TOWARDS A HIGHER-FIDELITY PHASE SPACE RECONSTRUCTION IN ACCELERATORS

By

Anthony Duy Tran

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Physics—Doctor of Philosophy

2026

ABSTRACT

Modern accelerator facilities frequently rely on simple RMS-based diagnostics and Gaussian approximations to characterize beam distributions. However, as accelerators push toward higher beam power and increased brightness, these simplified models become insufficient. Characterizing and controlling the beam at the level of individual particle distributions is now essential for maximizing power and preventing beam loss on sensitive components. This thesis addresses this challenge by developing novel algorithms for efficient phase-space reconstruction and implementing them experimentally at the Argonne Tandem Linac Accelerator System (ATLAS).

The work evaluates three distinct approaches to achieve high-fidelity beam characterization. First, leveraging Machine Learning (ML), we developed models to optimize beam transmission and provide virtual diagnostics. Second, we created a robust analysis code for a pepper-pot detector, enabling accurate, "one-shot" direct characterization of the phase space. Finally, we investigated maximum entropy tomography as an indirect method, showing it provides higher-fidelity reconstruction compared to traditional quadrupole scans.

ACKNOWLEDGEMENTS

(Optional) Acknowledge advisors, collaborators, funding sources, and family.

TABLE OF CONTENTS

LIST OF SYMBOLS

Roman Symbols

A	Design matrix for LLSQ
A^+	Pseudoinverse of design matrix
$a = \sqrt{M_{11}^2 + M_{12}^2}$	Scaling factor for beam tomography
a_k	Row of design matrix A for LLSQ
a_n, b_n	Multipole coefficients
B_0	Main dipole field strength
$B_1 = \partial B_y / \partial x$	Quadrupole gradient function
B_k	Main dipole field strength of kicker magnet
B_p	Pole field at r_p
B_x, B_y, B_z	Magnetic field components
$B\rho = p_0/e$	Momentum rigidity of the beam
$f = \lim_{L\to 0} 1/KL$	Quadrupole focal length
$f(\mathbf{x})$	Probability distribution
$\hat{f}$	Reconstructed distribution
$\tilde{f}(k_x, k_y)$	2D Fourier transform of $f(x, y)$
$H(f) = -\iint f \ln f \, d\mathbf{x}$	Entropy
$h, \tilde{h}$	Lagrange factor or h-functions
I	Beam current
(i, j, m)	Iteration variables
J	Total number of cells
J_k	Jacobian
K	Number of measurements
$K(s) = B_1/B\rho$	Effective focusing function
$K_x(s) = 1/\rho^2 + K(s)$	Horizontal focusing function
$K_y(s) = -K(s)$	Vertical focusing function
k, k'	Measurement index
$k_L = \frac{B_{sol}}{2B\rho}$	Larmor wave number

L	Length
ℓ	particle loss
$M(s)$	Betatron transfer matrix of a transport section
N	Number of particles or dimensions
N_p	Total number of pixels
n	Particle index ($n = 1, \ldots, N$)
P	Total number of line integrals
$P^{(e_1)}$	Projection along unit direction e_1
P_θ	Projection of $f(x, y)$ along θ
$\tilde{P}_\theta$	1D Fourier transform of P_θ
$p(\cdot)$	Projection
$\hat{p}$	Simulated projection
Q	Dimensionless perveance
R	Modified matrix used for normalized-space reconstruction
$\mathbf{r} = A\mathbf{x} - \mathbf{b}$	Residual vector
r_b	Beam radius
r_p	Beam pipe radius
$\mathbf{u} = M\mathbf{x}$	Transformed coordinates
(w, φ)	Longitudinal phase space coordinates (energy, phase)
w_{ji}	Contribution of cell to integral
$\mathbf{x}$	State vector
$(\hat{x}, \hat{y}, \hat{s})$	Frenet-Serret coordinate system
(x, x')	Horizontal betatron phase space coordinates
(x_N, x'_N)	Normalized phase-space coordinates
(y, y')	Vertical betatron phase space coordinates
(z, δ)	Longitudinal phase space coordinates

Greek Symbols

α, β, γ	Twiss parameters
$\beta_r = v/c$	Relativistic velocity

$\gamma_r = 1/\sqrt{1-\beta_r^2}$	Lorentz factor
Δx_q	Quadrupole misalignment
ϵ	Beam emittance
θ	Rotation angle
$\theta_k = \Delta B_k L/(B\rho)$	Dipole kick angle
λ	Line density
$\lambda_k(u_k)$	Lagrange multipliers
$\rho(s)$	Bending radius
σ	Beam matrix
Φ	Rotation angle for solenoid
ψ	Phase advance
Ψ	Lagrangian functional
ω	Relaxation parameter

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ALS	Advanced Light Source
APS	Advanced Photon Source
ART	Algebraic Reconstruction Technique
AT	Accelerator Toolbox
ATLAS	Argonne Tandem Linac Accelerator System
AWA	Argonne Wakefield Accelerator
BPM	Beam Profile Monitor
cDVAE	Conditional Diffusion Variational Autoencoder
CNN	Convolutional Neural Network
CSR	Coherent Synchrotron Radiation
DCT	Discrete Cosine Transform
DOE	U.S. Department of Energy
ECR	Electron Cyclotron Resonance
EEX	Emittance Exchange
ELU	Exponential Linear Unit
ESRF	European Synchrotron Radiation Facility
FCC	Future Circular Collider
FCP	Faraday Cup (ATLAS device tag prefix)
FEL	Free-Electron Laser
FRIB	Facility for Rare Isotope Beams
GANIL	Grand Accélérateur National d'Ions Lourds
GPSR	Generative Phase Space Reconstruction
GPU	Graphics Processing Unit

HLLHC High-Luminosity Large Hadron Collider

HPC High-Performance Computing

HIT-LEBT Heidelberg Ion Therapy Low Energy Beam Transport

IBIC International Beam Instrumentation Conference

KV Kapchinskij–Vladimirskij

LCLS Linac Coherent Light Source

LEBT Low Energy Beam Transport

LHC Large Hadron Collider

LINAC Linear Accelerator (conference series acronym)

LLSQ Linear Least Squares

MAE Mean Absolute Error

MAD-X Methodical Accelerator Design (code)

MCMC Markov Chain Monte Carlo

MENT Maximum Entropy Tomography

ML Machine Learning

MOPA Michigan State University High-Performance Computing cluster name (as used in figures)

MSE Mean Squared Error

MSU Michigan State University

NAPS Normalized Adaptive Projection Sampling

NEC National Electrostatics Corporation

NN Neural Network

NUPS Normalized Uniform Projection Sampling

ORNL Oak Ridge National Laboratory

PAC Particle Accelerator Conference

PCA Principal Component Analysis

PII Positive Ion Injector II

PMP Profile Monitor (ATLAS device tag prefix)

QDP Quadrupole Doublet (ATLAS device tag prefix)

QSP Quadrupole Singlet (ATLAS device tag prefix)

RAON Rare isotope Accelerator complex for ON-line experiments

ReLU Rectified Linear Unit

RF Radio Frequency

RFQ Radio-Frequency Quadrupole

RIBF RIKEN Radioactive Isotope Beam Factory

RMS Root Mean Square

SART Simultaneous Algebraic Reconstruction Technique

SNS Spallation Neutron Source

SPIRAL2 Système de Production d'Ions Radioactifs Accélérés en Ligne 2

SRF Superconducting Radio Frequency

SSIM Structural Similarity Index Measure

STP Steerer (ATLAS device tag prefix)

SWFA Structure-based Wakefield Accelerator

TDS Transverse Deflecting Structure

TESLA TeV-Energy Superconducting Linear Accelerator

TRIUMF Tri-University Meson Facility

UQ Uncertainty Quantification

VAE Variational Autoencoder

CHAPTER 1

INTRODUCTION TO ACCELERATORS

This section provides an introduction to the history of particle accelerator and the major advances which motivate the need of a higher-fidelity phase space reconstruction.

1.1 A Brief History and Overview

A particle accelerator is a device that accelerates, transports, and manipulates particles, including electrons, protons, heavy ions, muons, and even neutrons [1]. Accelerator and beam physics is the science of the motion, generation, acceleration, manipulation, prediction, observation, and use of charged particle beams.

The most fundamental particle accelerator consists of a static electric field generated between two parallel charged plates. The electric field between the plates accelerates a charged particle; however, this is constrained by the breakdown voltage, as well as the ability to effectively inject and extract particles. The following highlights the evolution of accelerators from the Cathode-Ray Tube in 1897 to the Large Hadron Collider (LHC) in 2008.

- **1926 – The X-Ray Tube:** William David Coolidge achieved 900-keV electron beam energy with 3 X-ray tubes in series [2].

- **1929 – The Cyclotron:** Ernest Lawrence invented the cyclotron, which accelerates particles, first by stabilizing them in a constant magnetic field, and then repeatedly accelerating them with an RF source in a small gap between the magnets [3].

- **1932 – The Cockcroft-Walton Generator:** John Cockcroft and Ernest Walton reached 400-kV and achieved the first man-made nuclear transmutations: $p + Li \rightarrow 2He$ [4]

- **1940 – The Betatron:** Donald Kerst built and operated the first betatron, achieving 2.3 MeV. It was based on the idea that a changing magnetic field induces an electric field [5].

- **1945 – Phase Stability and the Synchrotron:** Vladimir Veksler [6] and Edwin McMillan [7] independently discovered the principle of phase stability in rf acceleration. This

overcame the relativistic limits of the cyclotron, in which, due to the relativistic mass effect, particle synchronism is destroyed, limiting the proton's maximum kinetic energy to about 12 MeV.

- **1947 – First Synchrotron Light:** General Electric observes synchrotron radiation, light emitted by accelerated charged particles, for the first time in a 70 MeV electron synchrotron. This discovery led to a revolutionary scientific tool, laying the foundation to probe matter at the atomic scale [8].
- **1952 – Strong Focusing:** Courant, Snyder, and M.S. Livingston discovered the principle of "strong focusing." This significantly reduced the required magnet size and enables multi-GeV accelerators [9].
- **2008 – The Large Hadron Collider (LHC):** This is the largest accelerator to date at the time of this writing. In 2012, it enabled the discovery of the Higgs Boson [10, 11].

At the start, novel acceleration methods allowed for significantly higher beam energies. Eventually, performance became limited by beam stability and control. The introduction of strong focusing allowed for more efficient designs with fewer magnets. In the current era, beam power is primarily limited by collective and non-linear effects. This work presents research towards a better understanding of the beam distribution to model these effects more effectively. Accelerators continue to evolve, driving innovations that facilitate new technologies and scientific breakthroughs.

1.2 Grand Challenges in Accelerator Science

According to the 2022 Accelerator and Beam Physics Roadmap Report [12], there are currently 4 grand challenges in accelerators that help define the field's continued progress and make accelerators cheaper, safer, and more reliable. The four great challenges are: beam intensity, beam quality, beam control, and beam prediction.

1.2.1 Grand Challenge 1: Beam Intensity

As beam intensities are increased by an order of magnitude, novel physical constraints emerge that must be addressed. Modern accelerators have evolved to a regime where collective effects

and particle loss are the dominant performance limitations. While negligible at low intensities, collective effects severely constrain high-intensity beam operations. One prominent collective phenomenon is the space-charge effect. Space-charge forces perturb the beam, potentially driving it into resonant states, amplifying betatron oscillations, and ultimately inducing beam loss. For a very intense beam, even a part-per-million loss can damage equipment and cause radiation events.

Simulation models were developed to model these collective effects. There exist models that are successfully benchmarked against theoretical predictions that account for space-charge effects, but these are limited by uncertainties in the beam and accelerator, particularly the input distribution. It has become necessary to have a precise knowledge of the full 5-6D input phase space, along with all the lattice details and diagnostic tools in order to measure the distribution of beam loss to under 1 ppm.

Measuring the phase-space has been proven benifical. The first 6D phase-space direct measurement was performed in 2018 [13] and revealed hidden intensity-dependent correlations not observed in lower-dimensional projection measurements; however, a multi-hour scan was required it only yielded a reconstruction with approximately 10-20 points per dimension. The goal set by the community is to achieve a high resolution of more than 1000 points per dimension within a 1-2-hour scan time.

This motivates the search for an efficient method of phase-space reconstruction. Some examples include, tomography and indirect measurement technique, as well as phase-space painting which creates the phase-space distribution instead of measuring it [14].

1.2.2 Grand Challenge 2: Beam Quality

This challenge focuses on increasing the beam's phase-space density by orders of magnitude toward the quantum-degeneracy limit, where the phase-space volume occupied by the beam is minimized. In the ideal case, such compression could produce a *crystalline* beam, enabling unprecedented beam quality, including higher luminosity and the generation of copious coherent radiation [15]. Beam density is a key parameter because it directly affects two primary performance metrics. For high-energy colliders, this figure of merit is luminosity, whereas for light sources it is

brightness.

Luminosity

Luminosity is defined as the number of particle collisions per unit area per unit time. This is used to quantify the performance of collider accelerators. An approximate formula can be formulated by considering two identically Gaussian distributed beam bunches [16].

$$\mathcal{L} = \frac{f N_1 N_2}{4\pi \sigma_x \sigma_y}. \tag{1.1}$$

- N_1, N_2 = number of particles per bunch in each beam
- f = collision frequency
- σ_x, σ_y = transverse RMS beam sizes at the interaction point

Luminosity usually degrades over time as particles collide. As a result, it is often more informative to measure integrated luminosity, the total number of events over a given period of time.

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \tag{1.2}$$

Brightness

Brightness quantifies the performance of a light source. It determines its usability and measures the concentration of the charged particle beam. It is defined as the number of photons, N_{ph}, per unit time, per unit bandwidth, per phase-space volume ($d\Omega\, dS \approx \sigma_x \sigma_{x'} \sigma_z \sigma_{z'}$). In other words, it describes the intensity, collimation, and spectral purity of the light. High brilliance is desirable in many fields of science. Since it is inversely proportional to the emittance, high brightness demands a beam with a small emittance.

$$\mathcal{B} = \frac{d^4 N}{dt d\Omega dS (d\lambda/\lambda)} = \frac{d^4 N_{\text{ph}}}{dt \sigma_x \sigma_{x'} \sigma_y \sigma_{y'} (d\lambda/\lambda)} \tag{1.3}$$

$$\text{Units: } \frac{\text{Photons}}{\text{s} \cdot \text{mm}^2 \cdot \text{mrad}^2 \cdot 0.1\% \text{ bandwidth}}$$

1.2.3 Grand Challenge 3: Beam Control

Currently, most beam-level controls remain at the macroscopic scale. This involves assuming a Gaussian beam distribution and characterizing it using statistical moments and root-mean-squared (RMS) emittance estimates. Individual-particle-level control and detection would aid in the realization of crystalline beams. At present, significant efforts are directed towards mesoscopic-scale design, including phase-space painting and tomography. Again, we see the critical need for efficient techniques for characterizing the 6D phase-space distributions.

1.2.4 Grand Challenge 4: Beam Prediction

With recent advancements in machine laerning (ML) and artificial intelligence (AI), data-driven approaches may be increasingly used to model beamlines. According to the roadmap report, computer modeling is an essential component of advanced beam physics, particle accelerator design, and operation. Future developments aim to yield models sufficiently sophisticated in physics and engineering to enable a virtual accelerator. This requires the creation of virtual test stands in order to push the frontiers of theory and modeling. To this end, there is a need to integrate ML/AI into accelerator simulations and control systems, leveraging high-performance computing, advancing underlying mathematical models and algorithms, and reducing technical risk while increasing efficiency. This is a community-wide effort, and these resources should be accessible to the entire community. Again, since accurate modeling of the beam dynamics requires both an accurate model of the beamline and the inital beam distribution, effective measurement of the phase space will become necessary.

1.3 Impact of Phase Space in Various Accelerator Facilities

This section will cover various accelerator facilities and their overall goals. Understanding the phase space can support verification of the existing beam distribution, prevent beam loss, and optimize various accelerator facility goals.

1.3.1 Synchrotron Light Sources and Free-Electron Lasers (FELs)

Synchrotron light sources are circular accelerators that accelerate electrons to relativistic speeds. When these particles are bent by magnets, they emit synchrotron radiation, a form of light charac-

terized by several unique properties. This light is extremely bright and collimated, offering a wide energy spectrum that spans from infrared light to hard X-rays. Because the beam is highly tunable, researchers can isolate an intense beam of any selected wavelength. This radiation can be utilized to resolve atomic structures of atoms in material science or the structure of proteins in life science.

Prominent facilities like the Advanced Photon Source (APS) at Argonne National Laboratory, the Advance Light Source (ALS) at Berkeley, and the European Synchrotron Radiation Facility (ESRF) serve as hubs for this research. These facilities employ specialized insertion devices, such as wigglers and undulators, which induce periodic transverse oscillations in the electron beam to increase photon emission. These facilities support a diverse array of disciplines, including chemistry, materials science, pharmaceutical sciences, condensed matter physics, and earth sciences. A comprehensive overview of industrial applications can be found here [17].

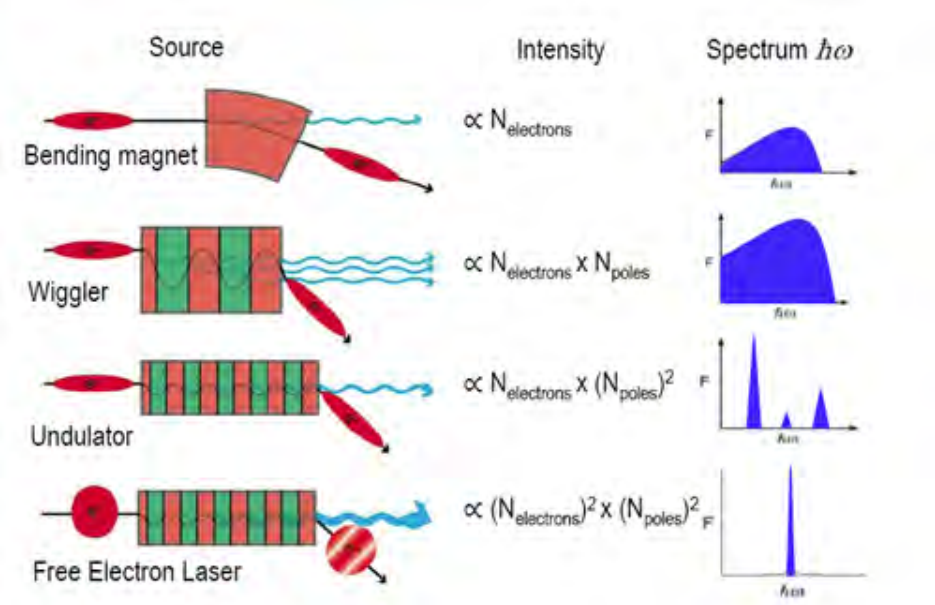


Figure 1.1 Forms of Synchrotron Radiations [18].

In the early 1970s, it was found that electrons in an undulator can exchange energy with a laser beam along the same direction. With the right relationship, the laser can be amplified at the expense of the electron beam. In the 1980s it was found that if the beam current were increased by three orders of magnitude, the electrons would undergo microbunching, leading to coherent spontaneous emission. To create this, the beam was compressed to a much smaller space. As seen

in figure 1.1, FEL produces light that is more coherent, both temporally and spatially. As stated in [19] regarding the Linac Coherent Light Source (LCLS) at SLAC, "They carry a million times more pulse energy than synchrotron x-rays, are 10 000 times shorter, and have coherence that can produce focused x-ray beams with intensities up to 10^{20} W/cm^2, more than a billion times greater than any previously achieved."

Users of these facilities require high repetition rates to maximize data acquisition and enable the study of ultrafast time-resolved dynamics, such as bond formation and cleavage. Attosecond pulses $(10^{-}18)$ enable ultra-fast process observation and time-resolved studies while minimizing sample damage.

Since brightness is inversely proportional to emittance, the study of phase-space manipulation, specifically maintaining phase-space density with low emittance, is of significant interest. Collective effects, such as space charge and coherent synchrotron radiation (CSR), dilute emittance. Consequently, the prediction and experimental validation of beam phase space, down to the halo particle level, remains a fundamental area of research [12].

