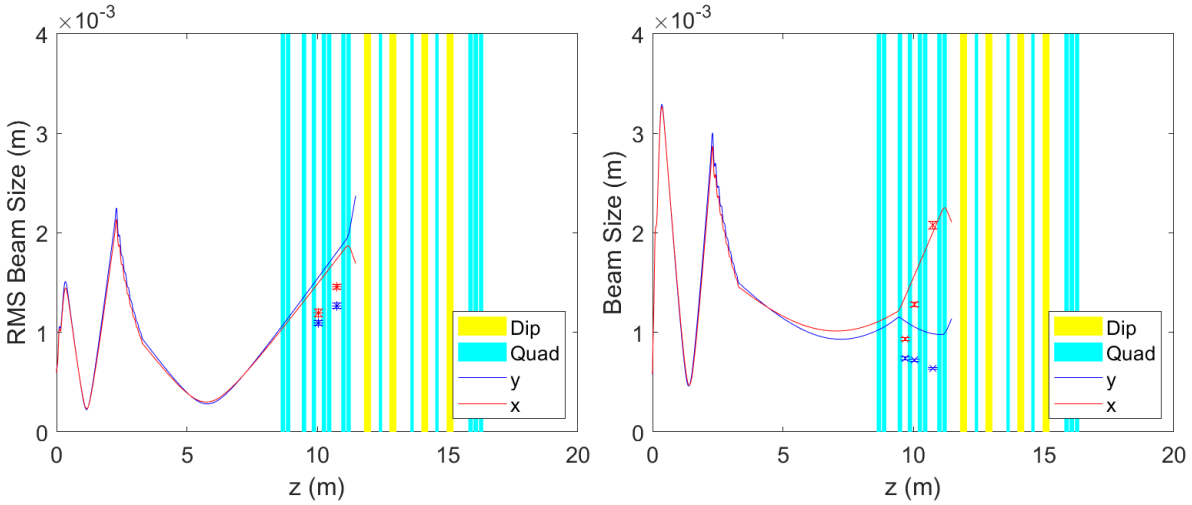


Figure 3.15: Downstream beam second order moments with GPT predictions for an approximately 100pC beam (left) and an approximately 700pC beam (right). Measurements were taken at X108, X109 and X111 (see Fig. 4.3).

3.3.4 Discussion

A preliminary GPT model of FAST is presented. While the model generally qualitatively matches the behavior observed in experiment, there are still discrepancies that will continue to be improved.

For example, the a scan of the bucking solenoid does not yield the same dynamics as the model indicates. A bucking solenoid scan at constant charge and main solenoid setting was taken on X101. While it shows a similar beam size to what would be expected in simulation, the behavior as the solenoid is increased does not match.

This could suggest several issues, including that the main solenoid position is closer to the cathode, or that the X101 position has been incorrectly fit. However, these positions allow the model to match the downstream behavior of the beam. Additionally, the presence of quadrupole focusing is evident in the solenoid scan data, but is not included in the model, and could have some effect on the bahavioral discrepancy.

This asymmetry could be one of the main reasons for a larger emittance value in the experiment than in simulation. A skew quadruople would increase the projected emittances.

In the future, these quadrupoles in the solenoid will be incorporated into the model, as will CC2, which was inoperative for the majority of these tests. Additional survey data would greatly help in finalizing the model. Additionally, absolute calibration of integrating current transformers measuring the beam charge should be confirmed. Relative charge calibration has been confirmed using beam image intensities.

3.4 Conclusion

Physics-based models are generally useful to a class of diagnostics and control problems in beam physics. Despite their slower speed than other tools available, they are reliable and versatile. This can lead to numerous applications, some of which are showcased in this section, including for control, as at FAST and for understanding underlying physics, as

shown at HiRES, Pegasus (simulation) and in the future, FAST for photocathode studies.

Notably, for MCMC, the effect of the slowness of the physics-based models was mitigated by using surrogate models for rapid execution, as MCMC is a sample-inefficient method. The tradeoff is that MCMC gives more information about the underlying probability distribution than other methods. Of course, the use of a surrogate model also has its tradeoffs; the price to be paid for using these surrogate models is that the working range of simulation must be defined in advance, and as such is not as general as working on the simulation itself.

This chapter focused on matching these physics models to the real behavior of the beamline. This was accomplished by MCMC and coarser methods for FAST. One of the main goals is the same: to have a model high-fidelity physics model, complimentary to the experimentally driven models shown in the other chapters of this thesis.

CHAPTER 4

Modeling in the Presence of Fluctuations

4.1 Motivation & Background

As described in previous chapters, accelerator models can be particularly useful for improving diagnostics: both offline for data interpretation and analysis (see Chapter 3) and online for virtual diagnostics (see Chapter 2). In this chapter, modeling toward control and its augmentation with model-independent methods for further automation are discussed.

In general, models for control guide the accelerator settings needed in order to obtain certain beam properties. Note that this problem is the inverse of most physical models, in which accelerator settings yield beam properties. In general, for most simulation codes, a solution to the inverse problem must be found. These solutions include optimization, but could also include sampling techniques such as MCMC, as shown in the previous chapter.

This chapter attempts to automate and improve control processes in a realistic space, where existing models of a more idealized system — where fluctuations are not fully taken into account — are used to predict a real system, which is laden with such fluctuations. These models can be used for optimization to find the accelerator parameters need to produce certain beam properties (model-based control). Any differences between the desired beam and the beam resulting from model-based control is compensated by model-independent techniques, which are techniques that optimize the beam without using physical knowledge of the system. Since physics-based simulations are generally slow to execute, fast fluctuations necessitate a faster online modeling approach. Thus, ML-based, fast surrogate models are beneficial.

However, the price to be paid for using such a fast tool is the limitation in the range of possible parameters. While ML models are generally faithful to the true behavior inside the span of parameter space of the training data, the problem of extrapolation under so-called distribution shift is an open research question. Solutions including a "soft landing" [60, 81] outside the span of the training set have been proposed.

In this work, a so-called "warm start" approach is shown, where a ML-based surrogate model is trained on experimental data, but due to anticipated system drift around the boundaries of the training set and untracked machine fluctuations, the model is treated only as an approximate guide. The predictions are used as an approximate solution, something nearby to the true solution in parameter space that can be found quickly using online model-independent tools.

In the first part of this chapter, the optimization of the round-to-flat beam transform (FBT) at Pegasus is discussed. At the time of this experiment, Pegasus had numerous short- and long-term fluctuations, including in RF and magnet systems, which have since been mitigated through major hardware overhauls, including the replacement of the LLRF systems with a digital clock and the entire RF amplification chain with a state-of-the-art solid-state modulator and klystron. Further improvements to software, magnet alignment and data logging have made operations more consistent. Other improvements include the introduction of a Bergoz integrating charge transformer and beam charge monitor and an X-band linearizer [113]. The work toward FBT in the previous, fluctuation-laden environment will be described in the following sections, leading to the need for compensation with model-independent methods. This chapter borrows heavily from [107] and in-preparation materials.

In the second part of this chapter, work toward automating the electron beam setup for UED in a dynamic environment at HiRES is shown, further demonstrating the utility of model-independent optimization techniques for the "warm start" approach. This section borrows from [114].

4.2 Flat Beam Transform at Pegasus

4.2.1 Flat Beam Theory

While maximizing 4-D or 6-D (discussed in Chapter 1) beam brightness is often critical, for certain cases, maximizing 2-D beam brightness is sufficient. This can be accomplished by minimizing the emittance in one transverse plane at the expense of the other, leading to asymmetric projected emittances. Applications of a beam with large 2-D brightness include slab-geometry dielectric laser acceleration [115], as the acceptance in one spatial dimension is small, while the acceptance in the other dimension is orders of magnitude larger. Other applications of flat beams include those that require intense line foci, such as those in electron microscopy with certain sample geometries or lithography.

There is a fundamental, physical reason to minimize the 2-D beam brightness associated with Gromov's Non-Squeezing Theorem [116]. This theorem states in symplectic transformations, such as those of linear beam transport, there is a limit to the minimum emittance in any dimension. Indeed, the eigenemittances are a second set of invariants for beam transport, beyond the phase space volume [117, 118]. Thus, the eigenemittance is the lowest limit of the projected emittance corresponding to any spatial dimension. It is the goal of the FBT to minimize one such eigenemittance (at the expense of the other) and then to retrieve it as the projected emittance in a preferential direction.

Eigenemittances are defined as follows from the beam matrix (see Eq. 1.11) [93, 119]:

$$\epsilon_1 = \frac{1}{2}\sqrt{-\mathrm{Tr}[(\Sigma^{4D}J)^2] + \sqrt{\mathrm{Tr}^2[(\Sigma^{4D}J)^2] - 16\det[(\Sigma^{4D})]}} \tag{4.1}$$

$$\epsilon_2 = \frac{1}{2}\sqrt{-\mathrm{Tr}[(\Sigma^{4D}J)^2] - \sqrt{\mathrm{Tr}^2[(\Sigma^{4D}J)^2] - 16\det[(\Sigma^{4D})]}}, \tag{4.2}$$

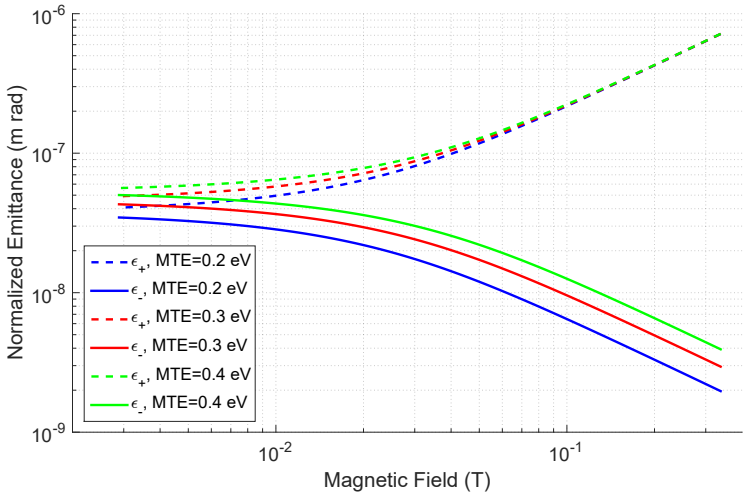


Figure 4.1: Analytical scaling of large and small projected emittances following the properly-optimized FBT with respect to the applied magnetic field at the photocathode and the MTE characteristic of the photocathode.

where J is the symplectic matrix,

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{4.3}$$

In order to make one eigenemittance small, the beam is born in a longitudinal magnetic field on the photocathode. This magnetic field gives the beam angular momentum, is preserved as canonical angular momentum during transport. A beam born from a magnetized cathode has a beam matrix as follows:

$$\Sigma^{4D} = \begin{pmatrix} \sigma_c^2 & 0 & 0 & \mathcal{L} \\ 0 & \sigma_c'^2 + \frac{\mathcal{L}^2}{\sigma_c^2} & -\mathcal{L} & 0 \\ 0 & -\mathcal{L} & \sigma_c^2 & 0 \\ \mathcal{L} & 0 & 0 & \sigma_c'^2 + \frac{\mathcal{L}^2}{\sigma_c^2} \end{pmatrix} \tag{4.4}$$

where σ_c is the RMS beam size on the cathode, σ_c' represents the initial angles in the beam due to the photocathode MTE in the absence of a magnetic field and

$$\mathcal{L} = \frac{eB_c}{2\gamma\beta mc}\sigma_c^2 = \frac{L}{\gamma\beta} \tag{4.5}$$

is proportional to the angular momentum of the beam, where B_c is the magnetic field on the cathode, and γ and β are the relativistic factors. L is also proportional to the angular momentum of the beam.

By inserting Eq. 4.4 into Eqs. 4.1 and 4.2, it can be shown that the two normalized eigenemittances, hereafter referred to as ϵ_+ and ϵ_- are given by [33]:

$$\epsilon_\pm = \sqrt{\epsilon_{th}^2 + L^2} \pm L \tag{4.6}$$

where

$$\epsilon_{th} = \sqrt{\frac{MTE}{mc^2}\sigma_c^2} \tag{4.7}$$

where ϵ_{th} is the normalized thermal emittance of the cathode and depends on the Mean Transverse Energy (MTE) of the photoemission process at the cathode.