1.3.2 Particle Collider

Colliders are accelerators that collide particles together at extremely high energies to probe fundamental physics. An example is the Large Hadron Collider (LHC) at CERN. Researchers have used these machines to discover new particles, such as the Higgs boson, and to test theories The next upgrade would lead to the High-Luminosity LHC (HL-LHC) project to increase luminosity by a factor of 10, and after that the Future Circular Collider (FCC).

As beam power and intensity increase, having a precise initial distribution becomes critical. For example, it is generally accepted that 1 W/m of uncontrolled beam loss is an acceptable standard [20]. Given an average power P, the fractional accepted beam loss is $\approx \frac{1}{P}$. The average power of modern hadron accelerators has exceeded 1 MW and now reaches 10 MW. This corresponds to $f \approx 10^6$. This is the level of particles on the outer edge of the beam distribution, the beam halo. These have the potential to damage beam components. Thus, this motivates the need to characterize a beam with a dynamic range of up to 10^6. From a machine-protection perspective, a better

understanding of phase-space reconstruction is critical for the LHC and future colliders in order to model and mitigate these effects.

1.3.3 Spallation Sources

Like photons, neutrons are able to penetrate deeply into matter in order to probe the structures. Unlike photons which mainly interact with the electrons, neutron interact with the nuclei of atoms, offering a complementary view of the atomic structure. Spallation sources are designed to generate neutrons via the spallation process [21], in which a proton beam impinges upon a heavy-metal target, such as tungsten or liquid mercury. At the Spallation Neutron Source (SNS) at Oak Ridge National Laboratory, protons are accumulated in a ring to high intensities before being extracted onto a liquid mercury target in short pulses. To meet the primary objective of maximizing neutron production power, phase-space painting is employed to accumulate the beam from the linear accelerator into the ring, enabling the delivery of world-leading power to the target.

Currently, the limits on beam intensity and power are primarily dictated by collective effects. While the Facility for Rare Isotope Beams (FRIB) serves as a notable exception, where collective effects are not expected to be the limiting factor even for the 400 kW upgrade, these phenomena remain the dominant constraint for most high-power facilities. The theories and models governing these dynamics are typically validated through simulation; however, an accurate initial distribution is required to faithfully reproduce the observed impacts [12].

As mentioned before, it has been demonstrated that a 6D phase-space measurement is necessary to fully characterize the beam's complete phase-space structure [22]. Although the first 6D phase-space measurement was recently achieved at the SNS [23], it remains impractical for routine operations due to its low resolution (10–20 points per dimension) and the multiple hours required for acquisition. The current strategic target is to achieve a resolution of approximately 1000 points per dimension within 1–2 hours [12]. To bridge this gap, efficient phase-space reconstruction techniques are needed.

1.3.4 Structured-Wakefield Accelerators

It is also important to measure and control the phase space at intermediate stages of the beamline, particularly for applications like Structure-based Wakefield Accelerators (SWFA) driving light sources. Conventional Radio Frequency (RF) cavities, which typically operate with gradients of approximately 20–$50\,\mathrm{MV/m}$, are the standard for particle acceleration. However, as the demand for higher beam energies increases, more efficient high-gradient methods are required to prevent accelerators from becoming prohibitively large. Advanced technologies, such as SWFA, can achieve gradients approaching $400\,\mathrm{MV/m}$ [24]. Several schemes are currently under investigation, and prior studies indicate that shaping the drive-beam current profile can significantly improve acceleration efficiency [25].

1.3.5 Heavy Ion Accelerators

Finally, the Heavy-Ion Accelerator is the last type of accelerator to be covered in this introduction. Some of them include:

- ATLAS, Argonne Nat. Lab (USA)
- RIBF, RIKEN Nishina Center (Japan)
- FRIB, Michigan State Univ. (USA)
- GANIL and SPIRAL2, Caen (France)
- TRIUMF, Vancouver (Canada)
- RAON, Daejeon (South Korea)

FRIB, as seen in Fig. 1.2, at the time of this writing, is a world-leading high-power machine, with a baseline key feature of 400 kW beam power on target for all ions from protons to uranium, with energy $\geq 200\,\mathrm{MeV}/\mu$. It offers in-flight isotope separation to provide fast development times for any isotope, beams of all elements, and fast, stopped, and re-accelerated beams [26].

These accelerators operate at high beam power, and minimizing beam loss is the highest priority to prevent radiation and structural damage. Simple optimization assuming a Gaussian beam is insufficient because halo particles, if not accounted for, can damage sensitive and expensive components such as Superconducting Radio Frequency (SRF) cavities.

In another example, the Argonne Tandem Linac Accelerator System (ATLAS) utilizes distinct ion species for various experiments, often changing ion species more than once per week. Retuning a beamline for these changes is time-consuming. One common tuning method involves scaling the

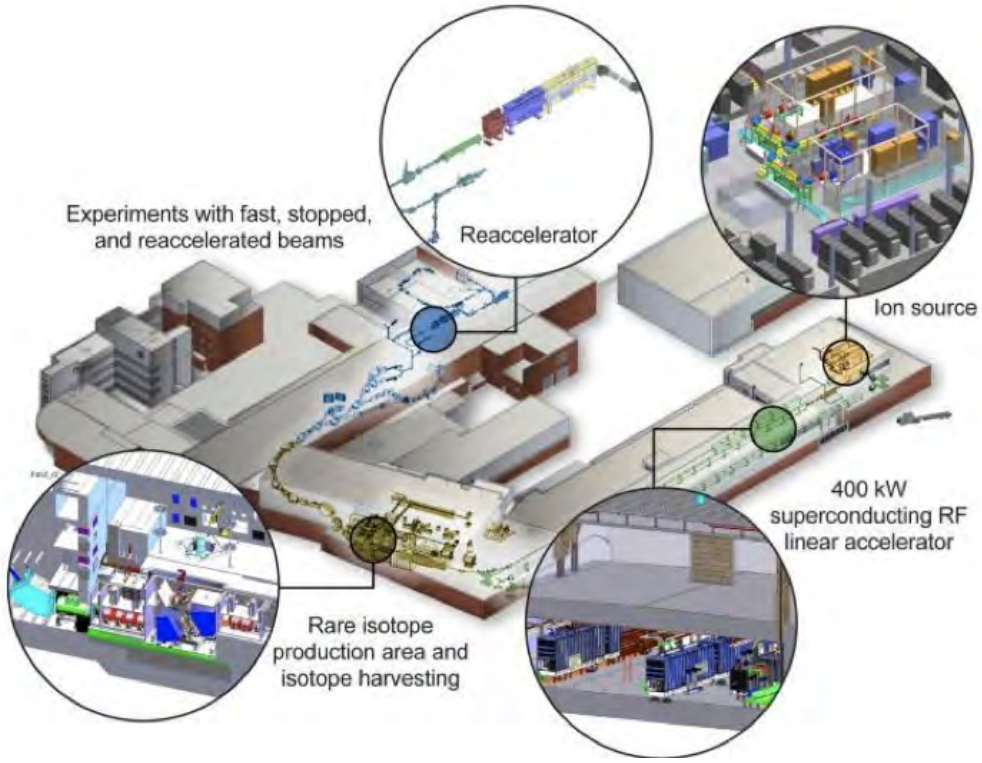


Figure 1.2 Layout of FRIB.

previous tune according to the new beam's rigidity; however, this still often results in significant residual beam loss. A primary factor contributing to this loss is the unknown initial distribution. Even an approximate Gaussian model that accounts for beam offset would improve the efficiency of this method. At ATLAS, efforts are underway to match experimental data to a virtual model created in TRACK [27]. Research has demonstrated that fitting misalignments and the initial beam distribution using a Gaussian model can match experimental transmission with an average error of 6% [28]. Using a mesoscopic initial distribution would likely further reduce this discrepancy.

1.3.6 Medicine

In oncology, Hadron therapy utilizes the unique physical properties of high-energy protons to treat cancer. This procedure requires extremely precise phase-space control for dose control and for validating the initial distribution assumptions. First proposed by Wilson [29, 30], this method leverages the Bragg peak (Fig. 1.3) to deposit maximum energy at a precise depth, targeting tumors while sparing healthy tissue. As of 2023, the demand for this high-precision treatment has led to over 100 operating proton centers worldwide [31].

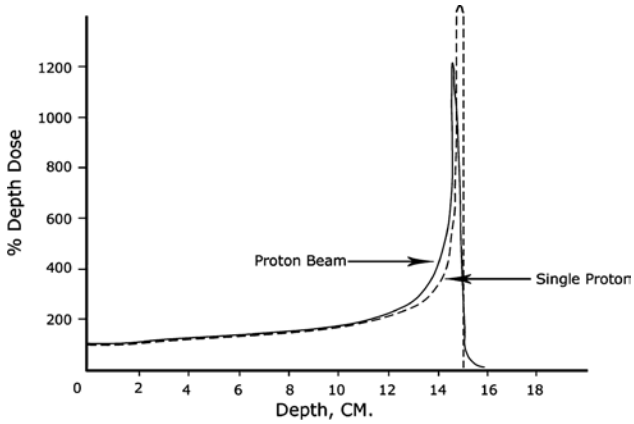


Figure 1.3 The calculated depth dose curve for a 140 MeV proton beam in tissue and a single proton [29].

1.3.7 Industry

Industrial applications often require precise control of beam phase space. In the semiconductor industry, ion implantation is used to set dopant concentrations with extremely high precision, down to the single-ion level, for advanced transistors and photovoltaic devices.

Beyond microelectronics, high-power electron beams are used for rapid sterilization of medical devices, wastewater and waste treatment, and food irradiation (“cold pasteurization”). Particle accelerators also support security applications such as cargo screening, as well as cultural-heritage science through non-destructive materials analysis. For example, the AGLAE facility at the Louvre employs MeV-scale ion beams to determine artwork composition without damaging the objects.

1.4 Phase-Space Reconstruction in Accelerators

In many accelerator systems and analysis, an accurate initial beam distribution is frequently unknown. This thesis examines the use of various methods to obtain a high-accuracy phase-space distributions. While the Gaussian distribution is often utilized as a primary approximation, alternative models, such as the "waterbag" distribution, offer a distinct advantage by providing a hard-edged boundary. This removes the infinite tails inherent in Gaussian profiles, resulting in a more realistic representation. The Kapchinskij–Vladimirskij (KV) distribution [32] is widely used in space-charge studies because it produces linear self-fields. As beam energy and intensity continue to increase, these conventional macroscopic descriptions are becoming increasingly

inadequate for the demands of next-generation facilities.

Phase-space reconstruction algorithms, which can utilize the high dynamic range of 1D measurements from beam profile monitors, currently targeted to achieve a dynamic range of at least 10^6 [12] in the future. Initially, these methods required multiple measurements, substantial computing resources, and several hours of runtime for higher-dimensional reconstructions. Recent advances in algorithms and computing power, however, have enabled the reconstruction of 6D data in a few hours using as few as 20 measurements [33]. This has proven useful at the Argonne Wakefield Accelerator (AWA) for resolving the beam distribution from its Emittance Exchange (EEX) beamline.

Phase-space reconstruction algorithms for accelerators are not yet sufficiently mature for routine, practical use in beamline operations. This work evaluates several approaches aimed at advancing these methods toward operational deployment.

1.5 Thesis Outline

This thesis investigates advancements in particle accelerator techniques aimed at improving the understanding of the initial beam distributions. Direct, indirect, and machine-learning methods are applied to reconstruct these initial distributions and optimize beam transmission along the beamline. This work combines novel algorithmic development with experimental validation at the ATLAS facility to advance this field.

Chapter 2 provides a foundational introduction to accelerator physics, detailing the basic theory necessary for the subsequent chapters. It then examines various methods for phase-space estimation to identify the regimes where beam tomography is most advantageous. Finally, the chapter discusses current methodologies for quantifying uncertainty in these reconstructions.

Chapter 3 focuses on computational simulations for beamline modeling and phase-space reconstruction. It first presents a data-driven approach to predict beam loss and project phase-space distributions at various locations along the beamline. Motivated by the necessity of an accurate initial distribution for this model, the chapter explores methods for generalizing various reconstruction algorithms to higher dimensions and introduces a newly developed algorithm to optimally select

measurements for more efficient tomographic reconstruction.

Chapter 4 details experiments conducted at the ATLAS facility to measure the transverse phase space, leveraging its characteristics as a primarily uncoupled system. It begins with studies utilizing a Pepper-Pot Emittance Meter, a direct measurement technique, before transitioning to indirect methods. This includes an initial quadrupole scan to estimate the distribution, followed by beam tomography to achieve a higher-fidelity reconstruction.

Finally, Chapter 5 concludes the dissertation with a summary of the primary research contributions, an overview of the experimental results, and a discussion of the future outlook for this work.

CHAPTER 2

PHASE-SPACE MEASUREMENT METHODS

This chapter discusses the fundamental accelerator physics concepts used throughout the rest of this thesis, current accelerator diagnostics for direct and indirect measurements, and data-driven measurements. It is not intended to be an exhaustive review, but rather a summary of standard procedures that help readers place this work within the broader field. General machine learning concepts are not included in order to keep the focus on beam physics, but readers are directed to [34] as a useful resource.

2.1 Primer on Accelerator Physics

Confining and accelerating a particle beam is fundamentally a stability problem. Particles drift away from the center, horizontally or vertically, and are focused back. In many accelerator designs, this is accomplished with quadrupole magnets arranged in an alternating sequence, which produces an overall net focusing effect in both transverse planes.

A key advantage of this setup is that the beamline can be treated as a sequence of discrete elements, each with a well-defined linear effect on the transverse phase space. This lets us model the dynamics piecewise and represent each element by a transfer matrix, so that the behavior of an entire section of the accelerator is obtained by multiplying the matrices in order. This matrix formalism will be the main tool used in the analysis that follows.

2.1.1 Frenet-Serret Coordinate System

It is useful to consider a coordinate system along a reference orbit rather than the laboratory frame. Figure 2.1 shows the coordinate system that we will use. The variables x and y correspond to the horizontal and vertical displacement in the transverse plane, while the variable s, the reference path length, is the independent variable. If this were not used, complex time-dependent functions would be needed to describe particles' motion in the lab frame. As a result, the motion behaves like a harmonic oscillator around the reference orbit $(0, 0, 0)$.

In beam dynamics calculations, we assume a paraxial beam, a beam where $x' \ll 1$, $y' \ll 1$, $p \approx p_z$. In this new coordinate system, the conjugate momenta are with respect to s, the path of

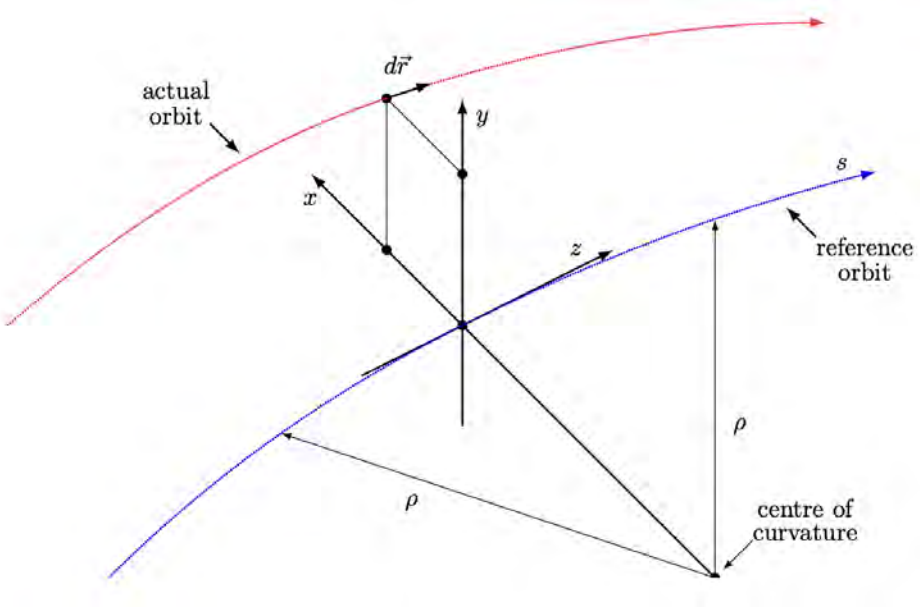


Figure 2.1 Frenet-Serret coordinate system [35].

the reference orbit.

$$x' = \frac{dx}{ds}, \quad y' = \frac{dy}{ds} \tag{2.1}$$

2.1.2 Hill's Equation

The equations of transverse motion in an accelerator in the presence of dipole or quadrupole magnets are given by Eq. 2.2. These are known as the Hill's Equation [36], where $K_1(s) = \mp B_1(s)/B\rho$, $B_1 = \partial B_y/\partial x$ is the quadrupole field gradient, and $\rho(s)$ is the radius of curvature. For a straight section, $\rho(s) = \infty$ and thus the dynamics become similar in both planes, but opposite for nonzero values of K.

$$\begin{aligned} x'' + K_x(s)x = 0, \qquad & y'' + K_y(s)y = 0, \\ K_x = 1/\rho^2 + K_1(s), \quad & K_y = -K_1(s). \end{aligned} \tag{2.2}$$

The two main accelerator components are the drift sections, dipole magnets, and quadrupole magnets. These components, when considered separately, usually have uniform fields or can be approximated as uniform, leading to the focusing functions, $K_{x,y}$, being essentially piecewise. The

particle's motion can be solved piecewise, facilitating a matrix-based solution. Given a constant K, the solution to Eq. 2.2 is

$$x(s) = \begin{cases} a\cos(\sqrt{K}s + b), & K > 0, \\ as + b, & K = 0, \\ a\cosh(\sqrt{-K}s + b), & K < 0. \end{cases} \tag{2.3}$$

Here, x represents both the x and y coordinates. Let a state-vector $\mathbf{x}$ and transfer matrix be $M\,(s)$.

$$\mathbf{x}(s) = \begin{pmatrix} x(s) \\ x'(s) \end{pmatrix}; \quad \mathbf{x}(s) = M\,(s \mid s_0)\,\mathbf{x}\,(s_0)\,. \tag{2.4}$$

Let L be the beam element length, the transfer matrix for different values of constant K is given as

$$M\,(s) = \begin{cases} \begin{pmatrix} \cos\sqrt{K}L & \frac{1}{\sqrt{K}}\sin\sqrt{K}L \\ -\sqrt{K}\sin\sqrt{K}L & \cos\sqrt{K}L \end{pmatrix} & K > 0 : \text{focusing quad}, \\ \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} & K = 0 : \text{drift space}, \\ \begin{pmatrix} \cosh\sqrt{|K|}L & \frac{1}{\sqrt{|K|}}\sinh\sqrt{|K|}L \\ \sqrt{|K|}\sinh\sqrt{|K|}L & \cosh\sqrt{|K|}L \end{pmatrix} & K < 0 : \text{defocusing quad}. \end{cases} \tag{2.5}$$

The net transfer matrix through a sequence of n elements is the ordered product of the individual transfer matrices,

$$M(s) = M_n M_{n-1} \cdots M_2 M_1, \tag{2.6}$$

so that for a state vector $\mathbf{x}_0$ at the entrance,

$$\mathbf{x}(s) = M(s)\,\mathbf{x}_0, \tag{2.7}$$

with M_1 acting first and M_n acting last. This defines a trajectory for a state.

The set of points which contains all possible states of a dynamical system is called the phase space.

2.1.3 Transfer Matrices

While real quadrupole magnets have finite length L, it is often computationally and conceptually advantageous to treat them as 'thin' elements. Thin elements assumes $L \to 0$, where $f = \lim_{L\to 0} \frac{1}{|K|L}$. The transfer matrices from the previous section, and thin quadrupole elements, are given as

$$M_{\text{drift}} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}, \quad M_{\text{thin}} = \begin{pmatrix} 1 & 0 \\ \mp\frac{1}{f} & 1 \end{pmatrix}, \tag{2.8}$$

$$M_{\text{QF}} = \begin{pmatrix} \cos\sqrt{K}L & \frac{1}{\sqrt{K}}\sin\sqrt{K}L \\ -\sqrt{K}\sin\sqrt{K}L & \cos\sqrt{K}L \end{pmatrix}, \quad M_{\text{QD}} = \begin{pmatrix} \cosh\sqrt{K}L & \frac{1}{\sqrt{K}}\sinh\sqrt{K}L \\ \sqrt{K}\sinh\sqrt{K}L & \cosh\sqrt{K}L \end{pmatrix}. \tag{2.9}$$

Given magnetic field errors, the Hill's equations (2.2) obtain magnetic perturbation fields.

$$\begin{gathered} x'' + K_x(s)x = \frac{\Delta B_y}{B\rho}, \quad y'' + K_y(s)y = -\frac{\Delta B_x}{B\rho}, \\ \Delta B_y + j\Delta B_x = B_0 \sum_{n=0}^{\infty} (b_n + ja_n)(x+jy)^n. \end{gathered} \tag{2.10}$$

The magnetic fields B_y and B_x are expanded in a complex field representation known as the Beth representation. This is the U.S. convention where B_0 is the main dipole field strength, and b_n and a_n are the $2(n+1)^{th}$ multipole coefficients. The set of b coefficients is the normal component while the set of a is the skew component [37].

The result of this is a constant kick or misalignment. Using the matrix formalism, the state-vector, extended to $\mathbf{x} = (x, x', 1)^T$, and a 3x3 transfer matrix are used to model this effect as show in Eq. 2.11 for a thin-lens approximation. The dipole kick is $\theta_k = \Delta B_k L/B\rho$ where B_k is the magnetic field of the dipole, and the quadrupole misalignment is Δx_q [36].

$$M_{dipole} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \theta_k \\ 0 & 0 & 1 \end{pmatrix}, \quad M_{quad} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{f} & 1 & \frac{\Delta x_q}{f} \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.11}$$

The transfer matrix formalism can be extended to linear elements that introduce transverse coupling, such as solenoids and skew quadrupoles. These would be useful later in this work to expand 2D techniques to 4D. In these cases, the horizontal and vertical motions are no longer independent, necessitating a 4×4 matrix representation where the expanded state vector $\mathbf{x} = (x, x', y, y')$.

A skew quadrupole is a standard quadrupole magnet rotated by an angle θ about the longitudinal axis. In its upright orientation ($\theta = 0$), the transfer matrix is block-diagonal with zero coupling. To derive the matrix for the rotated element, we apply a coordinate rotation R to the standard thin-lens matrix M_{thin} using the similarity transformation $M_{\text{skew}} = R^{-1} M_{\text{thin}} R$. This results in the following coupled matrix [38]

$$R = \begin{pmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & \cos\theta & 0 & \sin\theta \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & -\sin\theta & 0 & \cos\theta \end{pmatrix}, \quad M_{\text{skew}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{f}\cos(2\theta) & 1 & \frac{1}{f}\sin(2\theta) & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{f}\sin(2\theta) & 0 & -\frac{1}{f}\cos(2\theta) & 1 \end{pmatrix} \tag{2.12}$$

A solenoid is a long, cylindrically wound coil of wire that produces a uniform longitudinal magnetic field. It performed simultaneous focusing in the horizontal and vertical planes due to the fringe fields, while also coupling the motions. It is used at FRIB to focus lower-energy beams. While a solenoid has non-uniform fields at the edge, a hard-edge equivalence can be used with sufficient accuracy, where $k_L = \frac{B_{sol}}{2B\rho}$ is the Larmor wave number and $\Phi = k_L L$ is the rotation angle to obtain [39]

$$M_{\text{sol}} = \begin{pmatrix} \cos^2\Phi & \frac{1}{2k_L}\sin(2\Phi) & \frac{1}{2}\sin(2\Phi) & \frac{1}{k_L}\sin^2\Phi \\ -\frac{k_L}{2}\sin(2\Phi) & \cos^2\Phi & -k_L\sin^2\Phi & \frac{1}{2}\sin(2\Phi) \\ -\frac{1}{2}\sin(2\Phi) & -\frac{1}{k_L}\sin^2\Phi & \cos^2\Phi & \frac{1}{2k_L}\sin(2\Phi) \\ k_L\sin^2\Phi & -\frac{1}{2}\sin(2\Phi) & -\frac{k_L}{2}\sin(2\Phi) & \cos^2\Phi \end{pmatrix} \tag{2.13}$$

While a higher-order Taylor map for transfer matrix representations exist using Tructed Power Series Algebra for higher-order elements such as sextupoles and octupoles, their treatment is beyond the scope of this thesis. The present work is restricted to the linear beam optics formalism.

2.1.4 Beam Statistics

To characterize the collective behavior of a particle beam, it is useful to compute statistical moments of the particle distribution. This analysis is primarily used to describe a Gaussian bunch beam, but it can also be applied to a non-Gaussian bunch beam. This section introduces the first and second moments, which fully describe a gaussian bunch, and relates them to the standard Courant-Snyder (Twiss) parameters and emittance. Let $x_i \in \{x, x', y, y'\}$

First Moments

The first moments of the distribution represent the beam centroid. This provide the average position and momentum of the particle bunch in phase space, effectively tracking the beam's center of mass.

$$\langle x_n \rangle = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n. \tag{2.14}$$

Second Central Moments

The second central moments measures the variance and covariance of the particle coordinates relative to the beam centroid. These quantify the statistical spread of the beam.

$$< x_j x_k >= \frac{1}{N}\sum_{n=1}^{N}(u_{j,n} - \langle u_j \rangle)(u_{k,n} - \langle u_k \rangle). \tag{2.15}$$

4D Beam Matrix

The moments can be arranged into a 4D beam matrix that describes the beam's state. This describes the transverse statistical properties of the beam. The beam is transversely decoupled when $\sigma_{xy} = \sigma_{yx} = 0$.

$$\sigma^{4D} = \begin{pmatrix} \langle x^2 \rangle & \langle xx' \rangle & \langle xy \rangle & \langle xy' \rangle \\ \langle xx' \rangle & \langle x'^2 \rangle & \langle x'y \rangle & \langle x'y' \rangle \\ \langle xy \rangle & \langle x'y \rangle & \langle y^2 \rangle & \langle yy' \rangle \\ \langle xy' \rangle & \langle x'y' \rangle & \langle yy' \rangle & \langle y'^2 \rangle \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} \tag{2.16}$$

Given the transfer matrix M, the evolution of the beam matrix is given as

$$\sigma_s^{4D} = M \sigma_{s0}^{4D} M^T. \tag{2.17}$$

Emittance

The emittance describes the statistical area occupied by the beam in phase space and can be directly calculated from the beam matrix moments. It is defined either as the projected emittance in a specific plane or as the total emittance derived from the full beam matrix

$$\epsilon_{rms}^x = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2}, \quad \epsilon_{rms}^{total} = \sqrt{\det \sigma^{4D}}. \tag{2.18}$$

According to Liouville's theorem, the geometric emittance ϵ_{rms} is invariant under symplectic transport, provided the beam energy remains constant. Consequently, to compare the emittance at different energies, we define the normalized emittance, which is an invariant quantity during acceleration.

$$\epsilon_{norm} = \beta_r \gamma_r \epsilon_{rms} \tag{2.19}$$

where $\beta_r = v/c$, $\gamma_r = 1/\sqrt{1 - \beta_r^2}$ are the relativistic velocity and Lorentz factor.

Twiss Parameters

Finally, the geometric properties of the beam distribution are described by the Courant-Snyder parameters, commonly referred to as the Twiss parameters. These parameters define the orientation

and shape of the phase space ellipse containing the beam particles. They are related to the statistical moments by

$$\alpha_x = -\frac{\langle xx' \rangle}{\epsilon_{rms}}, \quad \beta_x = \frac{\langle x^2 \rangle}{\epsilon_{rms}}, \quad \gamma_x = \frac{\langle x'^2 \rangle}{\epsilon_{rms}} \tag{2.20}$$

The corresponding values for y are obtained the same way. These parameters are not independent but are related through the geometric identity

$$\beta\gamma - \alpha^2 = 1 \tag{2.21}$$

The position-momentum phase space plot in Fig. 2.2 illustrates the beam's distribution. Intuitively, $\beta(s)$ defines the beam's width, whereas the $\alpha(s)$ describes the "tilt" of the beam.

Since $\alpha \equiv -\frac{1}{2}\beta'$, the sign of α indicates the slope of the beam envelope:

- $\alpha = 0$: The beam is at a waist, the minimum or maximum size.
- $\alpha < 0$: The beam is focusing toward a waist.
- $\alpha > 0$: The beam is defocusing away from a waist.

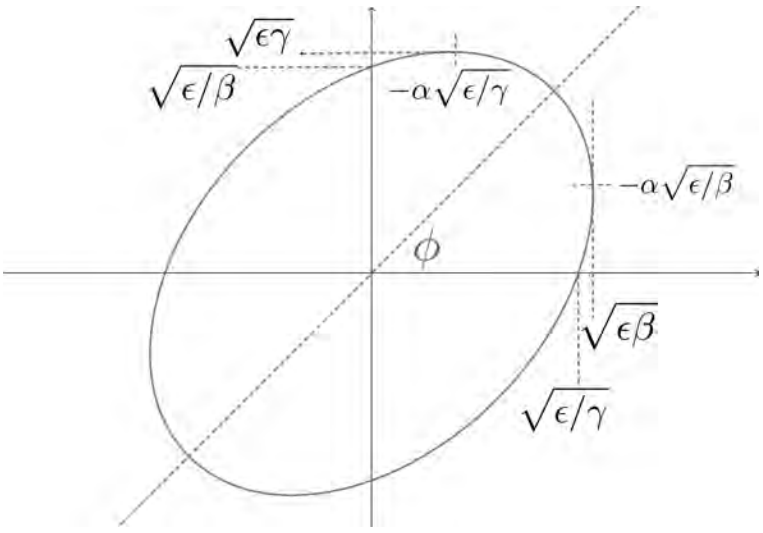


Figure 2.2 Courant Snyder Ellipse [40].

As the beam propagates through the lattice, the Twiss parameters evolve. Given a transfer matrix M, the Twiss parameters transform according to the following matrix relation [36]

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_1 = \begin{pmatrix} M_{11}^2 & -2M_{11}M_{12} & M_{12}^2 \\ -M_{11}M_{21} & M_{11}M_{22} + M_{12}M_{21} & -M_{12}M_{22} \\ M_{21}^2 & -2M_{21}M_{22} & M_{22}^2 \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_0 \tag{2.22}$$

The phase advance, ψ, represents the fractional number of oscillations a particle makes along the beamline. Adopting the convention for linear accelerators, it is given as

$$\psi = \arctan\left(\frac{M_{12}}{\beta_0 M_{11} - \alpha_0 M_{12}}\right). \tag{2.23}$$

In practice, calculating the phase advance requires knowledge of the Twiss parameters at the initial location and along the beamline to the end location.

2.2 Direct Phase-Space Measurements Methods

As noted in the previous chapter, one of the grand challenges is to measure the beam phase space with 1000 bins per dimension within a few hours. This is a large data size of at least 10^{12}, considering a 4-dimensional distribution. A high dynamic range is desired, with the ability to resolve halo particle distribution. Having this would aid beam control, manipulation, and controlled beam loss, as well as the simulation of collective effects. This first section provides an overview of common methods for measuring beam distribution.

System	Space	Mechanism	Sampling
Double-Slit	Transverse	Direct measurement	Scanning
Allison Scanner	Uncoupled Transv.	Slits + electrical sweep	Scanning
Pepper Pot	Full Transv.	Pinhole array + scintillator	Single-shot
Wire Scanner	1D Transv. Spatial	Sweeping wire	Scanning
Gas Sheet	X-Y plane	Beam-gas -> photons	Continuous
TDS	Longitudinal	Transv.-long. coupling	Single-shot
6D Scan	Full 6D	Integrated multi-stage optics	Scanning

Table 2.1 Summary of Phase Space Diagnostic Systems.

2.2.1 Transverse Phase Space Diagnostics

2.2.1.1 Double-Slit Scanner

The conventional method for directly measuring the beam phase space utilizes thin slits to slice the beam, allowing a small beamlet to propagate over a known drift length. As the beam propagates, it naturally diverges due to the inherent momentum distribution within the beamlet. A second slit is placed downstream to select a specific momentum, and a Faraday cup is used to measure the intensity. The slits are mechanically stepped across the beam to map the full transverse phase space. While high-resolution, this technique is time-intensive.

2.2.1.2 Allison Scanner

The Allison Scanner is a device capable of measuring the 2D transverse phase space of the beam in the horizontal and vertical planes. This device is composed of an entrance slit, a parallel plate with a sweeping voltage, an exit slit, and a Faraday cup, as seen in Fig. 2.3.

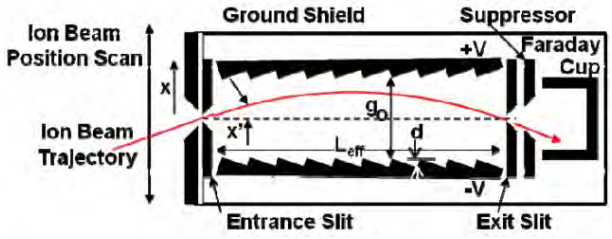


Figure 2.3 Schematic of an Allison Scanner [41].

The position coordinates are selected from the first slit. The voltage is adjusted or swept to choose the particle's angle x', y'. Finally, the intensity is recorded using the Faraday cup.

This is adaptable to a high dynamic range, limited by the Faraday cup's sensitivity, up to the point at which beam damage to the slits becomes significant. Measurements are taken in steps, with about 1-15 minutes required to complete a full scan, depending on the resolution [42].

2.2.1.3 Pepper Pot

A Pepper pot measures the 2D transverse phase space of the beam in the horizontal and vertical directions. According to the schematic shown in Fig. 2.4, the beam first passes through a thin metal sheet "peppered" with holes, splitting into many small beamlets. The holes can be in organized as a

grid, a single line, or points radiating from the center. A grid arrangement would allow cross terms to be calculated, thereby providing the full 4D transverse phase-space distribution. The holes represent the distribution's position coordinates. After passing through the holes, the beamlets propagate over a drift length L and then strike a scintillator (or a microchannel plate) to resolve the beamlets' location and spread as light. A camera then takes a picture. The spread of the beam and the distance L can be used to calculate the particle's momentum x', y'. With this, the full 4D phase space can be estimated.

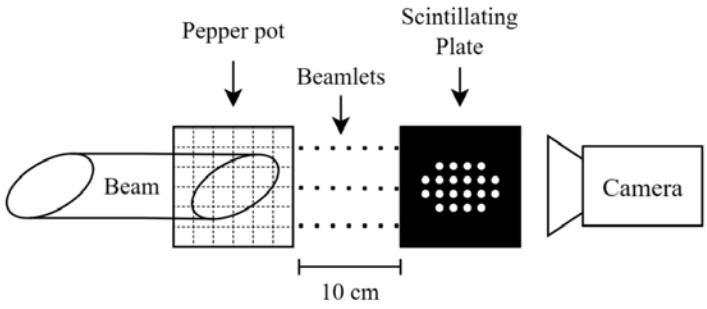


Figure 2.4 Schematic of Pepper Pot [43].

To measure the transverse phase space, only a single measurement is required, while multiple measurements can yield better statistics. Its resolution in the position dimensions (x, y) is limited by the hole spacing. Hole spacing is constrained by beamlet overlap, in which the beamlets at the end overlap with adjacent beams, making it difficult to align the beams with the holes. Parameters such as the drift distance, hole spacing and arrangement, and hole size are optimized to balance resolution and resolvability. The method, Virtual Pepper Pot [44], uses existing double-slit scanners to virtually generate data from a pepper pot, enabling a broader range of beams to be measured.

2.2.2 Wire Scanner

Although it cannot measure the transverse phase space directly, a wire scanner is a common diagnostic used to extract beam profiles in an accelerator.

By sweeping a single wire through the beam, it reduces cost and avoids the resolution limits imposed by the fixed spacing of multi-wire arrays. The wire is designed to withstand high temperatures without melting and can reach scanning velocities of up to 10 m/s using dedicated mechanics.

Because the measurement is not instantaneous, a steady-state beam distribution is required [45].

2.2.2.1 Gas Sheet Monitor

As beam power increases, intercepting diagnostics become infeasible due to the risk of thermal damage. Compared to a wire scanner, which measures a 1D projection of the beam, the wire scanner can still fail catastrophically under high-intensity conditions. A Gas Sheet Monitor addresses these limitations by utilizing a non-invasive interaction between the beam and a rarefied gas curtain. The beam particles excite the gas molecules, inducing fluorescence, which is imaged by a camera. Unlike solid targets, high beam power is advantageous here, as it increases the photon yield and improves the signal-to-noise ratio

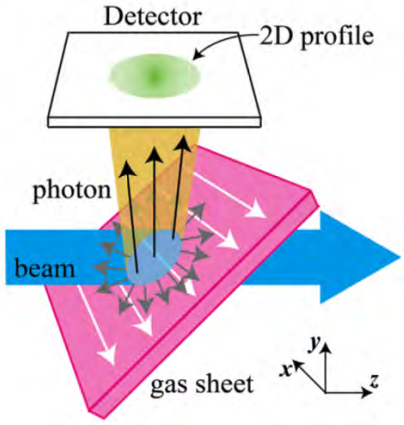


Figure 2.5 Schematic of a gas sheet bream profile monitor [46].

As seen in Fig.2.5, a thin gas sheet is injected into the beam pipe. Instead of filling the entire chamber, the gas flows across the beam path as a thin, directed stream. This gas is immediately pumped out to prevent disruption of the vacuum. The beam-gas geometry is oriented so that the measured photon distribution corresponds directly to the beam profile, enabling direct imaging.

2.2.3 Longitudinal Phase Space Diagnostics

2.2.3.1 Transverse Deflecting Structures

Transverse Diagnostics are essential for matching the beam distribution to the accelerator design. In high-gain Free-Electron Lasers (FELs), optimization of longitudinally compressed bunches affects the output light, thereby motivating the need for longitudinal diagnostics. Trans-

verse deflecting structures (TDS) are used to make time-resolved measurements. Short electron bunches, are on the order of femtoseconds [47], and these structures are able to provide strong coupling between the longitudinal and transverse dimensions, thus allowing measurement of the transverse image to calculate the longitudinal profile.

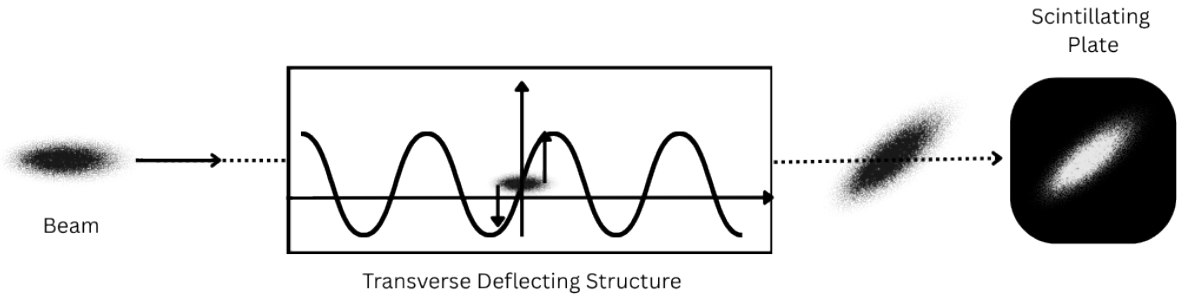


Figure 2.6 Schematic of A Transverse Deflecting Structure.

As seen in Fig. 2.6, a beam enters a TDS. The head of the beam is affected by a specific electric field in one direction, and the electric field varies rapidly enough that the tail of the beam receives the opposite force. This effectively streaks the beam. The beam is then measured, and the longitudinal profile can be extracted.

In a bit more detail, a dipole mode cavity is used. The dipole mode creates a longitudinal electric field that varies linearly with the transverse distance from the axis. A time-dependent magnetic deflecting field is also present. Given a state-vector $\mathbf{x} = (x, x', z, \delta)$, the "thin-lens" approximate of the cavity looks like [48]

$$\mathbf{M}_{\text{cavity}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & k & 0 \\ 0 & 0 & 1 & 0 \\ k & 0 & 0 & 1 \end{bmatrix}, \tag{2.24}$$

where $k = eV_0/aE$, and V_0 is the peak voltage and a is a constant characteristic of the cavity. Notice that x' will depend on the z coordinate, allowing one to encode the longitudinal profile into the transverse profile.