As such, if the magnetic field at the cathode is sufficiently large, $L >> \epsilon_u$ and

$$\epsilon_- \approx \frac{MTE}{eB_c c} \tag{4.8}$$

and

$$\epsilon_+ \approx 2L + \epsilon_-. \tag{4.9}$$

Equation 4.8 shows that a flat beam has a small eigenemittance that is independent of the initial spot size; in fact, it only depends (approximately) on the MTE of the photocathode and the magnetic field B_c that is applied to the photocathode. See Figure 4.1 to see how the large and small projected emittances following the optimized FBT scale with the B_c. Given the progress on advanced photocathodes that have very low MTE [120], one can thus obtain extremely small eigenemittances.

The last step of the FBT is to retrieve these eigenemittances as the projected emittances in either the vertical or horizontal dimension. This is accomplished by tuning a skew

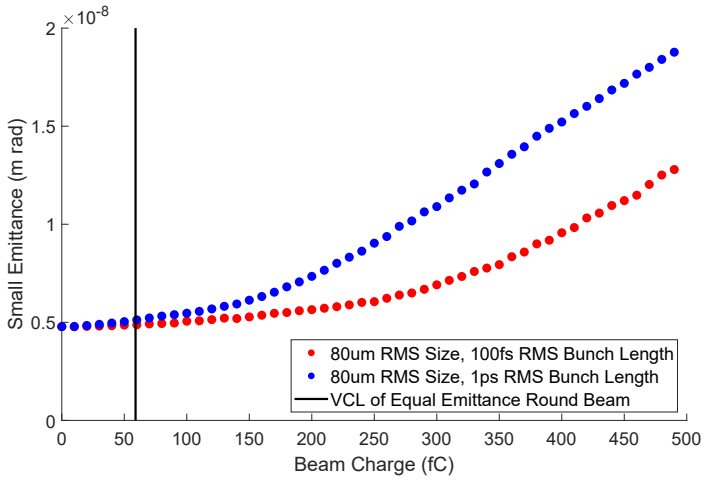


Figure 4.2: From simulation: minimum optimized small projected emittance as a function of charge. Comparison with a round beam of equal emittance to the small projected emittance in the zero-charge limit; the virtual cathode limit of such a beam is denoted in black.

quadrupole triplet to remove the angular momentum from the beam and suppress the coupling terms in the beam matrix shown in Equation 4.4. For the purposes of this chapter, FBT is a multidimensional optimization problem in a jittering environment. Thus, the remainder of the chapter will focus on this last step of the FBT: the tuning of the skew quadrupoles to find optimal values such that a flat beam is produced at the exit of the skew quadrupole triplet. Since the theory presented above is matrix-based and assumes zero charge, for a beam of a given energy, it can be used to solve for the three skew quadrupole currents. These assumptions generally do not hold for the real system. In fact, as shown in Figure 4.2, the minimal small emittance value changes as a function of charge, as do the optimal FBT quadrupole values as beam charge is increased. Thus, optimization that was already necessary because of drift, noise and incomplete knowledge of beam parameters at all times, becomes even more necessary at higher charge.

However, despite the requirement for optimization, the charge scaling for the 2-D beam brightness is generally better than a round beam, and makes it an attractive option. Figure 4.2 shows the scaling of the small projected emittance with charge from particle tracking simulation [56] for Pegasus with 400 meV MTE and 3.375 kG peak solenoid field, parameters

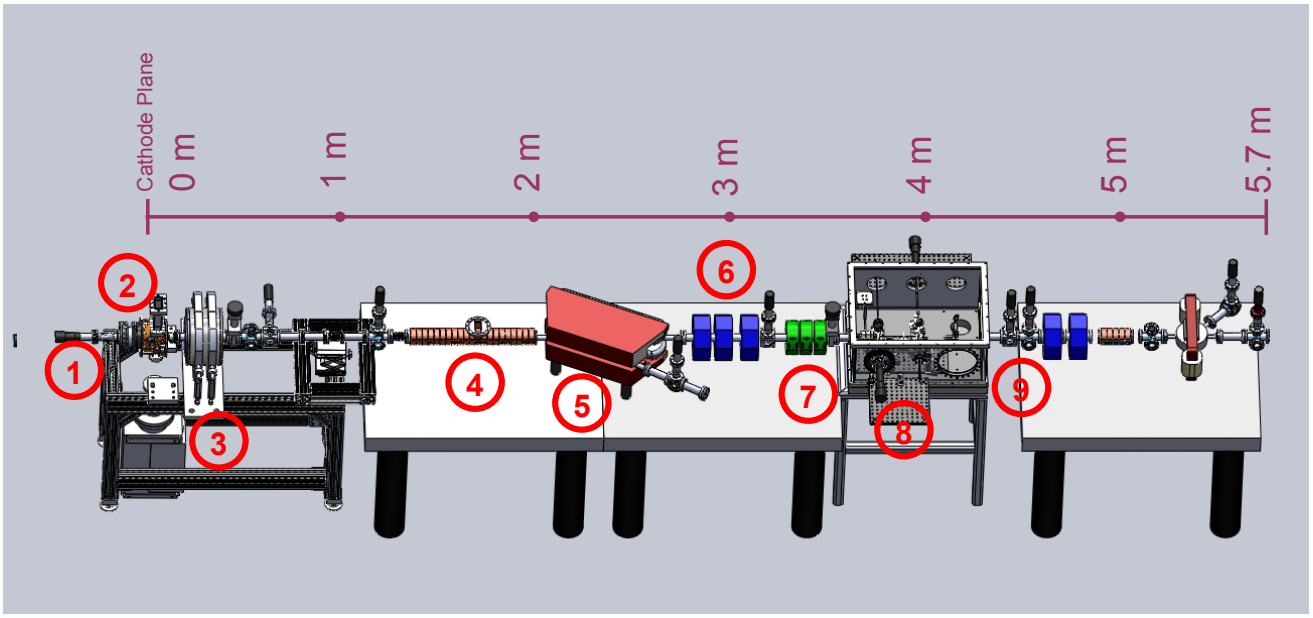


Figure 4.3: Model of the Pegasus Beamline at UCLA with selected distances. The elements are as follows: 1) Permanent magnet mounted on actuator to introduce a magnetic field on the cathode, 2) 1.6 cell RF gun, 3) Emittance compensation solenoid, 4) Booster linac (not used in this experiment), 5) Dipole spectrometer, 6) Flat beam adapter comprised of a skew quadrupole triplet, 7) normal quadrupole triplet for focusing, 8) TEM grids and pepperpots for emittance measurement, 9) Diagnostic screens

which are relevant for the experimental work discussed later. Without the FBT, to generate a round beam with an equivalent emittance, one would have to keep the laser RMS spot size on the cathode to less than 5.4 μm. The extracted charge for a beam this small would be limited by the high charge density. In Fig. 4.2, the vertical line indicates the virtual cathode limit [22, 121] for such a beam in the pancake regime at approximately 59 fC, given experimental phases. As can be seen, it is advantageous to maximize 2-D beam brightness with a flat beam compared to a round beam.

4.2.2 Experimental Setup

The three-skew-quadrupole approach for the FBT [34, 32, 33, 37, 35, 36, 38] at the Pegasus beamline in the ultralow charge regime is presented. At the Pegasus beamline (shown in Fig. 4.3), 800-nm-wavelength laser light is frequency-tripled and 266-nm-wavelegnth and 100-fs-rms laser pulses are incident on an adjustable iris. That iris is imaged onto the photocathode. The transverse spot size of the laser on the cathode can be controlled in the range 50-400 μm rms. The intensity of the light, and therefore the charge of the resulting electron beam, is adjustable using a polarizer. The beam charge was kept below 1 pC in this experiment. The emitted electron beam is accelerated in a 1.6 cell UCLA/SLAC/BNL type radiofrequency photoinjector (2.856 GHz) to approximately 3.65 MeV. A 35 mm x 35 mm x 17.4 mm neodymium (N38) permanent magnet is mounted directly behind the photocathode on an actuator with approximately 10 cm range of motion. When the permanent magnet is closest to the emitting surface of the photocathode, the maximum magnetic field is approximately 0.3375 T; when the magnet is farthest from the photocathode, the magnetic field is negligible — on the order of 10 Gauss. A plot of the on-axis field measurements is shown in Fig. 4.4. Per guidance given by particle tracking simulations using General Particle Tracer (GPT) [56], the emittance-compensation solenoid is set to focus the beam near the entrance of the skew quadrupole triplet that is located approximately 2.8 m from the photocathode. The skew quadrupoles used for the flat beam adapter have an effective length of 10.5 cm, a gradient of approximately 0.484 $\frac{\mathrm{T}}{\mathrm{A\ m}}$ and are followed by normal quadrupole triplet. These normal quadrupoles have an effective length of 7.68 cm and a gradient of approximately 0.45 $\frac{\mathrm{T}}{\mathrm{A\ m}}$. The beam was measured on two screens 4.46 m and 4.545 m downstream of the cathode. A focusing quadrupole triplet (not skew) followed the flat beam adaptor and a screen between the quadrupole triplets was used to quide optimization.

4.2.3 Optimization Considerations

To identify properly a flat beam, one must know the determinants (emittance squared) of the three unique blocks of the 4-D beam matrix, which contains all of the second-order

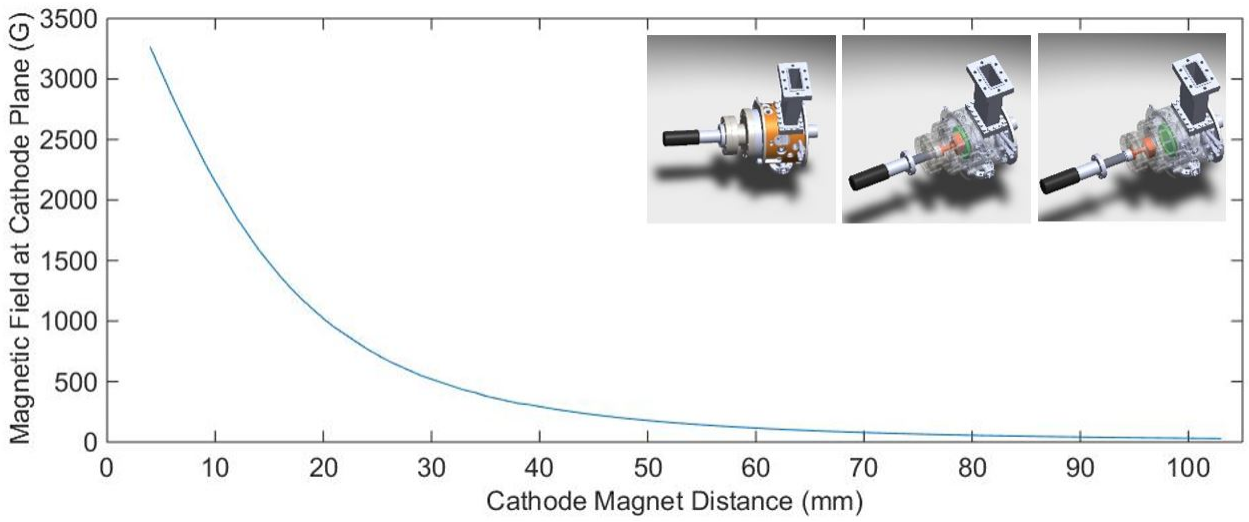
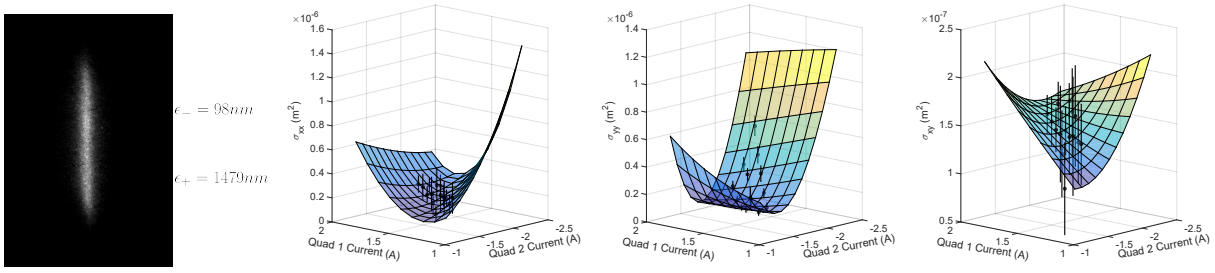