2.2.4 First 6D measurement of the Phase Space

The first direct 6D measurement of a beam distribution was achieved in 2018 at the SNS Beam Test Facility [23]. This is the transverse phase space (x, x', y, y') and the (w, φ) is the energy and phase relative to a design reference particle. This experiment utilized a sequence of collimating slits, bending magnets, and RF structures to isolate a specific hyper-volume in the 6D phase space.

To perform the scan, the system utilized upstream slits to define the transverse position $(x \pm \Delta x, y \pm \Delta y)$, followed by a drift and secondary slits to select the divergence $(x' \pm \Delta x', y' \pm \Delta y')$. A dipole magnet coupled with a slit selects the energy $w \pm \Delta w$. Different energies will bend differently in a dipole and are described by

$$M_{\text{dipole}} = \begin{bmatrix} \cos\theta & \rho\sin\theta & \rho(1-\cos\theta) \\ -\frac{1}{\rho}\sin\theta & \cos\theta & \sin\theta \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.25}$$

where ρ is the bending radius, and θ is the bend angle, and $\mathbf{x} = (D, D', 1)$ is the state-vector. The dispersion function D describes the transverse positional shift of a particle due to a deviation from the reference momentum. Consequently, a mechanical slit placed at a specific x coordinate in a dispersive region can be used to select a specific momentum slice of the beam. The energy term comes in through the bending radius from the beam rigidity equation $B\rho = p/q$, since the momentum p relates to energy.

An RF deflector is used to link the phase to the beams' arrival time. This is described similarly to the method shown in the previous section. An RF field is used to impart a time-dependent streak to the beam, allowing a final slit to select the phase $\varphi \pm \Delta\varphi$.

The final particles are collected in a Faraday cup. A scan of 5,675,740 points took 32 hours. The number of particles collected is small, with approximately 1 in 10 million collected at the beam core. The results revealed a previously unmeasured 'hidden' correlation within the beam core, attributable to space-charge forces, demonstrating that the conventional assumption of uncoupled phase space dimensions is invalid for high-intensity beams.

2.3 Indirect Phase-Space Measurements Methods

Direct measurement techniques, as detailed in the previous section, remain the standard for high-fidelity phase-space characterization due to their accuracy and model-independence. However, their limitations include being typically invasive, time-consuming, and requiring dedicated diagnostic lines that consume valuable beamline space. In contrast, indirect measurement methods leverage existing beamline components, such as quadrupole magnets and profile monitors, to reconstruct the full phase space distribution from a set of lower-dimensional projections. While these methods often rely on linear optics models and reconstruction algorithms, they offer the advantage of being non-invasive and compatible with existing operations.

To standardize the notation in this section, let the probability distribution be $f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^N$. Oftentimes, indirect measurement methods require taking measurements downstream of the reconstruction location. Let $M_k : \mathbb{R}^N \to \mathbb{R}^N$ be the linear transfer map representing the beamline between the reconstruction and measurement location for the k-th measurement.

For this chapter, let $N = 2$. The transformed coordinates $\mathbf{u}_k = M_k\mathbf{x}$, propagated through matrix M,

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix}. \tag{2.26}$$

describes the evolution of a beam distribution through a linear model of a beamline. After the transport, the measured projection $p_k(u)$, given in Eq. 2.27, is the integral of the phase space density over v. This is shown in Fig. 2.7.

$$p_k(u) = \int_{-\infty}^{\infty} f\left(M_k^{-1} \begin{bmatrix} u \\ v \end{bmatrix} \right) dv, \quad k = 1, 2 \dots K \tag{2.27}$$

Tomography is a technique to reconstruct a high dimensional object from its lower dimensional projections. A critical step in phase space tomography involves mapping the downstream measurements back to the reconstruction location. Standard tomographic reconstruction algorithms assume that all projections are derived from a static object. In a beamline, the beam distribution evolves as

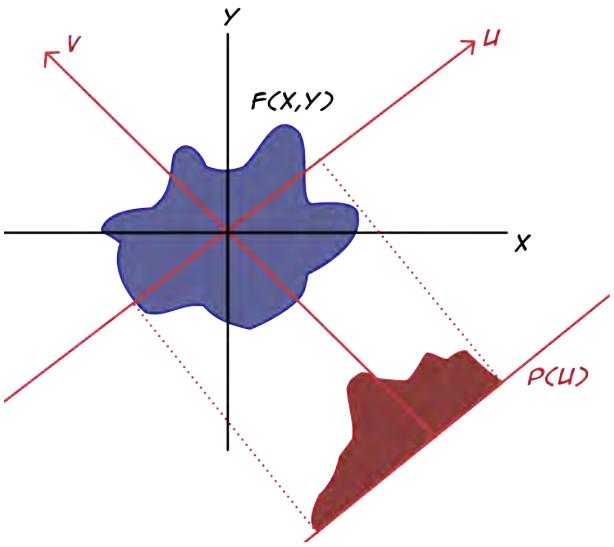


Figure 2.7 Projections $p(u)$ depends on the angle on which the projection was taken.

it propagates through magnetic elements. Thus, measurements taken at different locations or magnet settings represent projections of physically different distributions. A mapping can be derived from the symplectic nature of the beam transport, which ensures the conservation of phase-space density. The derivation of this relationship is detailed in [49] and is given by

$$\tan\theta = \frac{M_{12}}{M_{11}}, \quad a = \sqrt{M_{11}^2 + M_{12}^2}, \quad u = \frac{x}{a}, \quad p_A = a p_B, \tag{2.28}$$

where A represent the reconstruction location and B represents the measurement location. These equations determine the effective rotation angle θ and a scaling factor a, thereby allowing the downstream measurement to be mapped to a specific, scaled projection of the initial phase-space distribution. This set of scaled projections, p_k, can then be used with traditional tomography algorithms to reconstruct the initial distribution.

The following sections will describe 4 classical algorithms, Linear Least Squares (LLSQ), Filter Back Projection (FBP), Simultaneous Algebraic Reconstruction Technique (SART), and Maximum Entropy Reconstruction Technique (MENT), which can be employed to reconstruct the beam distribution from its projections.

2.3.1 Reconstruction in the Normalized Frame

It has been demonstrated that performing reconstruction in the normalized phase-space frame yields improved results [50]. In this normalized frame, a standard Gaussian beam distribution becomes rotationally symmetric, and the phase advance reduces to a simple angle of rotation.

Extending the reconstruction to normalized transverse phase space is straightforward but requires an initial estimate of the Twiss parameters, α and β, at the source. These can be obtained, for example, using the quadrupole scan method.

In the standard formulation, the coordinates at the measurement location are related to the reconstruction location via the transfer matrix M. To adapt this for normalized phase space, we introduce the transformation from normalized coordinates (x_N, x'_N) to real coordinates (x, x')

$$\begin{pmatrix} x \\ x' \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} & 0 \\ -\frac{\alpha}{\sqrt{\beta}} & \frac{1}{\sqrt{\beta}} \end{pmatrix} \begin{pmatrix} x_N \\ x'_N \end{pmatrix}. \tag{2.29}$$

Substituting Eq. 2.29 into the matrix equation Eq. 2.26 yields the mapping from the normalized frame to the real frame

$$\begin{pmatrix} u \\ v \end{pmatrix} = R \begin{pmatrix} x_N \\ x'_N \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} x_N \\ x'_N \end{pmatrix}, \tag{2.30}$$

where the effective transfer matrix R is defined as

$$R = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} \sqrt{\beta} & 0 \\ -\frac{\alpha}{\sqrt{\beta}} & \frac{1}{\sqrt{\beta}} \end{pmatrix}. \tag{2.31}$$

By simply replacing the matrix M with R, the preceding algorithm directly reconstructs in the normalized phase space using the same set of measured projections.

2.3.2 Linear Least Squares

A widely used technique for emittance estimation is the quadrupole scan. In this method, the magnetic strength of a quadrupole is varied, and the resulting beam profile is measured by a wire scanner after a drift section. These measurements are then used to estimate the initial Twiss

parameters and emittance, typically under the assumption of a Gaussian beam distribution. The Linear Least Squares (LLSQ) method generalizes this algorithm.

The statistical moments of the beam distribution are defined by Eq. 2.16, from which the Twiss parameters and emittance can be derived. As shown previously, the evolution of the beam matrix is governed by

$$\sigma_s^{4D} = M_k \sigma_{s0}^{4D} M_k^T, \quad k = 1, 2 \dots K. \tag{2.32}$$

Assuming the initial distribution, σ_{s0}^{4D}, remains unchanged during the scan, partial measurements of the final beam matrix, σ_s^{4D}, can be obtained for various transfer matrices. These measurements form a system of equations, $\mathbf{b} = A\mathbf{x}$, that can be solved to determine the unknown initial beam matrix, σ_{s0}^{4D}.

Certain combinations of measurements and beam manipulation can make this solvable or unsolvable. First, a specific case will be considered, then the conditions for solving this system in general.

Let us consider the specific case where we can only obtain the beam sizes $< x^2 >_s$ and $< y^2 >_s$. This is the standard output from common diagnostics, such as wire scanners. In this setup, the beam transport line consists solely of drift spaces and normal quadrupoles. Consequently, the transfer matrix can only be manipulated along its diagonal blocks. Because coupling cannot be measured or influenced on this beamline, we are forced to assume that the initial beam distribution is uncoupled. It should be noted that this is not always the case in a real machine.

To create the system of equations, we now expand Eq. 2.32 as a triple product.

$$\sigma_{i,j}^s = \sum_{p=1}^{4} \sum_{q=1}^{4} M_{i,p} \sigma_{p,q}^{s0} M_{j,q} \tag{2.33}$$

Our beam size measurement gives the following equations.

$$\sigma^s_{1,1} = \langle x^2 \rangle_s = \sum_{p=1}^{4}\sum_{q=1}^{4} M_{1,p}\sigma^{s0}_{p,q}M_{1,q} \tag{2.34a}$$

$$\sigma^s_{3,3} = \langle y^2 \rangle_s = \sum_{p=1}^{4}\sum_{q=1}^{4} M_{3,p}\sigma^{s0}_{p,q}M_{3,q} \tag{2.34b}$$

Upon expanding, we get

$$\langle x^2 \rangle_s = M^2_{1,1}\sigma^{s0}_{1,1} + 2M_{1,1}M_{1,2}\sigma^{s0}_{1,2} + M^2_{1,2}\sigma^{s0}_{2,2} \tag{2.35a}$$

$$\langle y^2 \rangle_s = M^2_{3,3}\sigma^{s0}_{3,3} + 2M_{3,3}M_{3,4}\sigma^{s0}_{3,4} + M^2_{3,4}\sigma^{s0}_{4,4} \tag{2.35b}$$

To solve for the initial beam parameters, we rearrange the expansion derived previously into a linear system of the form $\mathbf{b} = A\mathbf{x}$. The vector of unknowns, x, contains the independent elements of the initial beam matrix at location s_0.

$$\mathbf{x} = (< x^2 >_{s0}, < xx' >_{s0}, < x'^2 >_{s0})^T \tag{2.36}$$

The measurement vector, b, consists of the squared beam sizes measured at the downstream location s across k different quadrupoles.

$$\mathbf{b} = (< x^2 >^1_s, < x^2 >^2_s, \ldots, < x^2 >^k_s)^T \tag{2.37}$$

The design matrix, A, is constructed from the transfer matrix elements corresponding to each measurement k based on the expansion in Eq. 2.35.

$$A = \begin{bmatrix} (M^{(1)}_{1,1})^2 & 2M^{(1)}_{1,1}M^{(1)}_{1,2} & (M^{(1)}_{1,2})^2 \\ (M^{(2)}_{1,1})^2 & 2M^{(2)}_{1,1}M^{(2)}_{1,2} & (M^{(2)}_{1,2})^2 \\ \vdots & \vdots & \vdots \\ (M^{(k)}_{1,1})^2 & 2M^{(k)}_{1,1}M^{(k)}_{1,2} & (M^{(k)}_{1,2})^2 \end{bmatrix} \tag{2.38}$$

The index k in $\langle\cdot\rangle^k$ and $M_{i,j}^{(k)}$ denotes the beam's measured second moment and transfer matrix taken from the k^{th} measurement. While this derivation focuses on the horizontal (x, x') plane, an identical formulation applies to the vertical (y, y') plane using indices 3 and 4.

A minimum of 3 measurements is needed to solve this system of equations. Multiple measurements are usually taken to give an overdetermined system, which can be solved using the pseudo-inverse

$$A^+ = (A^T A)^\dagger A^T. \tag{2.39}$$

This will minimize the residual $\|\mathbf{r}\|_2^2 = \|A\mathbf{x} - \mathbf{b}\|_2^2$. In computational implementations, such as Python's `numpy.linalg.lstsq`, this system of equations is typically solved.

In the general case, the beam may be coupled and cannot be correctly solved using the current derivation. From Eq. 2.34, all elements of σ^{s0} are being used for any calculations of σ^s, thus to solve the full coupled system, either M must have non-zero off diagonal terms, coupling beam elements such as skew-quadrupoles or solenoids, or an additional measurement of coupling between the two planes is needed.

2.3.3 Filter Back Projection

Tomography is the field of study of algorithms that reconstruct a high-dimensional object from its lower-dimensional projections. Using the results of Eq. 2.28, tomography algorithms are useful and can be applied in accelerator physics to reconstruct phase space.

The most commonly used tomography algorithm is the Filtered Back-Projection (FBP) algorithm. It is a Fourier-domain reconstruction algorithm enabling the reconstruction of a 2D image from 1D projections. It is based on the **Fourier Slice Theorem**, which states that the following two are equal:

- The Fourier transform of a one-dimensional projection of a two-dimensional function.
- The two-dimensional Fourier transform of a slice through the origin, taken parallel to the projection direction.

Proof of Fourier Slice Theorem in 2D [51]

Let $f(x, y)$ be a two-dimensional function. The 2D Fourier transform of $f(x, y)$ is $\tilde{f}(k_x, k_y)$, given by

$$\tilde{f}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i(xk_x + yk_y)} \, dx \, dy. \tag{2.40}$$

Now, considering the (t, s) coordinate system, which is the original coordinate system rotated by angle θ

$$\begin{bmatrix} t \\ s \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \tag{2.41}$$

The projection along a line perpendicular to t is

$$P_\theta(t) = \int_{-\infty}^{\infty} f(t, s) \, ds. \tag{2.42}$$

Its Fourier transform of the projection at angle θ is

$$\tilde{P}_\theta(w) = \int_{-\infty}^{\infty} P_\theta(t) e^{-i2\pi wt} \, dt. \tag{2.43}$$

We will now show this is equal to Eq. 2.40. Using Eq. 2.42 and 2.43 we obtain

$$\tilde{P}_\theta(w) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t, s) \, ds \right] e^{-i2\pi wt} \, dt. \tag{2.44}$$

We now change the variables of integration from the rotated frame (t, s) back to the original frame (x, y). Since the transformation is a pure rotation, the Jacobian determinant is unity. Substituting $t = x\cos\theta + y\sin\theta$ results in

$$\tilde{P}_\theta(w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi w(x\cos\theta + y\sin\theta)} \, dx \, dy. \tag{2.45}$$

Comparing this to Eq. 2.40, we recognize that given the substitution of $k_x = w\cos\theta$ and $k_y = w\sin\theta$, the Fourier transform of the projection at angle θ is the same as the slice of its 2D

Fourier transform taken at the same angle through the origin. A visual example from a computer simulation is shown in Fig. 2.8.

$$\tilde{P}_\theta(w) = \tilde{f}(w\cos\theta, w\sin\theta) \tag{2.46}$$

□

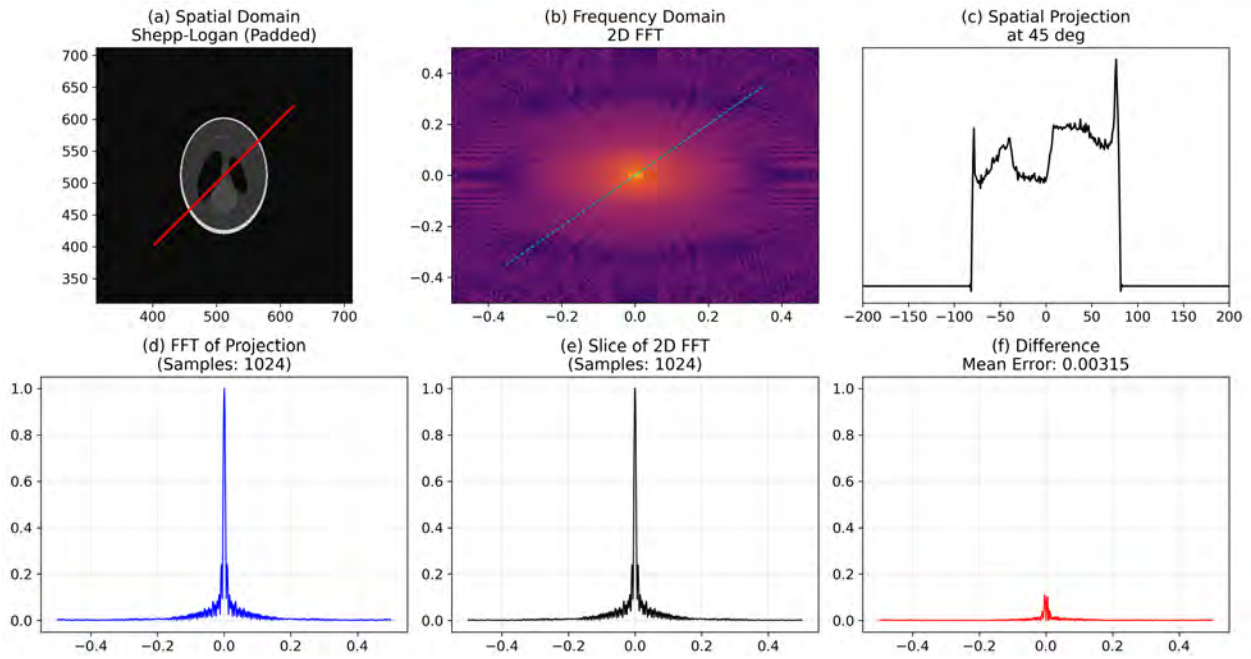


Figure 2.8 Verification of the Fourier Slice Theorem using the Shepp-Logan phantom. Top row: (a) The 2D test image, (b) its 2D Fourier Transform with a 45° slice indicated, and (c) the 1D spatial projection at 45°. Bottom row: (d) The 1D FFT of the projection, (e) the corresponding slice extracted from the 2D FFT, and (f) the difference between the two, demonstrating their equivalence.

The algorithm is implemented in Python using the scikit-image package [52]. As a consequence of this theorem, it follows that if one has an infinite number of projections from all angles, the 2D Fourier transform converges to a unique solution.

2.3.4 Simultaneous Algebraic Reconstruction Technique

A brute-force method is used to generate algebraic equations that link the unknown distribution to the measured projection data. As shown in Fig. 2.9, a grid is superimposed on the unknown image $f(x, y)$, and equations relating the projections to the measurements are described.

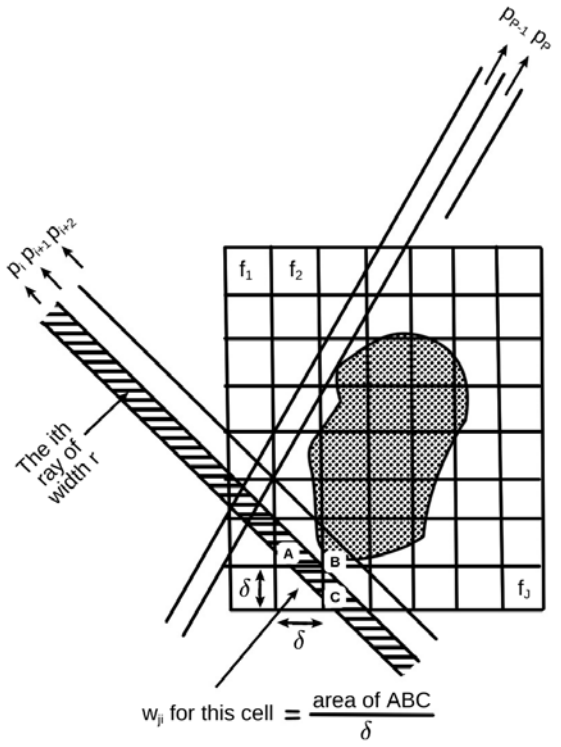


Figure 2.9 Algebraic Reconstruction grid over an unknown target image based on [51].