Figure 4.4: Magnetic field on the front face of the cathode as a function of distance from the front face of the nyodymium magnet. Inset (left to right): Gun with magnet actuator assembly, schematic of the magnet against the back of the cathode (green), schematic of the magnet positioned far from the cathode (green).

moments of the beam distribution function (see Eq. 1.11). In practice, one must find each of the second-order moments in order to find these determinants. Techniques that can reconstruct the 4-D beam matrix are well-developed. One such technique is the multi-shot, two-quadrupole scan [92]. The two-quadrupole scan is a multi-shot technique, which requires enough shots that it is impractical to use as an objective function for an optimizer, but can be used to validate the objective function. See Figure 4.5 for an image of the flat beam taken after skew quadrupole settings were chosen by a neural network, along with example quadrupole scan fits for $\langle xx \rangle$, $\langle xy \rangle$, and $\langle yy \rangle$ for an actual quadrupole scan. The normal quadrupoles that follow the skew quadrupole triplet are used for this multi-shot, two-quadrupole scan. Another emittance measurement technique is the mask-based emittance reconstruction: that is, to insert grids or pepperpots [122] into the beamline and perform a single-shot measurement downstream [93]. Generally, the single-shot, grid-based or pepperpot method is not well suited for a flat beam, as TEM grids or pepperpots are symmetric in the two dimensions and a flat beam by definition has a large emittance ratio;



(a) Image of flat beam (b) Example $\langle xx \rangle$ fit (c) Example $\langle yy \rangle$ fit (d) Example $\langle xy \rangle$ fit

Figure 4.5: a) Image of the resulting flat beam, and measured normalized projected emittances after NN optimization. b) Example quadrupole scan fit for $\langle xx \rangle$. c) Example quadrupole scan fit for $\langle yy \rangle$. d) Example quadrupole scan fit for $\langle xy \rangle$

the angles in one spatial dimension are much larger than the other. Generally, the flat beam can be measured in one dimension, but not the other with grids or pepperpots. For example, see Fig. 4.6 for an image of a flat beam as measured by the grids. In this case, the grid dimensions and beam transport are set such that the small emittance can be measured, but not the large dimension.

Thus, direct measurement of the emittance ratio of the flat beam is infeasible as an objective function for optimization. As such, it was necessary to find a suitable alternative. Using GPT [56] simulations as a guide, it was found that the optimal flat beam resulted from adjusting the skew quadrupoles to bring the beam to a tight focus a few cm downstream of the triplet. The beam is then divergent by the time it reaches the downstream screens. As such, a hallmark of a beam with a large emittance ratio is the large spot size in one transverse dimension (x or y) and small in the other. Thus, the chosen objective function is:

$$f(\sigma_1, \sigma_2) = \frac{\sigma_1}{\sigma_2} \tag{4.10}$$

where σ_1 is the user's choice of σ_x and σ_y, the one that is to be minimized, and σ_2 remaining of σ_x and σ_y. $f(\sigma_1, \sigma_2)$ is the objective function and is to be minimized.

The suitability of the objective function was confirmed in simulation. The contours of the objective function and the minimum emittance following the FBT is shown in Figure

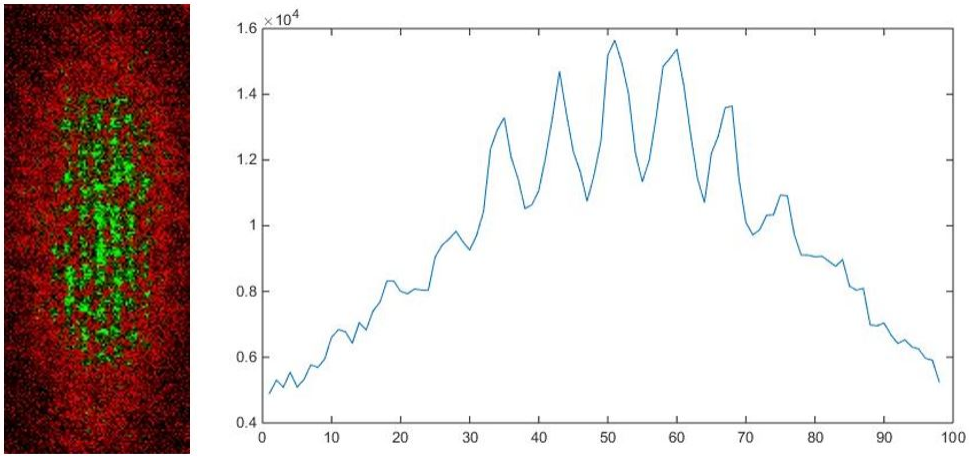
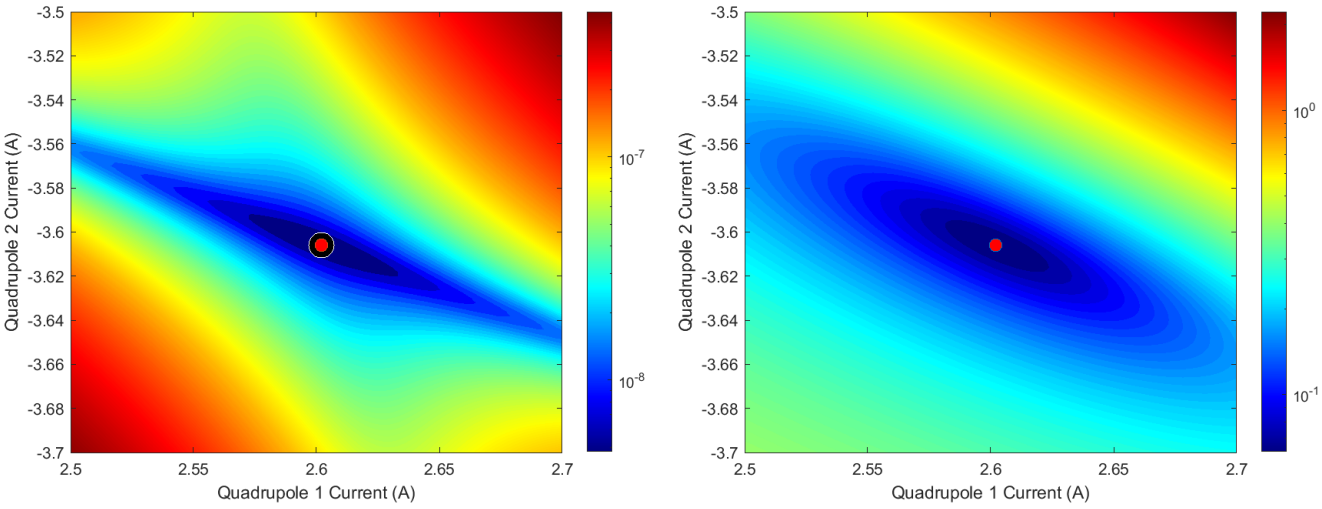


Figure 4.6: Left: Flat beam passing through TEM grids. Right: Projection of the flat beam image along the x-axis.



(a) Normalized Projected Emittance

(b) Objective Function

Figure 4.7: Comparison of the objective function and the minimum projected emittance as a function of quadrupole current in the first two quadrupoles in the skew quadrupole triplet. The currents in the solenoid and the third quadrupole are held to their optimal values. The minimum of the objective function is denoted in red, and the minimum of the small projected emittance is denoted in black in (a). These points are the same.

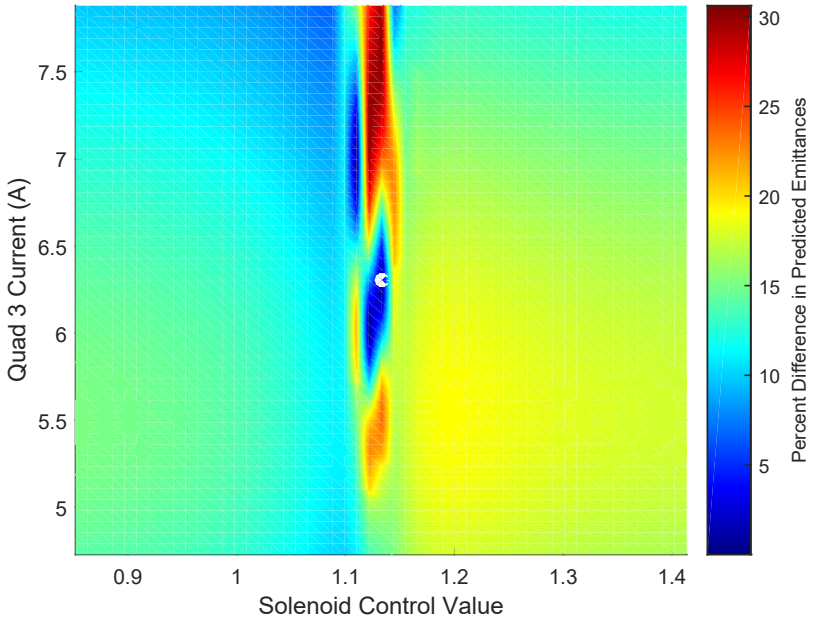


Figure 4.8: Comparison of the predicted minimum projected emittance by the objective function and the minimum projected emittance (in simulation). The percentage by which the predicted minimum emittance is greater than the true minimum is shown as a function of current in quadrupole 3 and solenoid setting. The white dot indicates the optimal settings for the third quadrupole and the solenoid.

4.7 as a function of the first two quadrupoles. To make this figure, the third quadrupole in the triplet is held constant at its optimal value, as it had the smallest impact on the final emittance, per GPT [56] simulation. The solenoid is also held at its optimal value, and the maximum magnetic field is placed on the photocathode.

The validity of the objective function was also tested in simulation for non-optimal values of the solenoid and the third quadrupole. For each combination of solenoid setting and quadrupole 3 current, the other two quadrupoles are optimized based on 1) the spot size ratio (experimental objective function) and 2) the flat beam condition. In Fig. 4.8, the percent difference between the projected emittance using the two objective functions are shown. Up to a 25% variation from their optimal values in simulation for both the solenoid and final skew quadrupole were explored, which more than covers the spread of values used

in experiment. As can be seen, within this range in parameter space, using the chosen, spot-size-based objective function consistently picks a point that is within 30 % of the minimum in terms of its projected emittance.

4.2.4 Online Modeling & Distribution Shift

In order to build a fast, online, ML-based surrogate model of the system, approximately 57000 experimental measurements were taken of the transverse spatial second order moments of the beam distribution function (σ_x, σ_y, σ_{xy}), as a function of upstream accelerator parameters. These parameters included the phase and amplitude of the RF in the gun and the electron optics downstream (dipole kicker magnets, emittance compensation solenoid, skew quadrupole triplet). A feed-forward neural network with 7 hidden layers each with a tanh activation function was trained to predict the final beam properties based on the upstream properties. These data points were taken over the course of three days, when coarse parameter scans were taken and the RF naturally varied each day.

Once the model was trained, multi-objective genetic optimization was used to optimize the model and provide Pareto-optimal solutions to maximize σ_y and minimize σ_x, thus finding solutions to Eq. 4.10, but also offering the ability to view the trade-off between the objectives. For this optimization, the other inputs were fixed to the experimental parameters. This last point deserves a bit more emphasis, as the determination of the "proper" experimental settings is not trivial in a fluctuating system. In general, fluctuations are another justification for using the "warm start" procedure as discussed below. Predictions are not expected to be perfect, as the optimal solutions are fluctuating with the system fluctuations. Fig. 4.9 shows the fluctuations during train and test sets.