The value within each cell j is a constant value equal to the mean value of the function $f(x, y)$ in that cell. It is denoted by f_j, for a total of all J cells used to enclose the distribution. When a projection is taken, a ray, modeled as a “fat line" with finite width, passes through the grid. The weight w_{ji} represents the contribution of the j-th cell to the i-th ray integral, typically calculated as the fractional area of the cell intersected by the ray. The relationship between the cell values f_j and the measured projection p_i is given by:

$$\sum_{j=1}^{J} w_{ij} f_j = p_i, \quad i = 1, 2, \ldots, P \tag{2.47}$$

where P is the total number of line integrals across all bins for all measurements. The number of equations grows as $O(N^3)$, necessitating the use of iterative methods.

The Kaczmarz method, illustrated in Fig. 2.10, is an iterative solver often used for this purpose. First, the problem, visualized geometrically. An image with J cells possesses J degrees of freedom and can be considered a single point in an J-dimensional space. Each linear equation in Eq.

2.47 defines a hyperplane in this space. A unique solution exists at the intersection of all these hyperplanes.

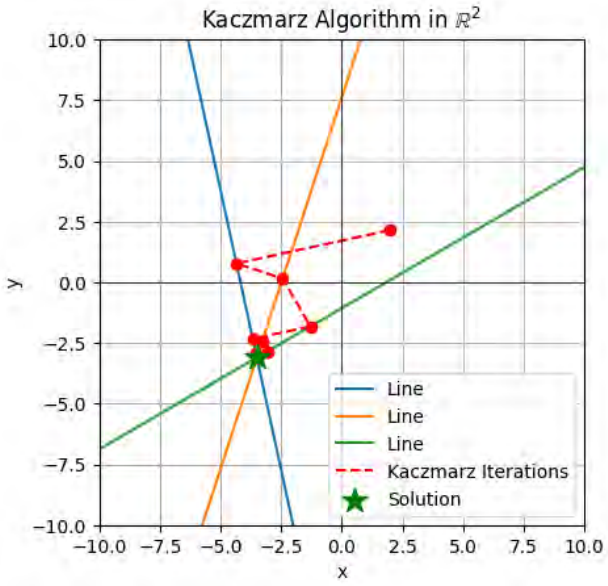


Figure 2.10 An illustration of the Kaczmarz algorithm [53].

Start with an arbitrary initial guess and project this point onto the hyperplane defined by the first equation. Then projects onto the next hyperplane and continues until all equations are satisfied. Continue repeating this whole process until convergence is reached. Several observations can be made regarding the optimization of this method:

1. **Convergence Speed:** If the hyperplanes are all orthogonal, the method arrives at the correct solution in exactly J steps. Conversely, if the angle between hyperplanes is small, many more iterations are required. This implies that projections taken at widely separated angles converge faster than those taken at clustered angles.

2. **Noise Sensitivity:** In an overdetermined system with measurement noise, the hyperplanes do not intersect at a single point. Consequently, the solution fails to converge to a stationary point and instead oscillates near the intersections.

3. **A Priori Information:** It is straightforward to incorporate prior knowledge by enforcing constraints during each iteration, such as non-negativity or setting values to zero outside a

known support region.

The Algebraic Reconstruction Technique (ART) and the Simultaneous Algebraic Reconstruction Technique (SART) are two specific algorithms derived from this framework. In standard ART, the image is updated ray-by-ray. However, to speed up computation, weights w_{ji} are sometimes simplified to binary 1s and 0s depending on whether the cell center lies within the ray. This approximation can introduce “salt-and-pepper" noise into the reconstruction due to the induced inconsistencies.

SART is a refined version of ART [54] which improves upon the algorithm by:

1. **Simultaneous Correction:** Instead of updating the image after every ray, SART computes the correction terms for all rays in a given projection simultaneously and applies their average. This acts as a smoothing filter, reducing noise.

2. **Hamming Window:** A longitudinal Hamming window, a bell-shaped curve, is applied to emphasize corrections near the middle of a ray compared to its ends, since calculations are most reliable in the center compared to the very edges.

3. **Bilinear Elements:** It uses bilinear basis functions, 3D structures placed on a 2D grid to linearly interpolate values, rather than simple square pixels to provide a better discrete approximation of the continuous image.

Similar to FBP, the SART algorithm can be implemented in Python using the `scikit-image` package [52].

2.3.5 Maximum Entropy Tomography

The Maximum Entropy Tomography (MENT) algorithm, introduced in 1978 [55], employs maximum-entropy techniques to reconstruct accelerator beam phase-space distributions. The method was motivated by the need to obtain reliable reconstructions from a limited number of views, typically fewer than 10. In accelerator physics, this constraint is critical, as wire-scanner measurements

can be time-intensive and the practicality of phase-space reconstruction depends on rapid data acquisition and analysis.

The derivation solves for the distribution that maximizes entropy subject to the constraints imposed by the measured projections.

$$H(f) = -\iint f(\mathbf{x}) \ln f(\mathbf{x})\, d\mathbf{x} \tag{2.48}$$

First, the entropy can be derived by counting possible microstates and applying Stirling's approximation as given by Eq. (2.48). Since a unique solution does not exist under the projection constraints alone, entropy is used to select the most statistically plausible distribution. If no constraints are imposed, the function $f(\mathbf{x})$ that maximizes entropy is the uniform distribution (or $f_i = 1/J$ in the discrete case, where J is the total number of grid cells). This assumption is reasonable for accelerator beams when no other prior information is available. It is possible to incorporate alternative priors to improve reconstruction accuracy, but this is a topic for future research.

Let $\hat{p}_k$ and $\hat{f}_k(\mathbf{x})$ represent our simulated projection and reconstructed distribution. The simulated projection $\hat{p}_k$ is defined in a similar way as Eq. 2.27.

$$\hat{p}_k(u_k) = \int_{-\infty}^{\infty} \hat{f}\left(M_k^{-1} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right) dv_k, \quad k = 0, 1 \ldots K \tag{2.49}$$

Our constraint equation enforces that our simulated projection matches our measurement projections for all u and k

$$\hat{p}_k(u) - p_k(u) = 0. \tag{2.50}$$

The method of Lagrange multipliers is used to combine both the constraint equation and entropy. The Lagrangian functional is defined as

$$\Psi(\hat{f}, \lambda) = H(\hat{f}) + \sum_{k=0}^{K} \int du_k\, \lambda_k(u_k) \left[\hat{p}_k(u_k) - p_k(u_k)\right] \tag{2.51}$$

where the integral and sum are used to ensure the constraint is enforced for all K measurements and all values of u. Here, the method of Lagrange multipliers is extended from scalar variables to functions (see Appendix A.1). We analyze the variation of the functional by considering a small perturbation of the multiplier:

$$\lambda_k(u_k) \rightarrow \lambda_k(u_k) + \epsilon\, \eta_k(u_k) \tag{2.52}$$

where $\epsilon \rightarrow 0$ is a scalar magnitude and $\eta_k(u)$ is an arbitrary smooth function. The condition for a stationary point is that the total variation of the functional Ψ must vanish:

$$\partial\Psi = \partial_{\hat{f}}\Psi + \sum_{k=1}^{K} \partial_{\lambda_k}\Psi = 0 \tag{2.53}$$

Since the variations with respect to f and λ_k are independent, this implies the simultaneous conditions:

$$\frac{\partial\Psi}{\partial\lambda_k(u_k)} = 0 \quad \forall k, \quad \text{and} \quad \frac{\partial\Psi}{\partial\hat{f}(\mathbf{x})} = 0. \tag{2.54}$$

Substituting Eq. 2.52 into the first condition and noting that $\partial\hat{f}(\mathbf{x})/\partial\lambda_k = 0$.

$$\frac{\partial\Psi}{\partial\lambda_k(u_k)} = \epsilon \int du_k\, \eta_k\, [\hat{p}_k(u_k) - p_k(u_k)] = 0. \tag{2.55}$$

Since η_k is an arbitrary function, this enforces a point-wise constraint for all u, giving the individual constraint conditions

$$\hat{p}_k(u_k) - p_k(u_k) = 0. \tag{2.56}$$

For the second condition, for $\frac{\partial\Psi}{\partial\hat{f}} = 0$, we get

$$\frac{\partial\Psi}{\partial\hat{f}(\mathbf{x})} = \partial_{\hat{f}} H(\hat{f}) + \partial_{\hat{f}} \sum_{k=0}^{K} \int du_k\, \lambda_k(u_k)\, [\hat{p}_k(u_k) - p_k(u_k)] = 0 \tag{2.57}$$

For the first term, we would get

$$-\frac{\partial}{\partial \hat{f}(\mathbf{x})} \iint \hat{f}(\mathbf{x}) \ln \hat{f}(\mathbf{x})\, d\mathbf{x} = -[1 + \ln \hat{f}(\mathbf{x})]. \tag{2.58}$$

For the second term, since $\hat{p}_k$ depends on $\hat{f}$ as well, the integral is undone to get

$$\frac{\partial}{\partial \hat{f}(\mathbf{x})} \sum_{k=0}^{K} \int du_k\, \lambda_k(u_k) \left[\int dv_k \hat{f} \left(M_k^{-1} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right) - p_k(u_k) \right]. \tag{2.59}$$

Since $p_k(u_k)$ is a constant, $\frac{\partial p_k(u_k)}{\partial \hat{f}} = 0$.

$$\frac{\partial}{\partial \hat{f}(\mathbf{x})} \sum_{k=0}^{K} \iint du_k dv_k\, \lambda_k(u_k) \hat{f} \left(M_k^{-1} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right). \tag{2.60}$$

Before this integral can be resolved, notice the hidden dependence of (u_k, v_k) in $\hat{f}(\mathbf{x})$. A change of variable needs to be done to convert $du_k dv_k$ to $d\mathbf{x}$, which leads to a Jacobian term J_k. This allows us to bring the integrand under both integrals.

$$\frac{\partial}{\partial \hat{f}(\mathbf{x})} \sum_{k=0}^{K} \iint d\mathbf{x} J_k\, \lambda_k(u_k(\mathbf{x})) \hat{f} \left(M_k^{-1} \begin{bmatrix} u_k(\mathbf{x}) \\ v_k(\mathbf{x}) \end{bmatrix} \right) = \sum_{k=0}^{K} J_k\, \lambda_k(u_k(\mathbf{x})) \tag{2.61}$$

Combining the two components, we obtain

$$\frac{\partial \Psi}{\hat{f}(\mathbf{x})} = -[1 + \ln \hat{f}(\mathbf{x})] + \sum_{k=0}^{K} J_k\, \lambda_k(u_k(\mathbf{x})) \; = 0. \tag{2.62}$$

Rearranging the equations, we obtain

$$\ln f(\mathbf{x}) = -1 + \sum_{k=1}^{K} J_k \lambda_k(u_k(\mathbf{x})) \tag{2.63}$$

To solve for the unknown $\lambda_k(u_k)$, we absorb the constant terms into unknown "Lagrange factors" or "h-functions," defined as $h_k(u_k) = \exp(J_k \lambda_k(u_k) - 1/K)$, to obtain the product form

$$\hat{f}(\mathbf{x}) = \prod_{k=1}^{K} h_k[u_k(\mathbf{x})]. \tag{2.64}$$

Thus, our distribution is a product of Lagrange factors. This solution is plugged back into the constraint equations to obtain K coupled nonlinear integral equations. No general closed-form solution exists; an iterative approach is required. Minerbo noted that $h_{k'}(u_{k'})$ can be factored out of its own constraint integral [55]

$$p_{k'}(u_{k'}) = h_{k'}(u_{k'}) \int d\mathbf{x} \prod_{k \neq k'} h_k[u_k(\mathbf{x})]. \tag{2.65}$$

This allows $h_k(u_k)$ to be expressed as:

$$h_{k'}(u_{k'}) = \frac{p_{k'}(u_{k'})}{\int d\mathbf{x} \prod_{k \neq k'} h_k[s_k(\mathbf{x})]} \tag{2.66}$$

In practice, direct inversion is unstable. A Gauss-Seidel relaxation method is used to determine the factors iteratively. For the k-th view at iteration $m+1$, we use the most recent updates for all other views:

$$h_{k'}^{(m+1)}(u_{k'}) = \frac{p_{k'}(u_{k'})}{\int d\mathbf{x} \left(\prod_{k=0}^{k'-1} h_k^{(m+1)}[u_k(\mathbf{x})] \cdot \prod_{k=k'+1}^{K} h_k^{(m)}[u_k(\mathbf{x})] \right)} \tag{2.67}$$

An under-relaxation method may also be applied to improve convergence stability.

$$h_{k'}^{(m+1)}(u_{k'}) = h_{k'}^{(m)}(u_{k'}) + \omega \left(\tilde{h}_{k'}^{(m+1)}(u_{k'}) - h_{k'}^{(m)}(u_{k'}) \right) \tag{2.68}$$

where $\tilde{h}_{k'}^{(m+1)}$ is the value calculated from Eq. 2.67 and ω is the relaxation parameter. The relaxation parameter can be adjusted to values less than one for better convergence.

2.4 Data-Driven Phase-Space Measurements Methods

Despite advancements in traditional tomography algorithms, challenges with computational speed, data acquisition, and limited computer memory remain. To address these constraints, recent research has leveraged machine learning to approximate phase-space reconstruction. By utilizing extensive training data to establish strong prior models or trained neural networks, these data-driven approaches can yield accurate results within an hour.

This chapter reviews three advances in the data-driven direction. For each study, the discussion will outline the motivation, methodology, and limitations.

2.4.1 Image Compression

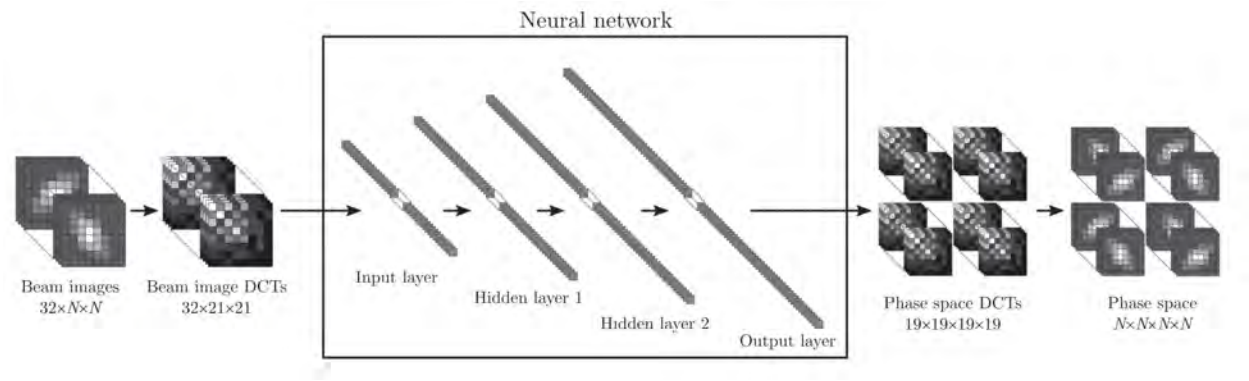


Figure 2.11 Visualization of using image compression for tomography [56].

Conventional tomography reconstruction methods face challenges, including artifacts, computational time, resource constraints, and algorithmic implementation details. As an example of compute time, the Algebraic Reconstruction Technique (ART) algorithm, a grid-based algorithm, scale like $O(N^4)$ for a 4-dimensional reconstruction. A resolution of 39 pixels per dimension, for example, would take a few minutes on a standard PC.

This algorithm was developed to improve speed, yielding results comparable to or superior to those obtained without the algorithm. It used a combination of image compression and ML to reconnstruct the 4D transverse phase space. As seen in Fig. 2.11, the algorithm is as follows:

1. Given an initial set of 2D images (x,y) of some high resolution, the image is compressed to 21×21 using a Discrete Cosine Transformation (DCT).

2. The resulting images are then used as the input layer for a neural network.

3. The output is the DCT of the 4D phase space.

4. The inverse DCT allows reconstruction of the entire 4D phase-space density at any resolution.

This was verified and compared with CLARA experimental data, yielding results comparable to those from ART. The dataset was trained on 3000 4D phase-space distributions and their corresponding sinograms, with 100 sets reserved for validation. Training takes several minutes on

a standard laptop, comparable to the time required to process a single dataset using ART; once trained, the neural network can, in principle, be applied to any dataset collected under the same quadrupole settings. A practical advantage is the ease of setup and the avoidance of complications associated with conventional tomography techniques, such as sinogram formatting.

A limitation of this method is the need for quality training data. It is trained on sets drawn from a Gaussian distribution, but tested on sets drawn from a realistic distribution; the results are far from ideal.

2.4.2 Generative Phase Space Reconstruction

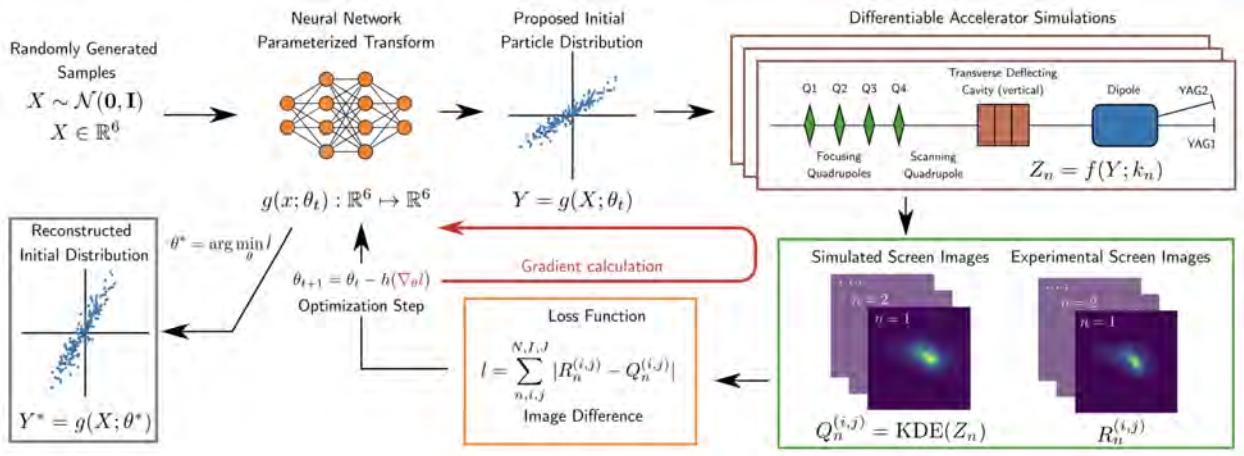


Figure 2.12 Visualization of the GPSR algorithm [33].

Generative Phase Space Reconstruction (GPSR) [33] is a novel technique that utilizes generative machine learning and differentiable beam-dynamics simulations to reconstruct detailed 6D phase-space distributions.

GPSR can resolve 6D phase space from 20 measurements in under 20 minutes computationally. It employs a generative approach similar to methods used in image generation and protein folding. By leveraging neural networks' capacity as universal function approximators, GPSR can generate particle distributions with nearly arbitrary structures in 6D. The reconstruction is formulated as a parameterization problem with approximately 1,000 free parameters. As shown in Fig. 2.12, a neural network parameterizes a randomly generated source distribution to minimize the loss between measured and simulated data. A differentiable accelerator simulation efficiently computes

gradients, enabling gradient-based optimization.

A significant advantage of this particle-based approach is that it avoids the scaling issues inherent in standard mesh-based methods. Additionally, it allows integration of data from multiple, potentially heterogeneous sources by directly incorporating them into the loss function.

However, the algorithm has limitations. It currently struggles to predict halo-beam distributions, even when the halo is approximated as Gaussian and while it performs well with nonlinear distributions, it struggles to reconstruct custom distributions with hard edges.