The first test of the model was performed the following day after training. While there were fluctuations in both the training and the test sets, note that the distribution drifted outside the span of the training set during the test sets. This problem of "distribution shift" led the model to extrapolate slightly. See Fig. 4.9 for more details. As such, predictions were not expected to be optimal, but considering that the distribution shift was small, the optimal

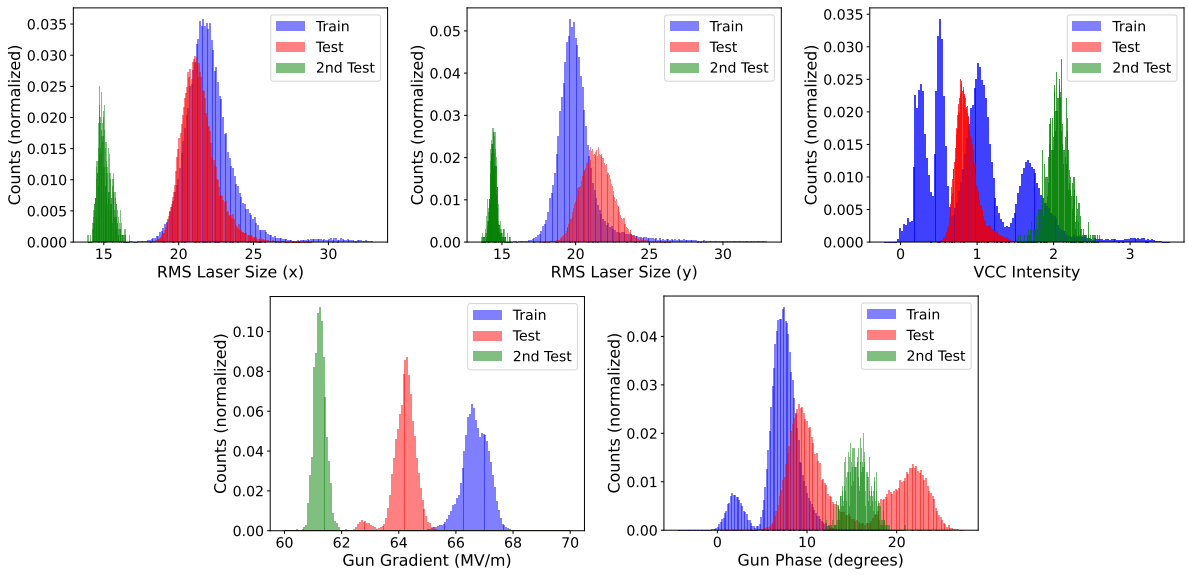


Figure 4.9: Recorded variation for some of the input parameters between training data set (over several days) and testing. The first test was done a day after recording the initial training data, and the second test was conducted about 5 months later. The pixel intensity sum and laser spot sizes are notable because these are correlated with changes in the charge of the beam.

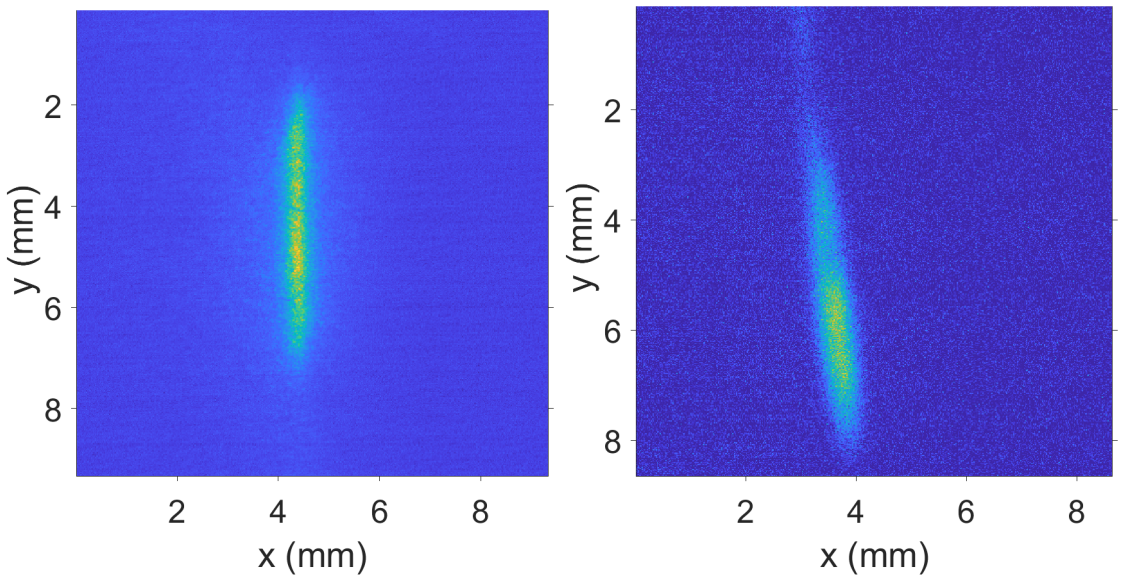


Figure 4.10: Left: Example beam produced from optimization over NN model one day after the last set of training data was obtained. Right: example beam produced 5 months later using the same procedure, under the distribution shift shown in Fig. 4.9.

parameters should be nearby. In particular, after more than five months of parameter drift, and with different slightly different input parameters (see Fig. 4.9), the model gave even worse predictions, despite still looking like a poorly optimized flat beam. See Fig. 4.10 for an example, where the beam appears to be a "rotated" flat beam on the screen. Considering that the surrogate model was predicting outside its training set, this is a relatively interesting result, as it suggests that the model could be used to give solutions nearby to the global optimum.

Assuming that the global optimum is the closest local minimum in the space, one can use local optimization to find the global optimum. As such, local, model independent techniques were used for such a "warm start" approach. Extremum seeking (ES) [123, 124, 125] is a powerful, model-independent optimization routine that can be applied to optimize quickly and control particle accelerators. It has myriad uses for particle accelerators, including optimizing an electron beam via automatic tuning of accelerator parameters [126] to tuning the latent space of a convolutional neural network-based surrogate model of the accelerator

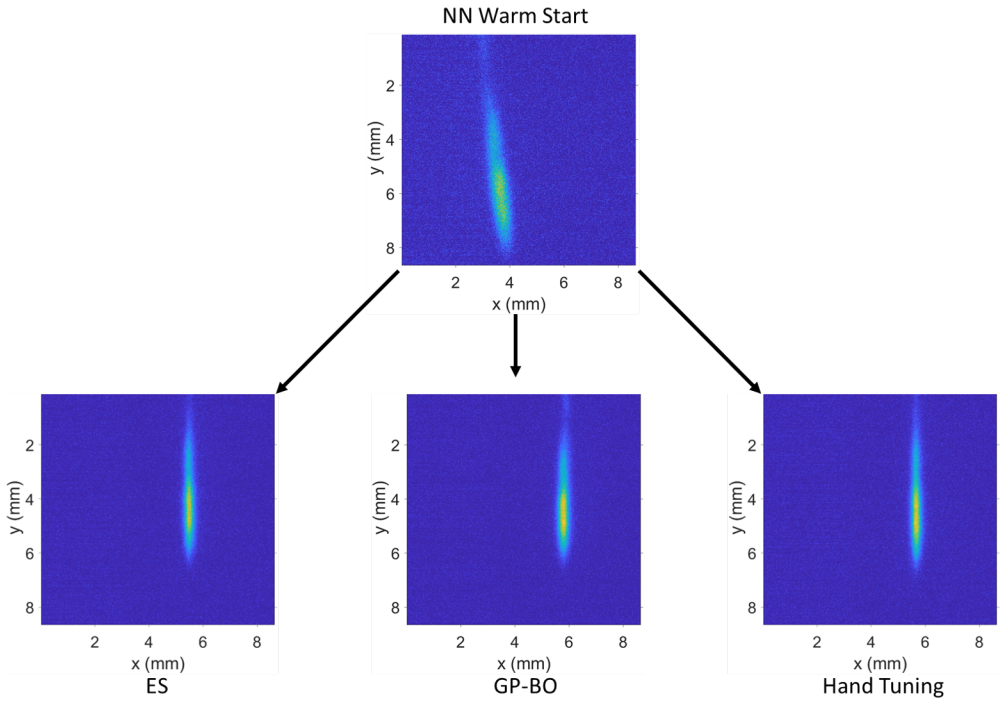


Figure 4.11: Starting from a "warm start" using a NN model trained 5 months earlier, ES solutions are compared to Gaussian Process-Bayesian Optimization (GP-BO) and hand tuned solutions.

to make it more robust to drifting parameters [59]. This work seeks to use ES to build a tool for adaptive, on-line control of accelerators.

In this case, ES was able to give an equivalent solution to hand tuning in an automated way. While the hand tuning approach with a well-trained operator was faster than ES, the ES approach is more repeatable, as it is fully automated. See Fig. 4.11 for an example of the "warm start" approach.

Note that — although not the focus of this chapter — Gaussian Process-Bayesian Optimization (GP-BO) also shows similar results. GP-BO is a tool that shows much promise for accelerators, as the sample efficient method builds a model of the system while optimizing. However, to keep the second part of the "warm start" prescription fully generalizable to systems that cannot be modeled beyond the initial, coarse NN model, due to fast, unmeasured jitter, model-independent methods such as ES are more the focus of this chapter.

4.2.5 Flat Beam Measurement and Discussion

A simple verification for the flat beam adapter was qualitatively carried out using the method of the three screens. Quantitative measurements were hindered by the large point spread functions of the screens used for this measurement. The scaling with the initial angular momentum was verified changing the laser spot size on the cathode (see Fig. (4.12)). As can be seen from Eqs. (4.6) and (4.7), the large emittance increases as the square of the initial laser spot size on the photocathode. Consequently, for constant beta functions at the screen, the large electron spot size varies approximately linearly with the laser spot size, since:

$$\sigma_{lg,screen} = \sqrt{\epsilon_+ \beta} \tag{4.11}$$

while the small electron beam spot size should stay constant as a function of the initial laser spot. The emittance of the beam was measured using two techniques: the two-quadrupole

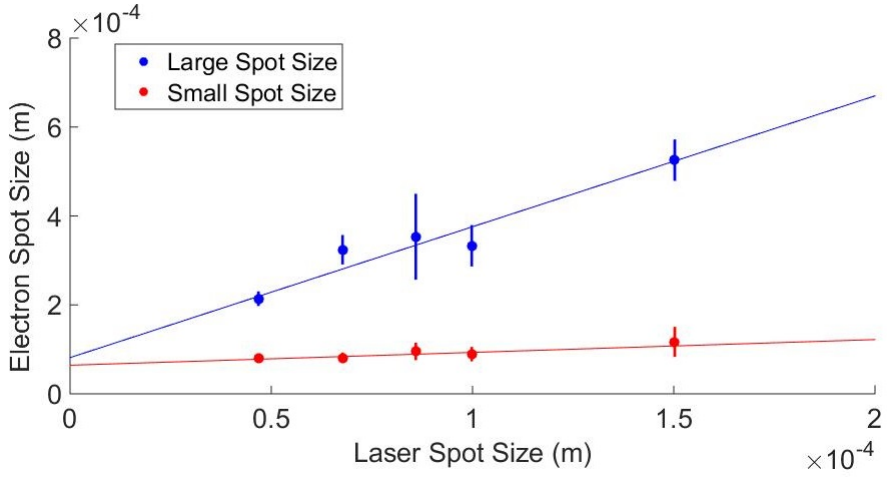


Figure 4.12: Flat beam rms spot sizes as a function of laser rms radius on the photocathode

scan [92] and the TEM grid method [93]. The two-quadrupole scan is based on varying the normal quadrupoles after the flat beam adapter and measuring the corresponding second order moments of the beam distribution at the detector using the procedure discussed in the previous chapter, in particular Eq. 3.8.

After a sufficient number of measurements, it is possible to fit all ten unique beam matrix

elements for the flat beam. The two-quadrupole scan technique is a multishot measurement which suffers from machine fluctuations and has a limited resolution on the small emittance value. The large emittance was found in very good agreement with the predictions (see Fig. (4.13)). In order to resolve the small emittance, the TEM grid [93] (see Fig. (4.6)) recon-

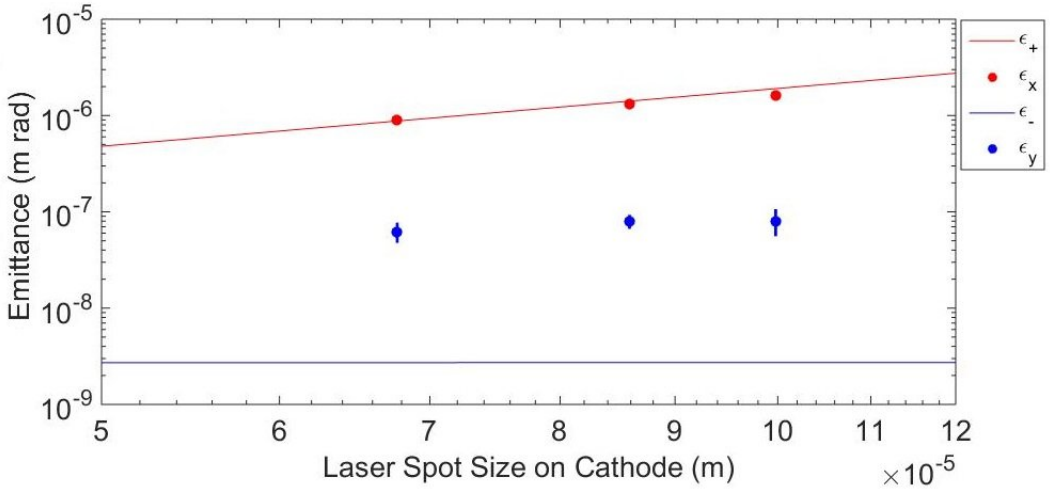


Figure 4.13: Flat beam emittances as measured by a quadrupole scan as a function of RMS laser spot size on the photocathode

struction technique was employed. As discussed above, the However, a 2-D reconstruction can be used for the small projected emittance.

The emittance measured using TEM is still larger than theory predicts by an order of magnitude. There are numerous reasons why the measured emittance would be larger than predicted in linear theory, or by simulations:

1. An erroneous quadrupole moment in the gun and/or solenoid [127], in which case, simulations suggest that a quadrupole coupling corrector [128, 129] could be used to retrieve the smallest eigenemittance as the normalized projected emittance.

2. Other nonlinear effects, such as chromatic aberrations in the quadrupoles, that are not incorporated in the linear theory of the flat beam transport, and not included in simulation.

3. A systematic error in the emittance measurements

4. Charge and RF jitter make it impossible to reach the optimum, as the optimal settings keep changing

While not an exhaustive list, any combination of the above factors could explain the larger emittance in measurement than in simulation or in analytical theory.

4.3 Model Independent Tuning at HiRES

The utility of a model-independent tuning approach was further demonstrated at HiRES. In order to optimize the real- and inverse-space resolution of the UED measurement, a tradeoff must be found between focusing the beam at the sample plane (EX in Fig. 2.1) and at the detector plane (DD). The beam is focused by the three magnets in the second quadrupole triplet in the UED line (Q2). The first and third quadrupoles in Q1 are constrained to minimize the dispersion and its derivative after the dogleg, leaving the second quadrupole in Q1 as a free parameter to tune the vertical beam size (due to symmetry, the beam must be approximately at a focus in the horizontal dimension at the second quadrupole plane, so this quadrupole has little effect in the horizontal dimension).

In a quadrupole tuning problem — similar to the FBT optimization — toward automation, ES was applied toward a solution to the above problem in a system with significant fluctuations. To demonstrate the utility to a problem such as the above, where there is a tradeoff between two parameters and therefore no overall best solution in all cases, ES was tested using a changing cost function.

For example, often, when operating an accelerator, it is necessary to change modes of operation. At HiRES, ES was demonstrated to be able to follow a moving cost function. In this example, data was taken on VS2 (as shown in Fig. 2.1). The quadrupoles in Q1 were the parameters on which ES was optimizing. To find a starting point, a coarse grid quadrupole scan was performed and interpolated using a four-layer feed-forward neural network. Inputs were the three quadrupole settings. Outputs were the root-mean-square transverse beam

sizes. Minimization of the NN of the following cost function was performed:

$$f(q_1, q_2, q_3) = |x_{rms}| + |y_{rms}| + |x_{rms} - y_{rms}| \tag{4.12}$$

to find the optimal quadrupole settings to make a small, round beam. Using the resultant point as a starting point, a sinusoidally varying cost function was introduced, as shown in Fig. 4.14. As can be seen, ES keeps the cost function low by varying the beam size as necessary. This shows the "warm start" approach in a case where the user-desired changes can be made automatically.

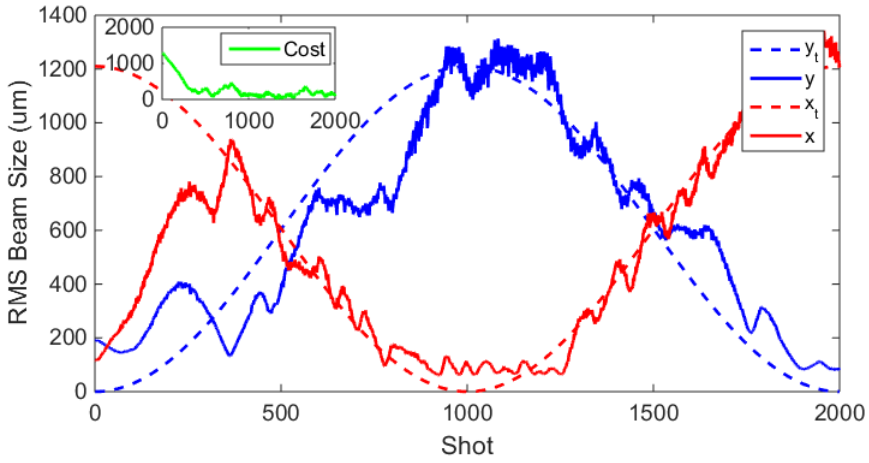


Figure 4.14: ES minimizing a variable cost function: RMS beam sizes and corresponding targets. Inset: cost function minimization

Another situation that often requires operator expertise is the compensation of large parameter drifts. In fact, severe enough drifts for a static, ML-based system will require operator intervention, if the system drifts outside of the range of the training set. As a trial, at HiRES, such a severe drift was induced by moving upstream parameters and was compensated by means of the ES algorithm. The data for this experiment were acquired using a PI-MAX 4 intensified camera at approximately 2 Hz on DD (see Fig. 2.1). The second quadrupole of the first quadrupole triplet in the dogleg (Q1 in Fig. 2.1) was varied in a sinusoidal pattern. As can be seen from Fig. 4.15, ES was able to keep the cost function (Eq. 4.12) minimized by only varying the three quadrupoles in the second triplet. The

nominal case of keeping the three quadrupoles constant is also shown. It is worth noting that beam sizes in the both dimensions remained small, so long as the rate of change of the drift was sufficiently small.

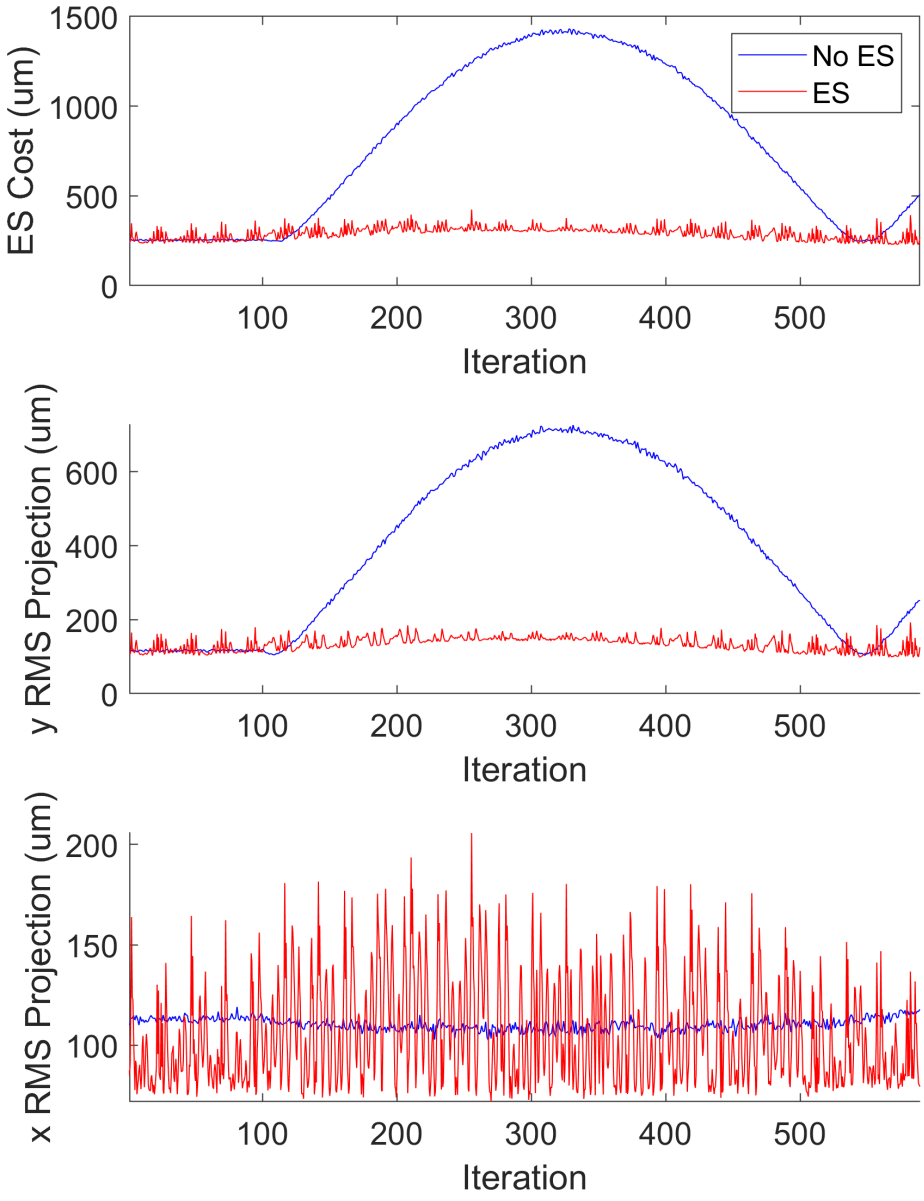


Figure 4.15: ES minimizing a static cost function (Eq. 4.12) in variable, drifting conditions. Top: Cost function with and without ES feedback. Middle: RMS beam size in y-dimension with and without ES. Bottom: RMS beam size in x-dimension with and without ES. Note that the feedback does not sacrifice beam size in x to make y smaller.

ES is shown to be a powerful tool for automatic, model-independent optimization of a changing system, in two cases: 1) if the system needs to change, and 2) if the system changes, but the objective stays the same. As such, this will be a powerful tool for automatic control after a "warm start" approach in the presence of system fluctuations.

Preliminary tests of four-quad, two-objective optimization using ES are underway, using a virtual diagnostic for the screen at the sample plane. In the future, this could be a powerful tool for automated adaptive control of beam parameters for UED, particularly in the face of drift and jitter. Currently, drift and jitter are ignored for hours at a time for UED measurements, but adaptive control could lead to more consistent beams throughout a UED measurement.

4.4 Conclusion

Accelerator models are a powerful tool for control. With a high-fidelity model, in an ideal environment, choosing the proper accelerator settings becomes a simple optimization or inversion problem. This chapter, however, details some of the work when the model is imperfect — in particular, when there are fluctuations that are not captured in the model. In this case, assuming that the fluctuations are small enough, the model can predict a solution nearby in parameter space to the true optimum and other model-independent methods can be used to compensate, in a fully-automated procedure.

The overall procedure is not that different from the current state-of-the-art, where in general, when a user of a particle accelerator requests a beam of a certain set of parameters, an operator will consult some model, whether a physics model or even a mental model, and select parameters that are nearby to the optimum. From there, the operator must rely on intuition and hand tuning to select the optimum. With a skilled operator, this approach has good results, but can lack reliability, repeatability, speed, and requires an operator of sufficient skill. Generally, this can also take dedicated time during an experiment, while the fully-automated methods discussed in this chapter can be used passively during an

experiment.

In this chapter, two examples of the "warm start" followed by automated tuning were shown: 1) tuning of the FBT at Pegasus and 2) tuning of the UED setup at HiRES. Both performed well in an environment laden with fluctuations. At Pegasus, the FBT was tuned automatically in an environment with relatively large fluctuations. At HiRES, upstream parameter drift was simulated and automated tuning was shown to follow. In the future, this approach has the potential to lead to far more reliable and less invasive tuning than current operator-based hand tuning.