2.4.3 Conditional Diffusion Variational Autoencoder

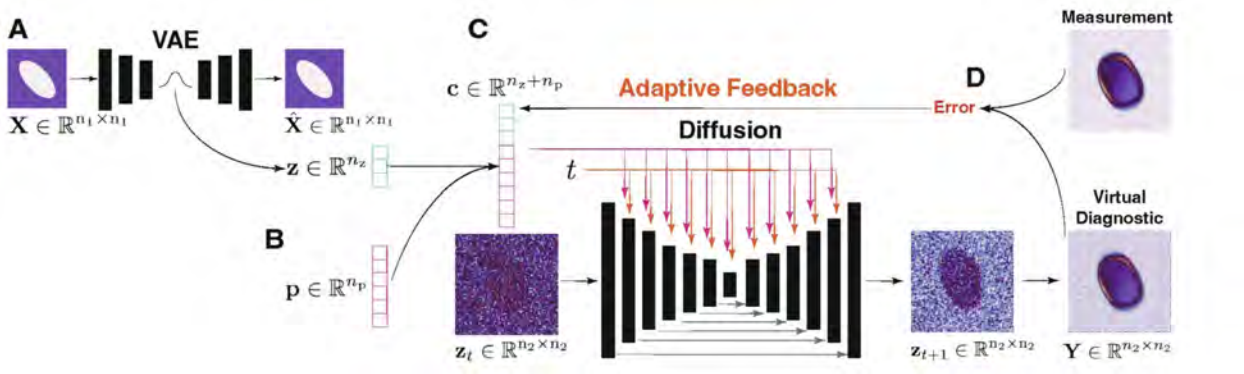


Figure 2.13 Visualization of the cDVAE algorithm [57].

To highlight another machine learning approach, this section introduces the Conditional Diffusion Variational Autoencoder (cDVAE), a generative architecture designed to perform “virtual diagnostics" for particle accelerators. The objective is to reconstruct high-fidelity projections of the 6D phase space using limited, non-invasive diagnostic inputs.

As in the previous two sections, the goal is to utilize machine learning to rapidly generate diagnostics. This work generates not only the full distribution but also all 15 unique projections of the 6D phase space. As illustrated in Fig. 2.13, the method proceeds as follows:

A. A Variational Autoencoder (VAE) encodes an input image (e.g., the initial (x, y) beam distribution) into a 3-dimensional latent embedding z.

B. The latent vector z is concatenated with a vector p, consisting of scalar parameters describing the accelerator components and beam properties, to create a conditioning vector c.

C. The vector c guides a diffusion process to generate 2D projections of the beam's 6D phase space. Generative diffusion works by gradually adding noise to an image over many steps; the model is then trained to iteratively denoise random noise until the final image is produced.

D. An adaptive feedback loop compares the measured and generated projections to tune the encoding, thereby tracking a time-varying beam.

However, this method is subject to the non-uniqueness problem inherent in tomographic reconstruction, as multiple distributions can yield the same projections. When using a single measured projection (e.g., (z, E)), adaptive feedback improves that projection but often degrades the accuracy of the other 14 projections. Furthermore, the model does not enforce physical consistency. For instance, the x-projection derived from an (x, y) image does not necessarily match the x-projection from an (x, x') image. Finally, the model's performance degrades when presented with out-of-distribution data.

In summary, three machine learning methods have been presented. A general theme is that a small loss in accuracy is sacrificed for substantial gains in speed. If an application does not require high-precision halo definition, for example, trajectory correction or envelope matching, this trade-off could be beneficial.

2.5 Concept of Uncertainty Quantification

Uncertainty quantification (UQ) is essential for making informed decisions. Direct measurement methods, such as the pepperpot emittance measurement, have been studied to produce UQ estimates [58]; however, to the best of our knowledge, UQ for beam phase-space tomography has not yet been demonstrated. While UQ for data-driven virtual diagnostics has been explored, it provides approximate results that may not necessarily ensure accuracy.

UQ is typically categorized into two types: epistemic and aleatoric uncertainty. Epistemic uncertainty refers to systematic, reducible uncertainty arising from limited knowledge of the model, such as missing data. In contrast, aleatoric uncertainty represents statistical, irreducible uncertainty stemming from experimental errors or inherent measurement noise.

Common methods for UQ in virtual diagnostics include Bayesian methods, bootstrapping, ensembling, and quantile regression [48]. Ensemble methods involve creating a collection of neural networks and combining their predictions, where the ensemble's variance serves as an estimate of uncertainty. Different ensemble members are generated by randomizing the initialization, using random subsets of data, or both, a process known as bootstrap aggregation, or bagging.

Quantile regression utilizes separate neural networks to predict different quantiles of the data distribution. For instance, a quantile of 0.10 indicates that 10% of the data lies below the predicted line, while a quantile of 0.50 represents the median prediction. Plotting multiple quantile lines simultaneously provides better insight into where the ground truth is likely to lie.

Validation of these methods relies primarily on simulation, given the scarcity of accurate, high-resolution phase space measurements in experimental settings. While preliminary tests often employ ideal lattice configurations, such scenarios fail to capture the complexities of a physical accelerator. A more rigorous approach involves generating an ensemble of machine configurations that introduce realistic errors and uncertainties. Evaluating the method against this ensemble yields a more reliable predictor of its performance in an operational environment.

CHAPTER 3

ADVANCED METHODOLOGIES IN BEAM LINE MODELING AND PHASE SPACE RECONSTRUCTION

This section will cover methods to advance beamline modeling and phase space reconstruction. First, a data-driven study was conducted on the front end of ATLAS, using a combination of NN and CNN to model the beam dynamics. Then studies were done to advance phase-space manipulations techniques.

3.1 Relating Initial Distribution to Beam Loss on the Front End of a Heavy-Ion Linac Using Machine Learning

In this section, we tested a data-driven approach to predict the beam loss using 2D projection measurements. The data was generated from virtual diagnostic instruments simulated using the beam dynamics code TRACK. The simulation data were obtained from a test lattice based on the LEBT of the ATLAS accelerator and were used to develop a convolutional autoencoder that encodes the data into a meaningful lower-dimensional representation, thereby relating phase-space information to beam loss.

3.1.1 Collecting the Data

The presented study used data generated by simulating ATLAS's LEBT. The virtual diagnostic instruments capture the 4D phase space of the beam. The locations are shown in Fig. 3.1 where the phase-space projections of the 4D phase space and losses are measured and saved. This amount of information is currently hard to achieve in a real accelerator but it is used to study the feasibility

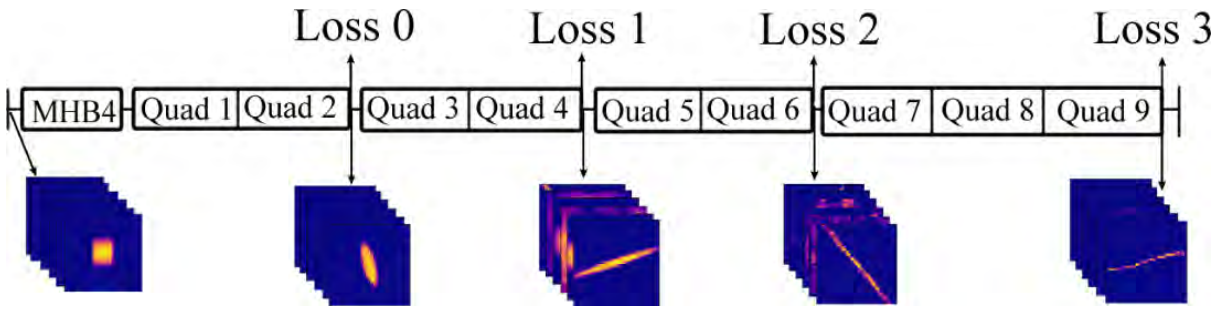


Figure 3.1 Cartoon of accelerator and beam measurements. The image shows the locations of the beam measurements.

Table 3.1 Parameter range used to generate the data set of the initial beam distributions and quadruple settings.

Input	
Voltages on Quadruples 1, 3, 5	uniform random number from [0,8] V
Voltages on Quadruples 2, 4, 6	uniform random number from [-8,0] V
Initial Distribution	random distribution from 9 built in distribution
$\epsilon_{x,y}$	$0.12 + Normal(\mu = 0, \sigma = 0.012)$ cm*mrad
$\alpha_{x,y}$	$Normal(\mu = 0, \sigma = 1)$ unitless
$\beta_{x,y}$	$100 + Normal(\mu = 0, \sigma = 10)$ cm/rad
Output	
Number of particles left	[0,10000] particles. Taken at 4 different points
Position of all particles	Taken at 5 different points

of the method.

3.1.1.1 Generating Data Using TRACK

TRACK is a ray-tracing or particle-tracking code that can: (1) represent external fields accurately within the aperture. (2) Calculate the particle coordinate at any point in the space. (3) determine beam loss in both the ideal case and in the presence of complex field errors and device misalignments [27].

TRACK simulations were used to collect data because machine data were unavailable. Over a million data point was generated on Michigan State University's high-performance computing cluster. This is necessary because a significant amount of data will be required to train autoencoders to high fidelity. The parameters for these simulations were varied according to Table 3.1 and were chosen within and interpertable range. The data were filtered so that the initial beam distributions remained within the beam aperture, yielding a final data set of approximately 430,000 simulation points.

2D phase-space projections where taken by depositing the particles onto an $n \times n$ grid using pairs of the coordinates axes, (x, x', y, y'). This resulted in 6 independent projections.

3.1.1.2 Non-linear Field

A separate data set was generated to assess the model's generalizability, which will be explained later. This was done by perturbing the initial distribution with a non-linear magnetic field, such as a sextupole, at the beginning of the simulation.

3.1.2 Creating the Model

3.1.2.1 Autoencoder

An autoencoder is a nonlinear data-reduction algorithm in machine learning. It comprises two parts: an encoder and a decoder. The encoder takes a high-dimensional input and reduces it to a lower-dimensional latent representation, while the decoder attempts to reconstruct the latent representation from the original input. The error, which is the difference between the original and reconstructed data, quantifies how well the latent dimension explains the original input. The advantage of compressing the data into a meaningful representation [59] makes it more efficient to train a neural network model on the reduced data.

In the model, a convolutional autoencoder implemented in PyTorch [60] was used to reduce the input dimensionality. A convolutional autoencoder uses a convolutional neural network as the encoder and decoder. A convolutional neural network is a type of neural network used to analyze visual information [34]. This has the advantage over principal component analysis[61], another data reduction algorithm, in that it includes spatial information, and can account for non-linear effects by using non-linear activation functions in the network. Activation functions map the input to a finite range. It was found that the ReLU and ELU activation functions were the best to use [62], which, in this case, helps the model train faster and be less likely to fail during training.

Each of the six 2D projections was given its own autoencoder. The decoder reproduced all the original projections with reasonable accuracy, thereby verifying that the projections were effectively encoded in the latent dimension. The latent dimension sizes used for this paper were 32 for the (x, x') and (y, y') projections, and 16 for the rest. Given that the original images were made to be 33×33 pixels, the inputs were significantly reduced.

3.1.2.2 Modeling

A neural network was used to create a surrogate model of the ATLAS front-end as shown in Fig. 3.2. The architecture is composed of, first, an encoder-decoder block that reduces, separately, each of the six phase-space projections into lower latent dimensions, and then concatenates them. The quad settings were also concatenated onto this vector. This vector is passed through a fully

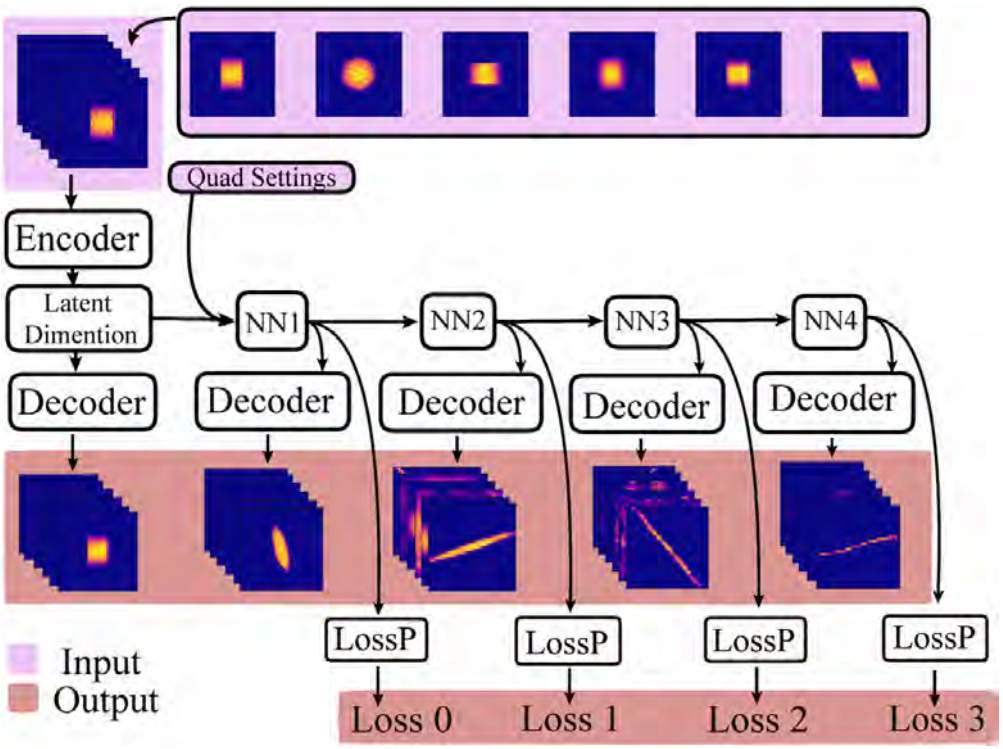


Figure 3.2 Cartoon of architecture. During training, the model takes all the 2D projections and loss value as input into the training. Only the initial 2D projections were given during testing, and the model predicts the loss values and 2D projections in addition.

connected layer that performs a phase-space transformation on the latent dimension. The output from this is passed to a decoder block to reconstruct the 2D phase-space projections at that location, to another fully connected layer to predict the number of particles remaining, and to another fully connected layer to repeat the process until the end.

The encoder-decoder block uses a convolutional autoencoder as described in the previous section. A decoder was not trained at every location but was combined across projections. This saves limited GPU memory and produces a more generalized decoder.

To estimate the number of remaining particles, a two-layer fully connected network was used. Again, the network was not trained at every location, but it was combined to make a generalized particle loss predictor for the same reasons stated above.

3.1.2.3 Training

Overall, the model encodes each initial phase-space projection into separate latent dimensions, then attempts to recreate the phase-space projections and predict the beam transmissions at the

other 4 locations, and then finally compares them to the ground truth from the dataset. Using both results in a training loop, the model will update itself using gradient descent [34] to better recreate the image in the next iteration.

The model used a loss function to quantify the difference between the predicted and ground truth results. Commonly, mean square error loss is used because it punishes large deviations; however, this results in an overflow in the gradient calculation during training—an error that occurs when a computer produces a number more significant than it can represent. This could have be resolved by normalizing the inputs and outputs, but a non-normalized dataset was used since that gave a better convergence. The model used absolute loss (L1 Loss) in those cases it would fail. This could also have been resolved by using a double or a long float, but this would use up valuable memory on the GPU.

To also aid in training, frozen layers were implemented to decrease training time and prevent gradient overflow. Frozen layers are layers where the gradient information was disconnected, thus preventing changes to that layer during training. This method was implemented in the following three-step training procedure:

1. LossP blocks were frozen and the learning rate was set to 0.01. Everything else was trained for 20 epochs. This correspond to the loss function

$$\mathcal{L}_{\text{distro}} = \sum_{i=0}^{5} \frac{1}{N_p} \sum_{d=1}^{N_p} \left(\hat{f}_{i,d} - f_{i,d} \right)^2, \tag{3.1}$$

   where N_p is the total number of pixels, $f_{i,d}$ and $\hat{f}_{i,d}$ is the ground-truth and reconstructed ith set of projections, where i is the projection set, and d is the pixel number.

2. LossP blocks were unfrozen, and everything else was frozen; thus, only LossP was trained for 10 epochs. The learning rate was also 0.01. The loss function here was

$$\mathcal{L}_{\text{loss\_pred}} = \sum_{i=0}^{4} \left| \hat{\ell}_i - \ell_i \right|, \tag{3.2}$$

   where ℓ and $\hat{\ell}$ is the ground-truth and predicted particle loss.

3. Everything was unfrozen and trained for 5 epochs at a lower learning rate of 0.001. The loss function was a combination of the two.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{distro}} + \mathcal{L}_{\text{loss\_pred}} \tag{3.3}$$

A problem that arises from using simulation is overfitting, which is a state where the model memorizes the training data rather than generalizing it. Because simulation generally differs from the actual machine due to installation and operation errors, if the model is not generalized or adjusted enough, it will perform poorly on the actual device.

The model used the last step of the training procedure to better merge the LossP block with the whole model. This also helps to prevent overfitting. A lower learning rate than the previous two steps discourages radical changes that may destroy the learned model in the first two steps.

3.1.2.4 Results

The model was tested on a newly generated dataset using the original parameters, as well as a non-linear dataset generated using a sextupole. Only the initial distributions were given, but the model would still predict the 2D projections and beam transmission at the other locations downstream. Then, to test the generalization of the model, a nonlinear field in the form of a sextupole [36] was added to the beginning of the simulation to generate a dissimilar subset of inputs.

In this study, the total number of simulated particles is 10,000. We define the prediction error as the absolute difference between the ground-truth and predicted values, divided by this total particle count. Furthermore, throughout this paper, any reported percentage error specifically evaluates the results bounded within two standard deviations of the mean. Because ATLAS operates as a low-power machine, an error of less than 1% within this 2-standard-deviation bound is considered sufficiently accurate for particle loss prediction. Figure 3.3A plots our obtained values as a correlation graph, where a perfectly straight line indicates zero error. Based on these metrics, the model achieved a 3% error on the original dataset using six projections.

This was then tested on the nonlinear sextupole distribution with fair results, an error of 5.5% as

shown in Figure 3.3C. The model was able to generalize fairly well, however, it is still far from the ideal case. In this case, a machine learning model mainly interpolates the results, so the accuracy of a model depends on how much data it can train on. The more data points a model has, the better the interpolation will be.

Due to the nature of hadron accelerators, many of the quadrupole configurations would produce a high particle loss because only a few configurations would allow most of the particles to pass. Thus, most of the dataset would be skewed towards high loss, resulting in higher accuracy in those cases since there is more data in those cases. To analyze this effect, the dataset was split into bins and as expected, the bin of particle loss between $9000 - 10000$ has an error around 2.5% and for the bin of particle loss between $0 - 1000$, the error was as high as 5%.

3.1.2.5 Testing on a Smaller Data Set

The same model was tested again, but with the (x, y'), (x', y'), and (y, x') projections removed. In Fig. 3.3B, the error predictions from the original data set show improved accuracy for "Loss: 0", whereas the other losses have approximately the same error. This is likely due to overfitting, as the predictions on the non-linear data set show a loss of overall accuracy, as seen in Fig. 3.3D; however, the model was shown to perform with half the image data used, making it more practical.

3.1.3 Conclusion

A proof-of-principle machine-learning-based model has been reported to test an ML-based 4D tomography using its 2D projections and to predict the beam transmission. The result shows that, given only three projections of the 4D phase space, these projections can be reduced to a lower latent dimension that captures the core information, which can then be used to predict downstream beam transmission. The latent dimension was verified to contain the core information by a decoder that correctly reconstructed the encoded images. This method generalizes fairly well to initial beam distributions with non-linear perturbations, showing robustness and the potential to model the real machine.