CHAPTER 5

Conclusion

5.1 Model Typology and Convergence

This thesis details the use of high-performance (fast and accurate) accelerator models to improve controls and diagnostics for high brightness beams. In general, the typology of such accelerator models is quite varied. One way to characterize models is to divide them into online and offline models, depending on whether they are suitable for use during an experiment. Online and offline models are complimentary and are generally used for different purposes. Example uses for online models appeared in Chapter 2 and Chapter 4, examples of the use of offline models were presented in Chapter 3.

Meanwhile, another distinction is physics-based models vs. data-driven models. Physics-based models are generally better suited to be offline models due to their execution time, but data-driven models are used in both. Data-driven models are discussed in all chapters, while physics-based models are discussed in Chapter 3 and Chapter 4. ML surrogate models occupy an interesting space, as they are data-driven models based on physics-based models and bridge the gap between the capabilities of a physics-based model and a ML model, at least within a predetermined range within parameter space. This was discussed in detail in Chapter 3.

In the future, this gap between physics-based and data-driven models may become even narrower. Efforts to inject more physics into ML models are underway, for example by using differentiable simulations [130], or by injecting information into the latent space of an autoencoder [60]. Additionally, physics-informed neural networks and physics-constrained neural networks inject physics directly into the neural network structure [131, 132, 133].

Further, to bridge the gap between offline and online ML models, transfer learning has been employed on surrogate models trained on simulation data in order to match more carefully the real outcomes [134].

5.2 Summary

In Chapter 2, a ML-based virtual diagnostic was trained using experimental data to be an online prediction tool for the electron beam time of arrival. The models shown ranged from the age-old linear regression model to the state-of-the-art ML forecasting model: the temporal fusion transformer. These models showed the ability to reduce the uncertainty in beam TOA due to drift to the level of the shot-to-shot fluctuations, showing potential to greatly enhance the resolution of long UED measurements.

In Chapter 3, offline, physics-based models were matched to experimental data to produce higher fidelity models. This was accomplished by using Markov chain Monte Carlo analysis at HiRES, and as a model-based reconstruction technique, has the potential to have even lower uncertainty in the matched parameters by injecting information about the fluctuations into the technique. This was demonstrated in a simulation study at Pegasus. Further, coarse matching of a model of FAST was discussed. This chapter sought to show a bit of "how the sausage is made" with regard to making high-fidelity models.

In Chapter 4, a common accelerator problem — large fluctuations — while working with an online model was discussed at HiRES and Pegasus. At Pegasus, the fluctuations made it challenging to optimize the flat beam transform, while at HiRES, transverse and — as discussed in Chapter 2 — longitudinal fluctuations can affect UED measurements. In particular, when optimizing an online model, a "warm start" solution was proposed. Methods for finding the optimum using model-independent methods were also discussed at HiRES and Pegasus.

An astute reader of this thesis will note the broad applicability of these techniques. Indeed, work was shown at HiRES (LBNL), Pegasus (UCLA) and FAST (Fermilab) on

three state-of-the-art photoinjectors. The points contained herein are even more general than these examples might suggest. While some of the applications listed in this thesis were specific, such as UED or the flat beam transform, the concepts are general. A virtual diagnostic was shown to work for UED, but time & energy stamping are general concepts in accelerator physics — even static emittance measurements require knowledge of the beam energy (e.g. [92, 93, 91]). The applicability of virtual diagnostics has been shown numerous times, as discussed in Chapter 2. Any beamline can make use of a physics-based model that matches real behavior and the MCMC and coarse matching methods can be applied anywhere. Finally, the "warm start" solution was shown to work at two very different beamlines for two different tasks, but is not limited only to these two. Applying these techniques was shown to help the problems shown in this thesis, but has greater applicability to numerous other problems.

In fact, for an ideal beamline, the work in each of these chapters could be applied to compliment each other. In general, the use of online, ML-based virtual diagnostics compliments having physics-based models to inform control policy. Online, ML models based on experimental data and surrogate models based on the physics-based simulation can be used for control and optimization. The surrogate model would have more range, while the experimentally-driven model likely would have higher fidelity. Finally, working with these models in a fluctuation-laden environment is a harsh reality of accelerator physics, and the work shown in the final chapter could be of use. In the future, greater application of these techniques has potential to improve even further controls and diagnostics for high brightness beams in experiment.

Bibliography

[1] D. Filippetto and H. Qian, "Design of a high-flux instrument for ultrafast electron diffraction and microscopy," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 49, no. 10, p. 104003, 2016.

[2] J. J. Thomson, "Cathode rays," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 44, no. 269, pp. 293–316, 1897.

[3] E. O. Lawrence, "Method and apparatus for the acceleration of ions," Feb. 20 1934, uS Patent 1,948,384.

[4] E. O. Lawrence and M. S. Livingston, "The production of high speed light ions without the use of high voltages," *Physical Review*, vol. 40, no. 1, p. 19, 1932.

[5] E. McMillan and P. H. Abelson, "Radioactive element 93," *Phys. Rev.*, vol. 57, pp. 1185–1186, Jun 1940. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.57.1185.2

[6] ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037026931200857X

[7] J. Fraser, R. Sheffield, and E. Gray, "High-brightness photoemitter development for electron accelerator injectors," in *AIP Conference Proceedings*, vol. 130. American Institute of Physics, 1985, pp. 598–601.

[8] R. L. Sheffield, "High brightness electron sources," in *Proceedings Particle Accelerator Conference*, vol. 2. IEEE, 1995, pp. 882–886.

[9] J. Fraser, R. Sheffield, and E. Gray, "A new high-brightness electron injector for free electron lasers driven by rf linacs," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 250, no. 1-2, pp. 71–76, 1986.

[10] P. Musumeci, J. G. Navarro, J. Rosenzweig, L. Cultrera, I. Bazarov, J. Maxson, S. Karkare, and H. Padmore, "Advances in bright electron sources," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 907, pp. 209–220, 2018.

[11] 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 ångstrom-wavelength free-electron laser," *nature photonics*, vol. 4, no. 9, pp. 641–647, 2010.

[12] C. Pellegrini, "The history of x-ray free-electron lasers," *The European Physical Journal H*, vol. 37, no. 5, pp. 659–708, 2012.

[13] X. Wang, D. Xiang, T. Kim, and H. Ihee, "Potential of femtosecond electron diffraction using near-relativistic electrons from a photocathode rf electron gun," *Journal-Korean Physical Society*, vol. 48, no. 3, p. 390, 2006.

[14] J. B. Hastings, F. M. Rudakov, D. H. Dowell, J. F. Schmerge, J. D. Cardoza, J. M. Castro, S. M. Gierman, H. Loos, and P. M. Weber, "Ultrafast time-resolved electron diffraction with megavolt electron beams," *Applied Physics Letters*, vol. 89, no. 18, p. 184109, 11 2006. [Online]. Available: https://doi.org/10.1063/1.2372697

[15] P. Musumeci, J. Moody, C. Scoby, M. Gutierrez, and M. Westfall, "Laser-induced melting of a single crystal gold sample by time-resolved ultrafast relativistic electron diffraction," *Applied Physics Letters*, vol. 97, no. 6, 2010.

[16] Y. Murooka, N. Naruse, S. Sakakihara, M. Ishimaru, J. Yang, and K. Tanimura, "Transmission-electron diffraction by mev electron pulses," *Applied Physics Letters*, vol. 98, no. 25, 2011.

[17] R. Li, C. Tang, Y. Du, W. Huang, Q. Du, J. Shi, L. Yan, and X. Wang, "Experimental demonstration of high quality mev ultrafast electron diffraction," *Review of Scientific Instruments*, vol. 80, no. 8, 2009.

[18] S. Weathersby, G. Brown, M. Centurion, T. Chase, R. Coffee, J. Corbett, J. Eichner, J. Frisch, A. Fry, M. Gühr *et al.*, "Mega-electron-volt ultrafast electron diffraction at slac national accelerator laboratory," *Review of Scientific Instruments*, vol. 86, no. 7, 2015.

[19] R. Bonifacio, C. Pellegrini, and L. Narducci, "Collective instabilities and high-gain regime in a free electron laser," *Optics Communications*, vol. 50, no. 6, pp. 373–378, 1984. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0030401884901056

[20] Z. Huang and K. Kim, "Review of x-ray free-electron laser theory," *Phys. Rev. ST Accel. Beams*, vol. 10, p. 034801, Mar 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.10.034801

[21] I. Bazarov, L. Cultrera, A. Bartnik, B. Dunham, S. Karkare, Y. Li, X. Liu, J. Maxson, and W. Roussel, "Thermal emittance measurements of a cesium potassium antimonide photocathode," *Applied Physics Letters*, vol. 98, no. 22, 2011.

[22] I. V. Bazarov, B. M. Dunham, and C. K. Sinclair, "Maximum achievable beam brightness from photoinjectors," *Phys. Rev. Lett.*, vol. 102, p. 104801, Mar 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.102.104801

[23] J. R. Rees, "Symplecticity in beam dynamics: An introduction," Citeseer, Tech. Rep., 2003.

[24] J. D. Bjorken and S. K. Mtingwa, "Intrabeam scattering," *Part. Accel.*, vol. 13, no. FERMILAB-PUB-82-47-THY, pp. 115–143, 1982.

[25] A. Piwinski, "Intra-beam scattering," in *Proceedings of the 9th International Conference on High Energy Accelerators*. Stanford, CA, USA: American Institute of Physics, 1974, pp. 405–409.

[26] T. Nakazato, M. Oyamada, N. Niimura, S. Urasawa, O. Konno, A. Kagaya, R. Kato, T. Kamiyama, Y. Torizuka, T. Nanba *et al.*, "Observation of coherent synchrotron radiation," *Physical review letters*, vol. 63, no. 12, p. 1245, 1989.

[27] S. Heifets, G. Stupakov, and S. Krinsky, "Coherent synchrotron radiation instability in a bunch compressor," *Physical Review Special Topics-Accelerators and Beams*, vol. 5, no. 6, p. 064401, 2002.

[28] M. Reiser, *Theory and design of charged particle beams*. John Wiley & Sons, 2008.

[29] K. Kim, "Rf and space-charge effects in laser-driven rf electron guns," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 275, no. 2, pp. 201–218, 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0168900289906888

[30] L. Serafini and J. B. Rosenzweig, "Envelope analysis of intense relativistic quasilaminar beams in rf photoinjectors:ma theory of emittance compensation," *Phys. Rev. E*, vol. 55, pp. 7565–7590, Jun 1997. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.55.7565

[31] Y. Sun, "Angular-momentum-dominated electron beams and flat-beam generation," Ph.D. dissertation, 2005.

[32] R. Brinkmann, Y. Derbenev, and K. Flöttmann, "A low emittance, flat-beam electron source for linear colliders," *Physical Review Special Topics-Accelerators and Beams*, vol. 4, no. 5, p. 053501, 2001.

[33] K. Kim, "Round-to-flat transformation of angular-momentum-dominated beams," *Physical Review Special Topics-Accelerators and Beams*, vol. 6, no. 10, p. 104002, 2003.

[34] Y. Derbenev, "Adapting optics for high energy electron cooling," *University of Michigan, UM-HE-98-04*, 1998.

[35] A. Burov, S. Nagaitsev, A. Shemyakin, and Y. Derbenev, "Optical principles of beam transport for relativistic electron cooling," *Physical Review Special Topics-Accelerators and Beams*, vol. 3, no. 9, p. 094002, 2000.

[36] A. Burov, S. Nagaitsev, and Y. Derbenev, "Circular modes, beam adapters, and their applications in beam optics," *Physical Review E*, vol. 66, no. 1, p. 016503, 2002.

[37] P. Piot, Y.-E. Sun, and K.-J. Kim, "Photoinjector generation of a flat electron beam with transverse emittance ratio of 100," *Physical Review Special Topics-Accelerators and Beams*, vol. 9, no. 3, p. 031001, 2006.

[38] Y.-E. Sun, P. Piot, K.-J. Kim, N. Barov, S. Lidia, J. Santucci, R. Tikhoplav, and J. Wennerberg, "Generation of angular-momentum-dominated electron beams from a photoinjector," *Physical Review Special Topics-Accelerators and Beams*, vol. 7, no. 12, p. 123501, 2004.

[39] B. E. Carlsten and K. A. Bishofberger, "Simple algorithm for designing skew-quadrupole cooling configurations," *New Journal of Physics*, vol. 8, no. 11, p. 286, 2006.