Before applying this method to a real machine, note that it is a simplified model of an actual accelerator. First, this model assumes that the accelerator elements can be modeled by a single pa-

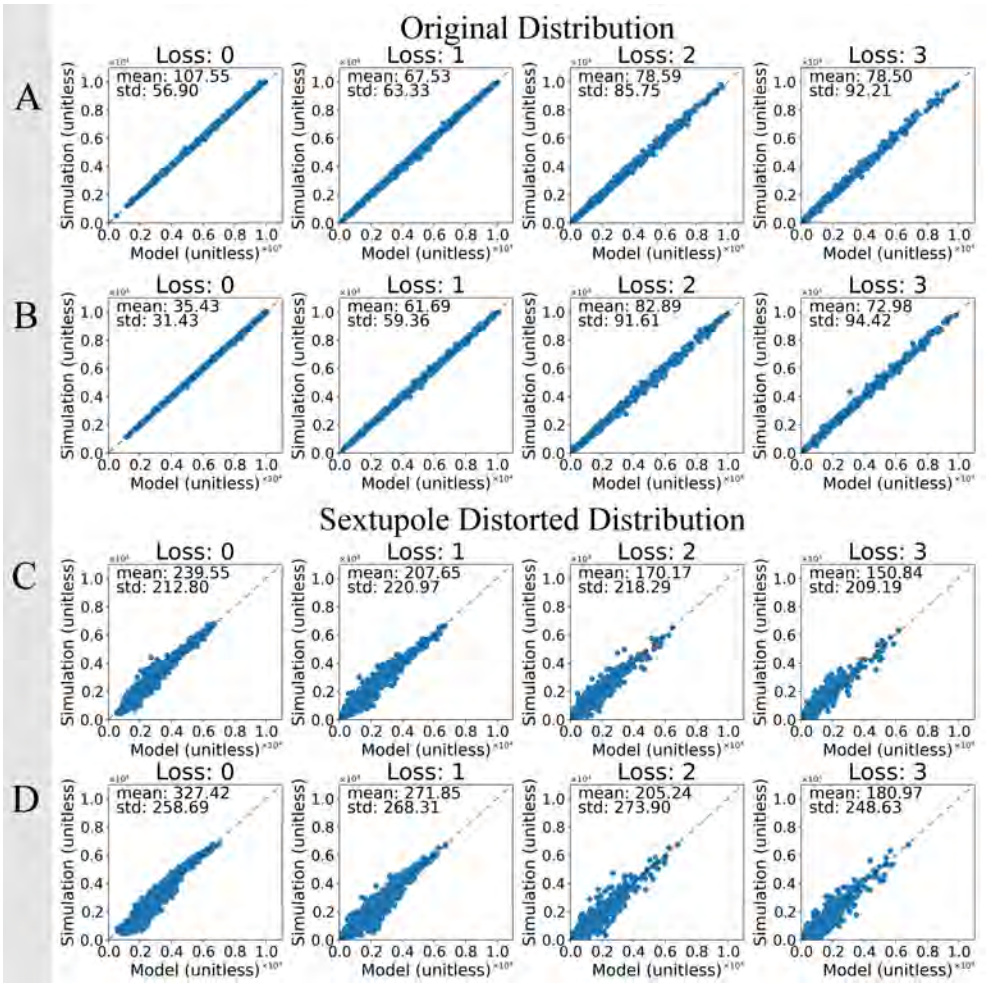


Figure 3.3 Histogram of original data set using six projections (A), and the same model but using three projections (B). Histogram of original data set using six projections (C), and the same model but using three projections (D).

rameter. Therefore, more complex effects, such as magnet misalignment and longitudinal overlap of transverse magnets, are not considered. Second, the model assumes that 2D projections can be measured precisely with no error. Measurement errors exist, but they can be reduced by taking multiple measurements with different optics settings, which was not done in our simplified model.

To use this method in experiments, methods known as "transfer learning" will have to be tested. This allows knowledge learned from the source dataset to be transferred to a target dataset [34]. This is done by freezing the model, adding an extra layer, and training that layer with the frozen model on the real machine. Afterward, the entire model can be unfrozen and trained with a much lower learning rate to fine-tune it. The positive results of this work give hope that incorporating this

knowledge may save time, increase sample efficiency, and further reduce the beam transmission error.

3.2 Comparing Different Phase Space Reconstruction Algorithms

TBD

3.3 Extending Phase Space Reconstruction Methods to Higher Dimensions

This section outlines the theoretical framework developed to extend phase space reconstruction algorithms to higher dimensions. First, a generalized beam matrix reconstruction algorithm utilizing LLSQ is introduced. This is followed by a generalized approach for mapping measured 1D projections along a linear beamline into arbitrary directions in higher-dimensional phase spaces. Finally, an extension of the MENT algorithm is detailed, providing a computational method to directly reconstruct a 4D phase-space distribution from 1D spatial projections.

3.3.1 Generalized Beam Matrix Reconstruction

This section generalizes the linear least squares (LLSQ) formulation (introduced in Sec. 2.3.2) to an N-dimensional phase space, with practical applications primarily for $N \in \{2, 4, 6\}$.

Let the state of a particle be defined by the phase-space coordinate vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$. The beam is statistically described by the symmetric $N \times N$ beam matrix (or covariance matrix), denoted by σ, whose elements consist of the beam's second central moments, $\sigma_{ij} = \langle x_i x_j \rangle$ defined in Eq. 2.15.

Consider the linear transport of the beam from an initial state to a final state, denoted by the subscripts 0 and f, respectively. Let M represent the $N \times N$ linear transfer matrix mapping the initial phase-space vector to the final phase-space vector ($\mathbf{x}_f = M\mathbf{x}_0$). The evolution of the beam matrix is thus described by the similarity transformation

$$\sigma_f = M\sigma_0 M^T. \tag{3.4}$$

Assume diagnostic measurements are performed at the final location to extract specific second moments of the beam profile. To condense the notation, let $s_{a,b} \equiv (\sigma_f)_{a,b}$ denote the measured moment between the a-th and b-th coordinates at the final location. By expanding the matrix

multiplication, these measured moments can be mapped directly to the unknown elements of the initial beam matrix

$$s_{a,b} = \sum_{i=1}^{N}\sum_{j=1}^{N} M_{ai}(\sigma_0)_{ij}M_{bj}, \tag{3.5}$$

where M_{ij} represents the elements of the transfer matrix M. Equation 3.5 forms the basis of the linear system necessary to reconstruct the initial phase-space distribution via the LLSQ method.

Since the beam matrix is symmetric, it has $N(N+1)/2$ unique elements. These are the elements to be solved, and the minimum number of measurements required to solve the system.

As mentioned at the end of Sec. 2.3.2, cases such as an uncoupled system, where there are no cross-dimensional couplings nor measurements, would not be able to probe all the degrees of freedom, leading to a rank-deficient matrix and and incorrect solution.

Let K be the total number of measurements, and index k representing the kth measurement. To formulate this as a linear least squares problem, each measurement provides one row in the linear system $\mathbf{b} = A\mathbf{x}$, defined as follows:

- $\mathbf{x}$: The column vector of length $N(N+1)/2$ containing the unknown unique initial beam moments, the upper-triangular elements of σ_0.
- $\mathbf{b}$: The column vector of length $K \geq N(N+1)/2$ containing the measured second moments.
- A: The $K \times N(N+1)/2$ design matrix. Each row corresponds to a single measurement.

Let $s_k = (\sigma_f^{(k)})_{a_k,b_k}$ be the k-th measurement taken after transport through matrix $M^{(k)}$ and $\mathbf{x}$, the upper triangular matrix, such that $\mathbf{x} = \{(\sigma_0)_{ij} | 1 \leq i \leq j \leq N\}$. Written out, the vectors are structured as:

$$\mathbf{x} = \begin{pmatrix} (\sigma_0)_{11} \\ (\sigma_0)_{12} \\ \vdots \\ (\sigma_0)_{ij} \\ \vdots \\ (\sigma_0)_{NN} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \\ \vdots \\ s_K \end{pmatrix} \tag{3.6}$$

To simplify the notation, recall again that s_k refers to the ab-moment of the kth measurement. This measurement generates the k-th row vector of the design matrix A, which we will denote as $\mathbf{a}_k$. By expanding Eq. 3.5, the specific coefficient in $\mathbf{a}_k$ are

$$(\mathbf{a}_k)_{ij} = \begin{cases} m_{ai}m_{bi} & \text{if } i = j \quad \text{(diagonal moments)} \\ m_{ai}m_{bj} + m_{aj}m_{bi} & \text{if } i < j \quad \text{(off-diagonal moments),} \end{cases} \tag{3.7}$$

where m represents the elements of matrix M_k, associated with the k-th measurement. Finally, stacking these row vectors for all K measurements constructs the full $K \times N(N+1)/2$ design matrix A

$$A = \begin{pmatrix} — & \mathbf{a}_1 & — \\ — & \mathbf{a}_2 & — \\ & \vdots & \\ — & \mathbf{a}_K & — \end{pmatrix} \tag{3.8}$$

After constructing A and $\mathbf{b}$ in this manner, the initial phase space vector $\mathbf{x}$ can be estimated using the pseudo-inverse, Eq.2.39.

3.3.2 Generalizing Projection Mapping to N-D

The fundamental principle of beam tomography relies on the deterministic mapping of particle trajectories throughout the beamline. Because the transport matrix is known, a projection measured at the end of the beamline can be mapped to a corresponding projection at the start of the beamline. For brevity, let location A denote the position of the initial beam distribution, and location B denote the downstream diagnostic position where the projection is measured.

Let

$$f_A(\mathbf{x}) = f_A(x_1, x_2, \ldots, x_N) \tag{3.9}$$

be an N-dimensional phase-space distribution of particles at location A. Similarly, $f_B(\mathbf{x})$ defines the distribution at location B.

We define a 1D spatial projection as

$$P^{(\hat{\mathbf{u}})}(s) = \int f(\mathbf{x})\delta(s - \hat{\mathbf{u}} \cdot \mathbf{x})\, d^N\mathbf{x}, \tag{3.10}$$

where $\hat{\mathbf{u}} \in \mathbb{R}^N$ is a unit vector defining the projection direction. Geometrically, the condition $s = \hat{\mathbf{u}} \cdot \mathbf{x}$ defines a parallel hyperplanes normal to the $\hat{\mathbf{u}}$ direction.

Let the linear transport from A to B be represented by the transfer matrix M. Since the beam-line transport is a symplectic map, it preserves the phase-space volume, i.e., $\det(M) = 1$. The coordinate transformation is given by:

$$\mathbf{x}_B = M\mathbf{x}_A. \tag{3.11}$$

Because the number of particles is conserved, the distributions are related by

$$f_B(\mathbf{x}_B)\, d^N\mathbf{x}_B = f_A(\mathbf{x}_A)\, d^N\mathbf{x}_A. \tag{3.12}$$

According to Liouville's Theorem, the phase space density is constant along the trajectories of the system. This means the Jacobian of a symplectic transformation is unity, $d^N\mathbf{x}_B = d^N\mathbf{x}_A$, the phase-space density is $f_B(\mathbf{x}_B) = f_A(M^{-1}\mathbf{x}_B) = f_A(\mathbf{x}_A)$.

Now, let $P_B^{(\mathbf{e}_1)}(s)$ be the 1D projection of the distribution f_B along the standard unit vector $\mathbf{e}_1$. We can map this measurement back to a projection taken along a different direction at A. Using Eq. 3.11 and Eq. 3.12, we expand the projection integral.

$$\begin{aligned} P_B^{(\mathbf{e}_1)}(s) &= \int f_B(\mathbf{x}_B)\delta(s - \mathbf{e}_1 \cdot \mathbf{x}_B)\, d^N\mathbf{x}_B \\ &= \int f_A(\mathbf{x}_A)\delta(s - \mathbf{e}_1 \cdot (M\mathbf{x}_A))\, d^N\mathbf{x}_A \end{aligned} \tag{3.13}$$

The dot product in the delta function can be rewritten using the transpose property: $\mathbf{e}_1 \cdot (M\mathbf{x}_A) = \mathbf{e}_1^T M\mathbf{x}_A = (M^T\mathbf{e}_1)^T\mathbf{x}_A = (M^T\mathbf{e}_1) \cdot \mathbf{x}_A$. Since the projection must be evaluated along a unit vector, we define a scaling factor α such that the vector is normalized.

$$\alpha = ||M^T\mathbf{e}_1|| \tag{3.14}$$

Substituting this back into the integral yields

$$P_B^{(\mathbf{e}_1)}(s) = \int f_A(\mathbf{x}_A)\delta\left(s - \alpha\left(\frac{M^T\mathbf{e}_1}{\alpha}\right)\cdot\mathbf{x}_A\right)\, d^N\mathbf{x}_A. \tag{3.15}$$

Applying the delta function scaling property, $\delta(cx) = \frac{1}{|c|}\delta(x)$ for $c \neq 0$, we can extract α

$$\begin{aligned} P_B^{(\mathbf{e}_1)}(s) &= \frac{1}{|\alpha|}\int f_A(\mathbf{x}_A)\delta\left(\frac{s}{\alpha} - \left(\frac{M^T\mathbf{e}_1}{\alpha}\right)\cdot\mathbf{x}_A\right)\, d^N\mathbf{x}_A \\ &= \frac{1}{|\alpha|}P_A^{\left(\frac{M^T\mathbf{e}_1}{\alpha}\right)}\left(\frac{s}{\alpha}\right). \end{aligned} \tag{3.16}$$

Setting $\mathbf{e}_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^N$, representing a standard profile measurement onto the x-axis, the scaling factor is

$$\alpha = ||M^T\mathbf{e}_1|| = \sqrt{M_{11}^2 + M_{12}^2 + \cdots + M_{1N}^2}. \tag{3.17}$$

This mathematically proves that a 1D projection measured at location B is equivalent to a projection taken at location A along the mapped direction $M^T\mathbf{e}_1$, scaled by the geometric factor α.

In practice, it is preferable to invert this relationship to scale the downstream projections at B back to the initial state at A, rather than to vary the reconstruction space. Following an identical derivation using the inverse map, the inverse relationship is

$$P_A^{(\mathbf{u}_1)}(s) = \frac{1}{\beta}P_B^{\left(\frac{(M^{-1})^T\mathbf{u}_1}{\beta}\right)}\left(\frac{s}{\beta}\right) \tag{3.18}$$

where $\mathbf{u}_1$ is the desired projection unit vector at A, and β is the normalization factor to normalize $(M^{-1})^T u_1$.

3.3.2.1 Results for $N = 2$

Previous work has derived these geometric relations for the $N = 2$ transverse phase space [49]. The corresponding phase-space rotation angle θ and scaling factor α are given by

$$\tan(\theta) = \frac{M_{12}}{M_{11}} \tag{3.19}$$

$$\alpha = \sqrt{M_{11}^2 + M_{12}^2} \tag{3.20}$$

We can show our generalized result gives back this derivation. Given the standard horizontal measurement axis $\mathbf{e}_1 = (1, 0)^\top$, Eq. 3.20 follows directly from the N-dimensional definition of α. Furthermore, Eq. 3.19 dictates the angle between $\mathbf{e}_1$ and the mapped vector $M^T\mathbf{e}_1$

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \quad \Rightarrow \quad M^T \mathbf{e}_1 = \begin{pmatrix} M_{11} \\ M_{12} \end{pmatrix} \quad \Rightarrow \quad \tan(\theta) = \frac{M_{12}}{M_{11}}$$

3.3.2.2 Verification in Simulation

This mathematical scaling was verified numerically via particle-tracking simulations. The following Python function approximates the analytical projection (Eq. 3.10) by binning the particles along the arbitrary direction $\mathbf{u}$ and applying Gaussian smoothing to approximate the Dirac delta function.

```
def projection(particles, u, s_values, epsilon=0.01):
    """
    Approximate projection using Gaussian smoothing
    of particles along direction u.
    """
    u = u / np.linalg.norm(u)
    projections = particles @ u

    # Approximate delta using Gaussian smoothing
    P = np.zeros_like(s_values)
    for z_proj in projections:
        P += np.exp(-0.5 * ((s_values - z_proj)/epsilon)**2) / (epsilon *
    np.sqrt(2*np.pi))
    return P / len(particles)  # normalize by particle count
```

A unique, asymmetric 4D distribution was generated at location A. The beam was transported to location B through a simulated lattice consisting of four standard quadrupoles and two skew quadrupoles. The strengths of all six magnets were systematically varied across 20 distinct settings to adequately rotate the phase space. For each setting, the linear transfer matrix M was recorded.

Spatial projections at the downstream location, P_B, were simulated along the standard $\mathbf{e}_1$ axis, while upstream projections, P_A, were taken along the theoretically mapped vector $M^T\mathbf{e}_1$. As dictated by the derived scaling laws, applying the amplitude and coordinate scaling factor α via the transformation

```
    plt.plot(s / alpha, P_B * alpha)
```

yields excellent agreement between the downstream and upstream projections, as shown in Fig. 3.4.

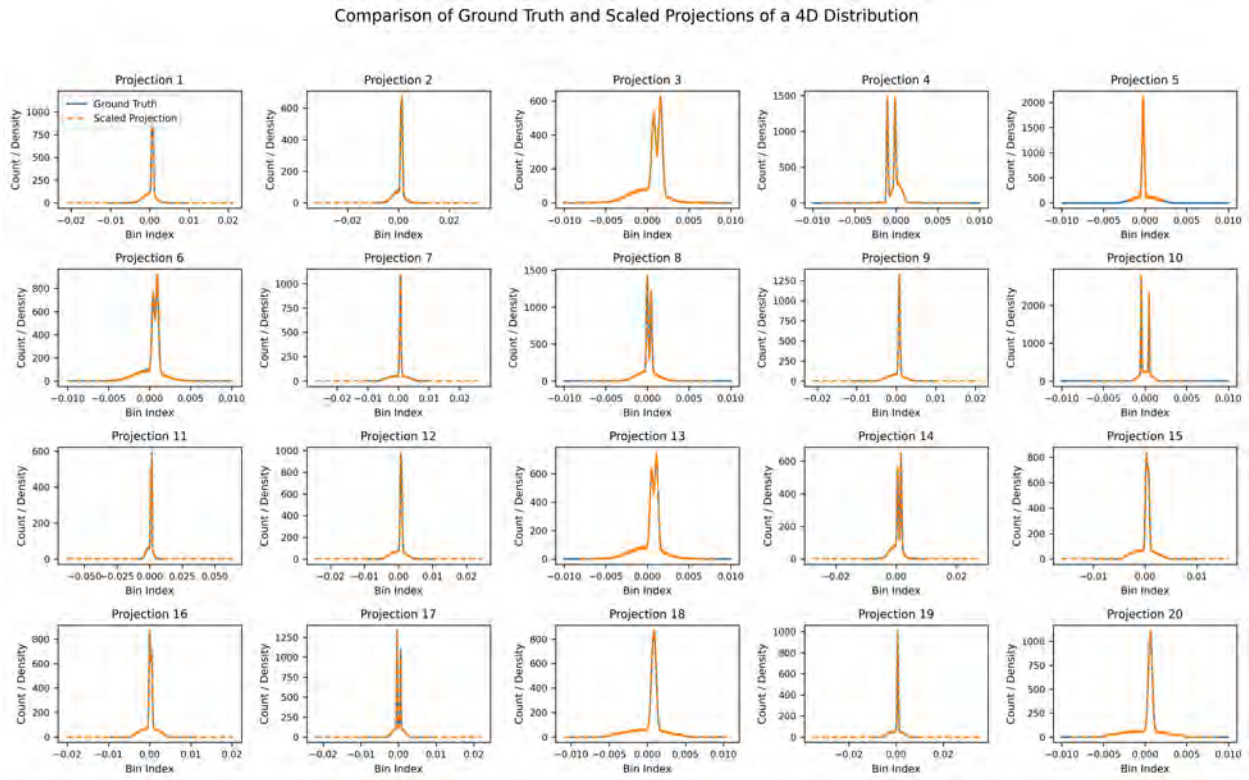


Figure 3.4 Comparison of ground truth (blue solid line) and scaled projections of a 4D distribution (broken orange line). Each subplot corresponds to a different view in the 4D phase space.

This numerical result successfully validates that any 1D profile measured at B corresponds exactly to a mathematically defined projection at A, scaled by the geometric factor α, provided the transport between the locations remains symplectic.

A major advantage of this generalized algorithm is that it enables the systematic optimization of measurement directions within an N-dimensional phase space. In the standard 2D formulation, specific phase-space projection angles can be targeted by optimizing the beamline's magnetic settings to produce the needed transfer matrix. This work successfully extends that principle to higher dimensions. By optimizing the N×N transfer matrices obtained from magnetic lattice optimization, one can select and measure projections along arbitrary hyperdimensional directions. This can be used in future research to optimize tomographic algorithms.