[40] W. K. H. Panofsky and W. A. Wenzel, "Some Considerations Concerning the Transverse Deflection of Charged Particles in Radio-Frequency Fields," *Review of Scientific Instruments*, vol. 27, no. 11, pp. 967–967, 12 2004. [Online]. Available: https://doi.org/10.1063/1.1715427

[41] J. T. Moody, P. Musumeci, M. S. Gutierrez, J. B. Rosenzweig, and C. M. Scoby, "Longitudinal phase space characterization of the blow-out regime of rf photoinjector operation," *Phys. Rev. ST Accel. Beams*, vol. 12, p. 070704, Jul 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.12.070704

[42] K. Floettmann and V. V. Paramonov, "Beam dynamics in transverse deflecting rf structures," *Phys. Rev. ST Accel. Beams*, vol. 17, p. 024001, Feb 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.17.024001

[43] A. H. Lumpkin and K. P. Wootton, "High-resolution longitudinal profile diagnostics for ultralow charges stored in a ring," *Phys. Rev. Accel. Beams*, vol. 24, p. 072806, Jul 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.24.072806

[44] A. Scheinker, S. Gessner, C. Emma, and A. L. Edelen, "Adaptive model tuning studies for non-invasive diagnostics and feedback control of plasma wakefield acceleration at facet-ii," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 967, p. 163902, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016890022030396X

[45] 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*, vol. 21, no. 11, p. 112802, 2018.

[46] O. Convery, L. Smith, Y. Gal, and A. Hanuka, "Uncertainty quantification for virtual diagnostic of particle accelerators," *Physical Review Accelerators and Beams*, vol. 24, no. 7, p. 074602, 2021.

[47] V. Yakimenko, L. Alsberg, E. Bong, G. Bouchard, C. Clarke, C. Emma, S. Green, C. Hast, M. Hogan, J. Seabury *et al.*, "Facet-ii facility for advanced accelerator experimental tests," *Physical Review Accelerators and Beams*, vol. 22, no. 10, p. 101301, 2019.

[48] A. Hanuka, C. Emma, T. Maxwell, A. S. Fisher, B. Jacobson, M. J. Hogan, and Z. Huang, "Accurate and confident prediction of electron beam longitudinal properties using spectral virtual diagnostics," *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.

[49] A. Scheinker and S. Gessner, "Adaptive method for electron bunch profile prediction," *Physical Review Special Topics-Accelerators and Beams*, vol. 18, no. 10, p. 102801, 2015.

[50] R. J. England, *Longitudinal shaping of relativistic bunches of electrons generated by an RF photoinjector*. University of California, Los Angeles, 2007.

[51] P. Denham and P. Musumeci, "Space-charge aberrations in single-shot time-resolved transmission electron microscopy," *Phys. Rev. Appl.*, vol. 15, p. 024050, Feb 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.15.024050

[52] M. Borland, "Elegant: A flexible sdds-compliant code for accelerator simulation," Argonne National Lab., IL (US), Tech. Rep., 2000.

[53] A. Adelmann, P. Calvo, M. Frey, A. Gsell, U. Locans, C. Metzger-Kraus, N. Neveu, C. Rogers, S. Russell, S. Sheehy *et al.*, "Opal a versatile tool for charged particle accelerator simulations," *arXiv preprint arXiv:1905.06654*, 2019.

[54] H. Grote and F. Schmidt, "Mad-x-an upgrade from mad8," in *Proceedings of the 2003 Particle Accelerator Conference*, vol. 5. IEEE, 2003, pp. 3497–3499.

[55] K. Flottmann, S. Lidia, and P. Piot, "Recent improvements to the astra particle tracking code," Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), Tech. Rep., 2003.

[56] "General particle tracer," *http://www.pulsar.nl/gpt/*.

[57] R. Huang, C. Mitchell, C. Papadopoulos, H. Qian, M. Venturini, J. Qiang, D. Filippetto, J. Staples, Q. Jia, and F. Sannibale, "Off-axis beam dynamics in rf-gun-based electron photoinjectors," *Phys. Rev. Accel. Beams*, vol. 19, p. 113401, Nov 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.19.113401

[58] S. Li, S. Alverson, D. Bohler, A. Egger, A. Fry, S. Gilevich, Z. Huang, A. Miahnahri, D. Ratner, J. Robinson, and F. Zhou, "Ultraviolet laser transverse profile shaping for improving x-ray free electron laser performance," *Phys. Rev. Accel. Beams*, vol. 20, p. 080704, Aug 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.20.080704

[59] A. Scheinker, "Adaptive machine learning for time-varying systems: low dimensional latent space tuning," *Journal of Instrumentation*, vol. 16, no. 10, p. P10008, 2021.

[60] 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*, vol. 107, p. 045302, Apr 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.107.045302

[61] D. Filippetto, P. Musumeci, R. Li, B. J. Siwick, M. Otto, M. Centurion, and J. Nunes, "Ultrafast electron diffraction: Visualizing dynamic states of matter," *Reviews of Modern Physics*, vol. 94, no. 4, p. 045004, 2022.

[62] K. Siddiqui, D. Durham, F. Cropp, F. Ji, S. Paiagua, C. Ophus, N. Andresen, L. Jin, J. Wu, S. Wang *et al.*, "Relativistic ultrafast electron diffraction at high repetition rates," *arXiv preprint arXiv:2306.04900*, 2023.

[63] P. Musumeci, J. Moody, and C. Scoby, "Relativistic electron diffraction at the ucla pegasus photoinjector laboratory," *Ultramicroscopy*, vol. 108, no. 11, pp. 1450–1453, 2008.

[64] B. J. Siwick, J. R. Dwyer, R. E. Jordan, and R. D. Miller, "An atomic-level view of melting using femtosecond electron diffraction," *Science*, vol. 302, no. 5649, pp. 1382–1385, 2003.

[65] G. Mourou and S. Williamson, "Picosecond electron diffraction," *Applied Physics Letters*, vol. 41, no. 1, pp. 44–45, 1982.

[66] A. Ischenko, V. Golubkov, V. Spiridonov, A. Zgurskii, A. Akhmanov, M. Vabischevich, and V. Bagratashvili, "A stroboscopical gas-electron diffraction method for the investigation of short-lived molecular species," *Applied Physics B*, vol. 32, pp. 161–163, 1983.

[67] O. Bostanjoglo, R. Tornow, and W. Tornow, "Nanosecond transmission electron microscopy and diffraction," *Journal of Physics E: Scientific Instruments*, vol. 20, no. 5, p. 556, 1987.

[68] F. Sannibale, D. Filippetto, C. Papadopoulos, J. Staples, R. Wells, B. Bailey, K. Baptiste, J. Corlett, C. Cork, S. De Santis *et al.*, "Advanced photoinjector experiment photogun commissioning results," *Physical Review Special Topics-Accelerators and Beams*, vol. 15, no. 10, p. 103501, 2012.

[69] D. Strickland and G. Mourou, "Compression of amplified chirped optical pulses," *Optics communications*, vol. 55, no. 6, pp. 447–449, 1985.

[70] C. Scoby, P. Musumeci, J. Moody, and M. Gutierrez, "Electro-optic sampling at 90 degree interaction geometry for time-of-arrival stamping of ultrafast relativistic electron diffraction," *Physical Review Special Topics-Accelerators and Beams*, vol. 13, no. 2, p. 022801, 2010.

[71] M. Othman, A. Gabriel, M. Hoffmann, F. Ji, E. Nanni, X. Shen, E. Snively, and X. Wang, "Terahertz driven compression and time-stamping technique for single-shot ultrafast electron diffraction," in *Int. Particle Accelerator Conf.(IPAC'21), Campinas, SP, Brazil*. JACOW Publishing, 2021.

[72] F. Cropp, L. Moos, A. Scheinker, A. Gilardi, D. Wang, S. Paiagua, C. Serrano, P. Musumeci, and D. Filippetto, "Virtual-diagnostic-based time stamping for ultrafast electron diffraction," *Physical Review Accelerators and Beams*, vol. 26, no. 5, p. 052801, 2023.

[73] T. Van Oudheusden, E. De Jong, S. Van der Geer, W. t Root, O. Luiten, and B. Siwick, "Electron source concept for single-shot sub-100 fs electron diffraction in the 100 kev range," *Journal of Applied Physics*, vol. 102, no. 9, 2007.

[74] T. Van Oudheusden, P. Pasmans, S. Van Der Geer, M. De Loos, M. Van Der Wiel, and O. Luiten, "Compression of subrelativistic space-charge-dominated electron bunches for single-shot femtosecond electron diffraction," *Physical review letters*, vol. 105, no. 26, p. 264801, 2010.

[75] J. Maxson, D. Cesar, G. Calmasini, A. Ody, P. Musumeci, and D. Alesini, "Direct measurement of sub-10 fs relativistic electron beams with ultralow emittance," *Physical review letters*, vol. 118, no. 15, p. 154802, 2017.

[76] T. P. Wangler, *RF Linear accelerators*. John Wiley & Sons, 2008.

[77] L. Zhao, Z. Wang, C. Lu, R. Wang, C. Hu, P. Wang, J. Qi, T. Jiang, S. Liu, Z. Ma *et al.*, "Terahertz streaking of few-femtosecond relativistic electron beams," *Physical Review X*, vol. 8, no. 2, p. 021061, 2018.

[78] D. Durham, K. Siddiqui, F. Ji, J. G. Navarro, P. Musumeci, R. Kaindl, A. Minor, and D. Filippetto, "Relativistic ultrafast electron diffraction of nanomaterials," *Microscopy and Microanalysis*, vol. 26, no. S2, pp. 676–677, 2020.

[79] K. M. Siddiqui, D. B. Durham, F. Cropp, C. Ophus, S. Rajpurohit, Y. Zhu, J. D. Carlstrom, C. Stavrakas, Z. Mao, A. Raja *et al.*, "Ultrafast optical melting of trimer superstructure in layered 1t'-tate2," *Communications Physics*, vol. 4, no. 1, pp. 1–7, 2021.

[80] F. Ji, D. B. Durham, A. M. Minor, P. Musumeci, J. G. Navarro, and D. Filippetto, "Ultrafast relativistic electron nanoprobes," *Communications Physics*, vol. 2, no. 1, pp. 1–10, 2019.

[81] A. Scheinker, F. Cropp, S. Paiagua, and D. Filippetto, "An adaptive approach to machine learning for compact particle accelerators," *Scientific Reports*, vol. 11, no. 1, p. 19187, Dec. 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-98785-0

[82] F. Sannibale, D. Filippetto, H. Qian, C. Mitchell, F. Zhou, T. Vecchione, R. K. Li, S. Gierman, and J. Schmerge, "High-brightness beam tests of the very high frequency gun at the advanced photo-injector experiment test facility at the lawrence berkeley national laboratory," *Review of Scientific Instruments*, vol. 90, no. 3, p. 033304, Mar. 2019. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.5088521

[83] G. Huang, L. R. Doolittle, Y. L. Xu, and J. Yang, "Low noise digitizer design for lcls-ii llrf," in *North American Particle Accelerator Conf.(NAPAC'16), Chicago, IL, USA*. JACOW Publishing, 2016, p. 3.

[84] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[85] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–41, jan 2022. [Online]. Available: https://doi.org/10.1145%2F3505244

[86] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, and A. Rush, "Transformers: State-of-the-art natural language processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics, Oct. 2020, pp. 38–45. [Online]. Available: https://aclanthology.org/2020.emnlp-demos.6

[87] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[88] W. Falcon et al., "Pytorch lightning," *GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning*, vol. 3, 2019.

[89] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[90] A. M. Dai and Q. V. Le, "Semi-supervised sequence learning," 2015. [Online]. Available: https://arxiv.org/abs/1511.01432

[91] M. G. Minty and F. Zimmermann, *Measurement and control of charged particle beams*. Springer Nature, 2003.

[92] E. Prat and M. Aiba, "Four-dimensional transverse beam matrix measurement using the multiple-quadrupole scan technique," *Physical Review Special Topics-Accelerators and Beams*, vol. 17, no. 5, p. 052801, 2014.

[93] D. Marx, J. G. 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*, vol. 21, no. 10, p. 102802, 2018.

[94] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelmann, "Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems," *Phys. Rev. Accel. Beams*, vol. 23, p. 044601, Apr 2020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.23.044601

[95] R. Roussel, J. P. Gonzalez-Aguilera, Y.-K. Kim, E. Wisniewski, W. Liu, P. Piot, J. Power, A. Hanuka, and A. Edelen, "Turn-key constrained parameter space exploration for particle accelerators using bayesian active learning," *Nature Communications*, vol. 12, no. 1, sep 2021. [Online]. Available: https://doi.org/10.1038%2Fs41467-021-25757-3

[96] F. J. Sacherer, "Rms envelope equations with space charge," *IEEE Transactions on Nuclear Science*, vol. 18, no. 3, pp. 1105–1107, 1971.

[97] P. Lapostolle, "Quelques proprietes essentielles des effets de la charge d'espace dans les faisceaux continus," Tech. Rep., 1970.

[98] P. M. Lapostolle, "Possible emittance increase through filamentation due to space charge in continuous beams," *IEEE Transactions on Nuclear Science*, vol. 18, no. 3, pp. 1101–1104, 1971.

[99] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: the mcmc hammer," *Publications of the Astronomical Society of the Pacific*, vol. 125, no. 925, p. 306, 2013.

[100] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The american statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[101] J. Goodman and J. Weare, "Ensemble samplers with affine invariance," *Communications in applied mathematics and computational science*, vol. 5, no. 1, pp. 65–80, 2010.

[102] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[103] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[104] M. J. Kochenderfer and T. A. Wheeler, *Algorithms for optimization*. Mit Press, 2019.

[105] F. Peiliang, H. Xiaozhong, Y. Liu, W. Tao, J. Xiaoguo, L. Yiding, Z. Xiaoding, W. Ke, and Y. Xinglin, "Simulation of the solenoid scan method used in overlapping field for thermal emittance measurement," *High Power Laser and Particle Beams*, vol. 33, no. 2, pp. 1–4, 2021.

[106] I. V. Bazarov, B. M. Dunham, Y. Li, X. Liu, D. G. Ouzounov, C. K. Sinclair, F. Hannon, and T. Miyajima, "Thermal emittance and response time measurements of negative electron affinity photocathodes," *Journal of Applied Physics*, vol. 103, no. 5, p. 054901, 2008.

[107] F. Cropp, N. Burger, P. Denham, J. G. Navarro, E. Liu, P. Musumeci, L. Phillips, A. Edelen, and C. Emma, "Maximizing 2-d beam brightness using the round to flat beam transformation in the ultralow charge regime," in *North Am. Particle Accel. Conf., Lansing, MI*, 2019.

[108] S. Antipov, D. Broemmelsiek, D. Bruhwiler, D. Edstrom, E. Harms, V. Lebedev, J. Leibfritz, S. Nagaitsev, C. Park, H. Piekarz, P. Piot, E. Prebys, A. Romanov, J. Ruan, T. Sen, G. Stancari, C. Thangaraj, R. Thurman-Keup, A. Valishev, and V. Shiltsev, "Iota (integrable optics test accelerator): facility and experimental beam physics program," *Journal of Instrumentation*, vol. 12, no. 03, p. T03002, mar 2017. [Online]. Available: https://dx.doi.org/10.1088/1748-0221/12/03/T03002

[109] P. Musumeci, R. Agustsson, L. Amoudry, D. Broemmelsiek, D. Bruhwiler, P. Denham, J. Edelen, A. Fisher, C. Hall, T. Hodgetts *et al.*, "Fast-greens: A high efficiency free electron laser driven by superconducting rf accelerator," in *13th International Particle Accelerator Conference (IPAC'22), Bangkok, Thailand, 12-17 June 2022*. JACOW Publishing, Geneva, Switzerland, 2022.

[110] F. Stephan, C. H. Boulware, M. Krasilnikov, J. Bähr, G. Asova, A. Donat, U. Gensch, H. J. Grabosch, M. Hänel, L. Hakobyan, H. Henschel, Y. Ivanisenko, L. Jachmann, S. Khodyachykh, M. Khojoyan, W. Köhler, S. Korepanov, G. Koss, A. Kretzschmann, H. Leich, H. Lüdecke, A. Meissner, A. Oppelt, B. Petrosyan, M. Pohl, S. Riemann, S. Rimjaem, M. Sachwitz, B. Schöneich, T. Scholz, H. Schulze, J. Schultze, U. Schwendicke, A. Shapovalov, R. Spesyvtsev, L. Staykov, F. Tonisch, T. Walter, S. Weisse, R. Wenndorff, M. Winde, L. v. Vu, H. Dürr, T. Kamps, D. Richter, M. Sperling, R. Ovsyannikov, A. Vollmer, J. Knobloch, E. Jaeschke, J. Boster, R. Brinkmann, S. Choroba, K. Flechsenhar, K. Flöttmann, W. Gerdau, V. Katalev, W. Koprek, S. Lederer, C. Martens, P. Pucyk, S. Schreiber, S. Simrock, E. Vogel, V. Vogel, K. Rosbach, I. Bonev, I. Tsakov, P. Michelato, L. Monaco, C. Pagani, D. Sertore, T. Garvey, I. Will, I. Templin, W. Sandner, W. Ackermann, E. Arévalo, E. Gjonaj, W. F. O. Müller, S. Schnepp, T. Weiland, F. Wolfheimer, J. Rönsch, and J. Rossbach, "Detailed characterization of electron sources yielding first demonstration of european x-ray free-electron laser beam quality," *Phys. Rev. ST Accel. Beams*, vol. 13, p. 020704, Feb 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.13.020704

[111] B. Aune, R. Bandelmann, D. Bloess, B. Bonin, A. Bosotti, M. Champion, C. Crawford, G. Deppe, B. Dwersteg, D. A. Edwards, H. T. Edwards, M. Ferrario, M. Fouaidy, P.-D. Gall, A. Gamp, A. Gössel, J. Graber, D. Hubert, M. Hüning, M. Juillard, T. Junquera, H. Kaiser, G. Kreps, M. Kuchnir, R. Lange, M. Leenen, M. Liepe, L. Lilje, A. Matheisen, W.-D. Möller, A. Mosnier, H. Padamsee, C. Pagani,

M. Pekeler, H.-B. Peters, O. Peters, D. Proch, K. Rehlich, D. Reschke, H. Safa, T. Schilcher, P. Schmüser, J. Sekutowicz, S. Simrock, W. Singer, M. Tigner, D. Trines, K. Twarowski, G. Weichert, J. Weisend, J. Wojtkiewicz, S. Wolff, and K. Zapfe, "Superconducting tesla cavities," *Phys. Rev. ST Accel. Beams*, vol. 3, p. 092001, Sep 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.3.092001

[112] S. van der Geer, A. Brynes, I. Setija, P. Smorenburg, P. Williams, and M. de Loos, "GPT-CSR: a New Simulation Code for CSR Effects," in *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018*, ser. International Particle Accelerator Conference, no. 9. Geneva, Switzerland: JACoW Publishing, June 2018, paper THPAK078, pp. 3414–3417, https://doi.org/10.18429/JACoW-IPAC2018-THPAK078. [Online]. Available: http://jacow.org/ipac2018/papers/thpak078.pdf

[113] P. Denham, A. Ody, and P. Musumeci, "X-band cavity based longitudinal phase space linearization at the ucla pegasus photoinjector," in *Proceedings of the North American Particle Accelerator Conference (NAPAC 2022)*, Albuquerque, New Mexico, USA, August 2022.

[114] F. Cropp, D. Filippetto, A. Gilardi, P. Musumeci, S. Paiagua, A. Scheinker, D. Wang *et al.*, "Toward machine learning-based adaptive control and global feedback for compact accelerators," in *13th International Particle Accelerator Conference (IPAC'22), Bangkok, Thailand, 12-17 June 2022*. JACOW Publishing, Geneva, Switzerland, 2022, pp. 991–994.

[115] A. Ody, P. Musumeci, J. Maxson, D. Cesar, R. England, and K. Wootton, "Flat electron beam sources for dla accelerators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 865, pp. 75–83, 2017.

[116] M. Gromov, "Pseudo holomorphic curves in symplectic manifolds." *Inventiones Mathematicae*, vol. 82, p. 307, 1985.

[117] A. J. Dragt, F. Neri, and G. Rangarajan, "General moment invariants for linear hamiltonian systems," *Phys. Rev. A*, vol. 45, pp. 2572–2585, Feb 1992. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.45.2572

[118] L. D. Duffy and A. J. Dragt, "Utilizing the eigen-emittance concept for bright electron beams," in *Advances in Imaging and Electron Physics*. Elsevier, 2016, vol. 193, pp. 1–44.

[119] C. Xiao, O. Kester, L. Groening, H. Leibrock, M. Maier, and P. Rottländer, "Single-knob beam line for transverse emittance partitioning," *Physical Review Special Topics-Accelerators and Beams*, vol. 16, no. 4, p. 044201, 2013.

[120] L. Cultrera, S. Karkare, H. Lee, X. Liu, I. Bazarov, and B. Dunham, "Cold electron beams from cryocooled, alkali antimonide photocathodes," *Phys.*

Rev. ST Accel. Beams, vol. 18, p. 113401, Nov 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.18.113401

[121] D. Filippetto, P. Musumeci, M. Zolotorev, and G. Stupakov, "Maximum current density and beam brightness achievable by laser-driven electron sources," *Phys. Rev. ST Accel. Beams*, vol. 17, p. 024201, Feb 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.17.024201

[122] M. Zhang, "Emittance formula for slits and pepper-pot measurement," Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), Tech. Rep., 1996.

[123] A. Scheinker and M. Krstić, "Minimum-seeking for clfs: Universal semiglobally stabilizing feedback under unknown control directions," *IEEE Transactions on Automatic Control*, vol. 58, no. 5, pp. 1107–1122, 2012.

[124] A. Scheinker, "Model independent beam tuning," in *Int. Partile Accelerator Conf.(IPAC'13), Shanghai, China, 19-24 May 2013*. JACOW Publishing, Geneva, Switzerland, 2013, pp. 1862–1864. [Online]. Available: http://accelconf.web.cern.ch/AccelConf/IPAC2013/papers/tupwa068.pdf?n=IPAC2013/papers/tupwa068.pdf

[125] A. Scheinker and D. Scheinker, "Constrained extremum seeking stabilization of systems not affine in control," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 2, pp. 568–581, 2018.

[126] 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," *Physical review letters*, vol. 121, no. 4, p. 044801, 2018.

[127] P. Denham, F. Cropp, and P. Musumeci, "Analysis of skew quadrupole compensation in rf-photoinjectors," *arXiv preprint arXiv:2003.00049*, 2020.

[128] D. H. Dowell, F. Zhou, and J. Schmerge, "Exact cancellation of emittance growth due to coupled transverse dynamics in solenoids and rf couplers," *Phys. Rev. Accel. Beams*, vol. 21, p. 010101, Jan 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.21.010101

[129] L. Zheng, J. Shao, Y. Du, J. G. Power, E. E. Wisniewski, W. Liu, C. E. Whiteford, M. Conde, S. Doran, C. Jing, C. Tang, and W. Gai, "Experimental demonstration of the correction of coupled-transverse-dynamics aberration in an rf photoinjector," *Phys. Rev. Accel. Beams*, vol. 22, p. 072805, Jul 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.22.072805

[130] 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.*, vol. 130, p. 145001, Apr 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.130.145001

[131] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021999118307125

[132] I. Lagaris, A. Likas, and D. Fotiadis, "Artificial neural networks for solving ordinary and partial differential equations," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 987–1000, 1998.

[133] A. Scheinker and R. Pokharel, "Physics-constrained 3D convolutional neural networks for electrodynamics," *APL Machine Learning*, vol. 1, no. 2, 04 2023, 026109. [Online]. Available: https://doi.org/10.1063/5.0132433

[134] L. Gupta, A. Edelen, N. Neveu, A. Mishra, C. Mayes, and Y.-K. Kim, "Improving surrogate model accuracy for the lcls-ii injector frontend using convolutional neural networks and transfer learning," *Machine Learning: Science and Technology*, vol. 2, no. 4, p. 045025, Oct 2021. [Online]. Available: https://dx.doi.org/10.1088/2632-2153/ac27ff