3.3.3 Generalizing the MENT algorithm

The derivation of MENT for 2D reconstruction was shown previously in Sec.2.3.5. The distribution $f(\mathbf{x})$ is given by a product of various "h-functions" given by Eq.2.64,

$$f(\mathbf{x}) = \prod_{k=1}^{K} h_k[u_k(\mathbf{x})],$$

while the simulated projections were given by Eq. 2.49,

$$\hat{p}_k(u_k) = \int_{-\infty}^{\infty} \hat{f}\left(M_k^{-1} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right) dv_k, \quad k = 0, 1 \ldots K.$$

To extend this to N dimensions, let the probability distribution be $f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^N$, $M_k : \mathbb{R}^N \to \mathbb{R}^N$ be a transfer map representing the beamline transport. Denote $\mathbf{u} = (u, \mathbf{v})^T$, where u is the axis being projected on, and $\mathbf{v}$ represents the axes being integrated over.

With this, the simulated projections are given by

$$\begin{aligned} \hat{p}_{k'}(u) &= \int_{\mathbb{R}^{N-1}} f(M_{k'}^{-1}(u, \mathbf{v})^T) d\mathbf{v} \\ &= h_{k'}(u) \int_{\mathbb{R}^{N-1}} \left[\prod_{k \neq k'} h_k\Big(u_k\big(M_{k'}^{-1}(u, \mathbf{v})^T\big)\Big) \right] d\mathbf{v} \end{aligned} \tag{3.21}$$

The key to this dimensional extension lies in understanding how the 1D h-functions (h_k) are coupled and integrated. Each h_k is a function of a single independent variable, u_k. If we expand the product inside the integral for the k'-th measurement, it takes the form:

$$\hat{p}_{k'}(u) = h_{k'}(u) \int_{\mathbb{R}^{N-1}} \left[h_0(u_0) \cdots h_{k'-1}(u_{k'-1}) h_{k'+1}(u_{k'+1}) \cdots h_K(u_K) \right] d\mathbf{v}. \tag{3.22}$$

The independent variables u_k do not share the same coordinate basis. Each depends on the specific transfer matrix M_k of its respective measurement. Therefore, to evaluate the integral, each h-function must be transformed back via a coordinate transformation into a common reference frame leading to the equation we see in Eq. 3.21.

The extension of MENT to higher dimensions is entirely facilitated by this mapping. The algorithm transforms the h-functions into a shared N-dimensional space and integrates over $d\mathbf{v}$ to calculate the expected projection.

Once the h_n are found, which match the measured projections $\hat{p}_k(u_k) - p_k(u_k) = 0$ for all u, then we would have found the solution that maximizes the entropy. Notice that each h-function has a related projection that is in the same coordinate system. This allows the k'th component to be factored out, enabling a relationship between the two as

$$\frac{p_{k'}(u)}{\alpha(u)} = h_{k'}(u), \quad \alpha(u) = \int_{\mathbb{R}^{N-1}} \left[\prod_{k \neq k'} h_k\Big(u_k\big(M_{k'}^{-1}(u, \mathbf{v})^T\big)\Big) \right] d\mathbf{v} \tag{3.23}$$

Where the simulated projection $\hat{p_{k'}}$ has been switched with the measured projection $p_{k'}$ to allow for one to solve for the h-functions. The problem then shifts from finding $f(\mathbf{x})$ that maximizes entropy and is consistent with the measured projections to efficiently solving these coupled integral equations.

This set of equations constitutes a system of coupled, nonlinear integral equations. This was originally solved using a Gauss–Seidel method as seen in the end of Sec. 2.3.5. A particle-sampling method using MCMC has also been used to effectively sample in higher dimensions[63]. Other methods, such as gradient descent and numerical methods, have yet to be explored to solve this system.

3.4 Selection Algorithm

While phase-space reconstruction has advanced significantly, identifying the new set of optimal measurements to improve reconstruction remains an open question. Uniform angular sampling from 0° to 180° [50] in normalized phase space is considered the gold standard for 2D cases. The extension to higher dimensions is the topic of this section.

In practice, magnet limitations and beam-loss concerns can prevent full angular sampling. Additionally, normalized coordinates require knowing the initial Twiss parameters, which are typically unknown until measurements are taken. Nevertheless, accurate reconstructions are possible without full sampling [64, 33]. An adaptive projection-angle sampling method can yield more

accurate distribution matching more efficiently over wider range of distributions than a uniform sampling method.

This work first addresses the ideal 2D case, assuming no prior knowledge and an ideal lattice. It will then incorporate more realistic effects, such as limitations due to quadrupole strengths and beam loss.

Sampling Conjecture

For the general tomographic reconstruction of an arbitrary 2D distribution $f(\mathbf{x})$, the Fourier Slice Theorem states that the 1D Fourier transform of a spatial projection taken at an angle θ corresponds directly to a 1D slice of the 2D Fourier transform $\tilde{f}(u, v)$, passing through the origin at that same angle θ.

In this frequency-domain representation, information loss is dominated by the largest angular gap of the unsampled space. Values not sampled are interpolated, and to bound the maximum interpolation error, the largest gap must be minimized. Thus, the sampled slices must be maximally and evenly separated from one another.

Thus, for a fixed number of K measurements, the optimal strategy is equiangular sampling, defined by an angular step size of $\Delta\theta = \pi/K$. This configuration provides uniform coverage of the frequency domain. For iterative reconstruction methods, where equiangluar sampling is not possible, one can iteratively select a set of measurement angles that minimizes the maximum angular spread between any two adjacent projections.

3.4.1 General Algorithm

A general algorithm for measurement selection in beam tomography generally falls between a black-box model of the beamline and a known model of the beamline. While black-box methods are useful for empirical tuning, they are incompatible with tomographic reconstruction, which requires a known forward model. Therefore, an accurate model is assumed for this theoretical development. In practice, there are often discrepancies between our understanding of the beamline and the actual beamline, but it is assumed sufficient to proceed with tomography.

Even with an ideal linear lattice, phase-space sampling is limited by physical constraints, such

as finite beam-pipe apertures and limits on magnetic gradients. To optimize tomographic reconstruction while systematically respecting these physical limits, the following iterative measurement algorithm is proposed

1. **Initialization and Twiss Estimation:** Compute the beam matrix and extract the initial Twiss parameters. This establishes the baseline beam envelope required to model aperture losses and defines the transformation matrix that maps to normalized phase space.

2. **Constrained Angle Optimization:** Transform previously acquired angles into their normalized-space equivalents. Within this space, apply the selection algorithm to compute K optimal new beamline configurations that maximize phase-space coverage bounded by the beamline constraints.

3. **Measurement and Reconstruction:** Apply these settings to the physical beamline, acquire the K measurements, and compute the N-dimensional tomographic reconstruction.

4. **Iterative Refinement:** Calculate the exact Twiss parameters directly from the newly reconstructed distribution. Substitute these updated parameters back into Step 2 to redefine the normalized space. Repeat Steps 2 through 4 until the change in the Twiss parameters falls below a defined convergence threshold.

The proposed algorithm has several limitations. An implicit limitation arises from relying on Twiss parameter refinement, which effectively assumes a Gaussian-distributed beam profile. For highly irregular beams or those with significant halos, this RMS-based boundary may not accurately represent the true physical limits of the beam. The required accuracy of the forward model is also not quantified. If an inaccurate beamline model is used, the algorithm fails to provide the optimized measurement selections.

3.4.2 2D Selection Algorithm Without Limitations

For the ideal case, consider a 2D phase-space distribution. Projections can be acquired from any angle without physical constraints, such as aperture beam loss or magnetic-field limits, typical

of a realistic lattice. The transfer matrix is effectively a simple rotation matrix defined as

$$T(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.24}$$

The projection is defined according to Eq. 2.27, where the transport matrix is given by $M_k = T(\theta_k)$. A unique measurement sample can be defined by its coordinate on a unit circle, or simply by a phase angle θ. In the 2D case, the angular representation is more convenient. Because projections taken 180° (π radians) apart contain redundant information, only a half-circle $[0, \pi)$ needs to be sampled. Thus, if K measurements are desired, the optimal equiangular sampling angles are defined as

$$\theta_k = \frac{k\pi}{K}, \quad k = 0, 1, \ldots, K-1. \tag{3.25}$$

The tomographic reconstruction is performed in the normalized frame according to Eq. 2.31. As illustrated in Figure 3.5, uniform angular sampling in the normalized phase space corresponds, upon mapping back to real phase space, to concentrating the measurement angles near the major axis of the elliptical distribution. This will be called the Normalized Uniform Projection Sampling (NUPS) algorithm.

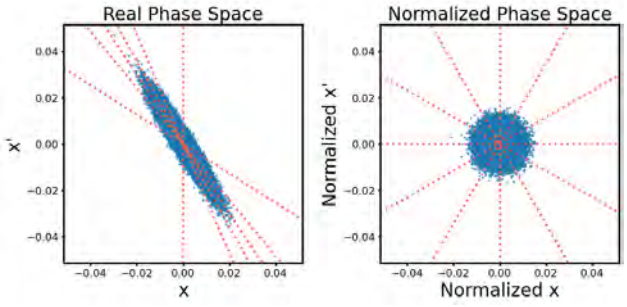


Figure 3.5 Example of uniform sampling in real versus normalized phase space.

3.4.3 2D Normalized Adaptive Projection Sampling Without Limitations

Sometimes it is not possible to sample uniformly, or multiple sets of measurements are required, as in cases where the distribution is non-Gaussian and multiple measurements are needed to correctly estimate the Twiss parameters.

3.4.3.1 Normalized Adaptive Projection Sampling (NAPS)

In an iterative algorithm, the previous measurement may still be utilized. These points will be called the fixed points. Let the sample points be the measurements proposed next. The goal is to generate K measurements which, with the fixed points, minimize the maximum angle.

To require samples only within $180°$, a trick for handling the edge conditions is to mirror the fixed points, generate $2K$ sample points, and then remove the mirrored half.

1. **Mirror the fixed angles.** If a profile is taken at an angle θ, it provides identical information as one taken at $\theta + \pi$; therefore, every fixed angle is copied to its mirror position, $\theta + \pi$, thus working on the full circle $[0, 2\pi)$. This eliminates the need to address edge conditions.

2. **Measure the gaps.** Sort all angles and sizes and measure the arc length, $\Delta\theta_j$, between consecutive angles. These arc lengths identify potential "gaps".

3. **Iteratively insert new points into gaps.** The $2K$ new points are placed as follows:

   a) Create a counter $k_j = 0$ for each gap j. This tracks the number of points assigned to gap j.

   b) Repeat $2K$ times to assign all points: In each iteration, identify the gap j^* that currently maximizes the metric $\Delta\theta_{j^*}/(1 + k_{j^*})$. Then, increment the counter k_{j^*} for this chosen gap.

   c) After all $2K$ points are assigned, for each original gap j, insert the k_j points allocated to it such that they are evenly spaced within that gap's original arc length $\Delta\theta_j$.

4. **Keep only the upper half-plane.** Discard those angles in the lower half (π to 2π), leaving exactly K angles in the interval $0 < \theta < \pi$ to use as additional projections.

Because each new angle is always placed in the largest available gap, the final set minimizes the largest gap.

3.4.3.2 2D NAPS on Guassian Beam

The setup for this analysis is identical to the previous section, where transfer matrices are given as $T(\theta_k)$ from Eq. 3.24.

NUPS and NAPS were applied to a Gaussian beam with a 2D X-plane (x, x') distribution. NAPS initially takes three measurements and then acquires four additional measurements in each subsequent iteration, updating the Twiss parameters with every iteration. NUPS also begins with three initial measurements to estimate the Twiss parameters, but this time, only one pass of a fixed number of N equally angled measurements is taken. The expected results for a Gaussian beam are that both should be similar, since the initial estimate of the Twiss parameters can be calculated precisely in the first iteration in an ideal scenario, whereas for a non-Gaussian beam, NAPS is expected to be superior.

The reconstructed images were compared with the ground-truth image using the Structural Similarity Index Measure (SSIM) [65], where a value of 1 indicates perfect structural similarity and 0 indicates strong dissimilarity. This is a well-known metric, and it is used here in preference to MSE or MAE because it focuses on structural information rather than the absolute error magnitude. The penalty for MSE and MAE would be higher if, for example, the reconstruction is identical but shifted one pixel to the right.

The result for a Gaussian beam is shown in Figure 3.6, which shows that the adaptive method performs on par with the uniform method. Figure 3.7 shows an example of the reconstruction.

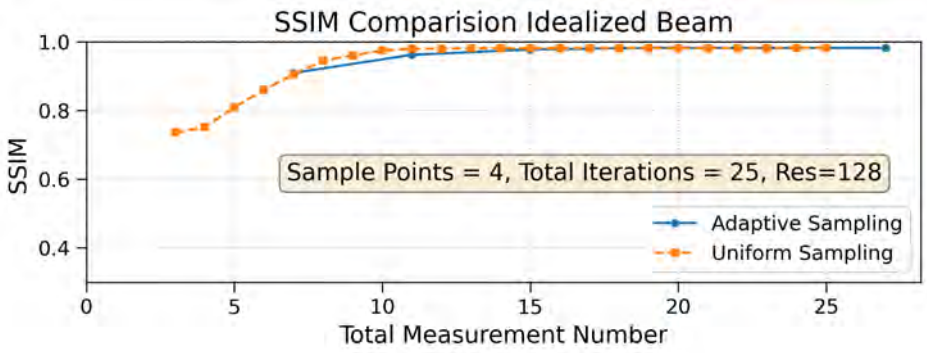


Figure 3.6 Convergence behavior for Gaussian beam in $(x - x')$ plane with $\beta = 2\frac{\text{m}}{\text{rad}}$, $\alpha = 3$, $\epsilon_{\text{rms}} = 1.5 \times 10^{-5}$ m $\cdot$ rad.

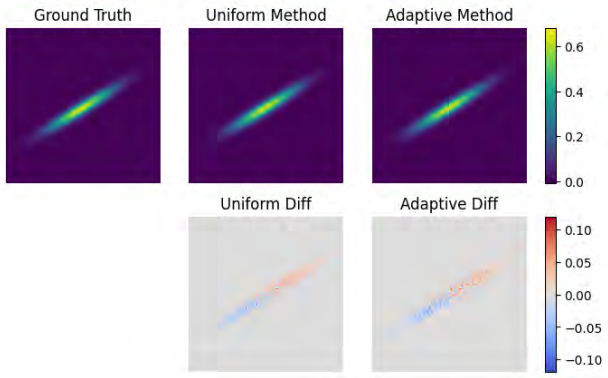


Figure 3.7 Results show reconstruction using a total of 11 measurements for a Gaussian beam distribution.

In Fig. 3.8 and Fig. 3.9, NUPS and NAPS were applied to another beam with a nonlinear beam-beam force kick applied to x' according to Eq. 3.26. This is to test the algorithms on distributions for which the Twiss parameters are not readily calculated from the first few measurements, thereby representing a more realistic scenario.

$$\delta x' = A(1 - e^{-x^2/(2\sigma^2)}), \quad A = 1 \times 10^{-4}, \quad \sigma = .003. \tag{3.26}$$

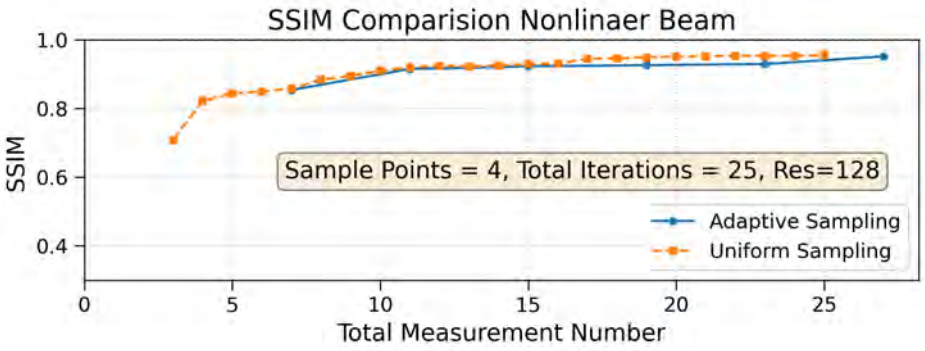


Figure 3.8 Convergence behavior for Gaussian beam in $(x - x')$ plane with nonlinear kick with $\beta = 1\frac{\text{m}}{\text{rad}}$, $\alpha = 0$, and $\epsilon_{\text{rms}} = 1.5 \times 10^{-5}$ m $\cdot$ rad.

In this case, NAPS yields errors similar to those of NUPS, but the results are inconclusive. A final test using a larger selection of initial distributions, with $\beta \in [0.1, 1.0]$ m/rad and $\alpha \in [0, 5]$, is shown in Fig. 3.10. NAPS and NUPS were used, and the results showing the mean and 10th-90th percentile spread of the SSIM are seen in Fig. 3.11. The beam was adjusted to be slightly larger than the aperture, yielding a similar nonlinear kick, with $A = 1.5 \times 10^{-4}$. This induced some beam

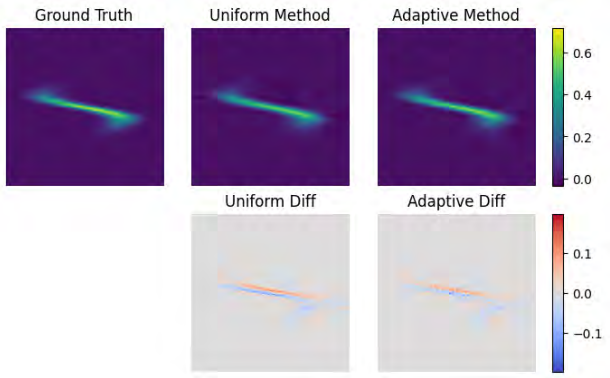


Figure 3.9 Results show reconstruction using a total of 11 measurements for non-linear beam distribution.

loss. This small beam-loss effect prevents SSIM from reaching 1, but NAPS was shown to be more robust than NUPS in this regard, either performing better or matching it. Results plateaued after 25 measurements and were removed.

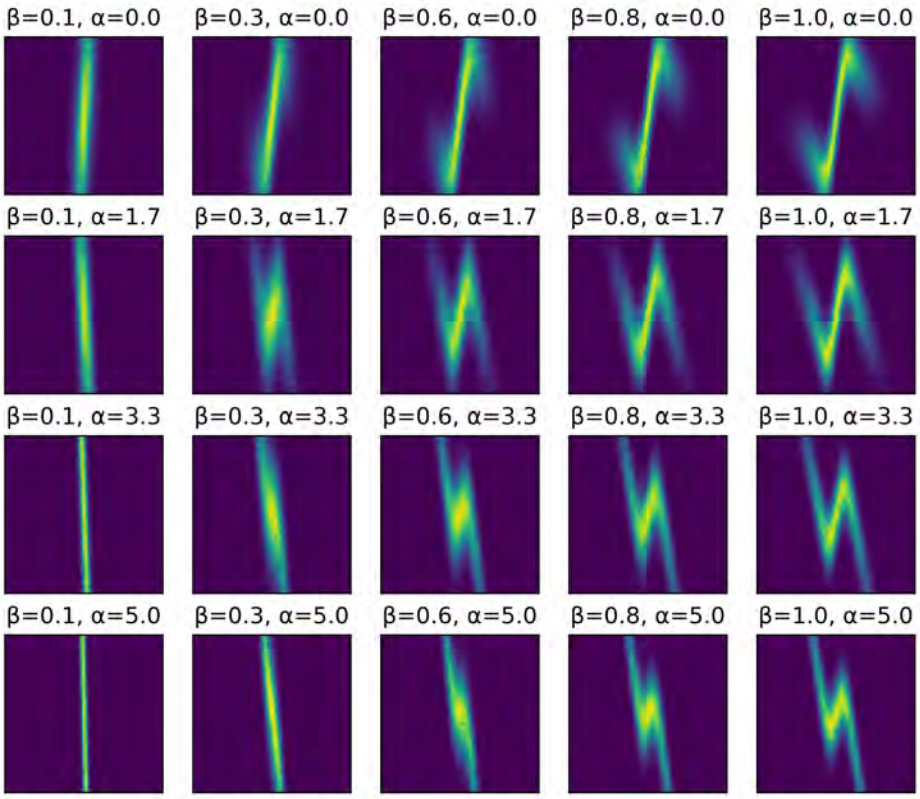


Figure 3.10 Ground truth distributions used to simulate statistics.

In the end, it was found that for a Gaussian beam, three equal-angle measurements are sufficient to accurately estimate the Twiss parameters, yielding comparable performance between the

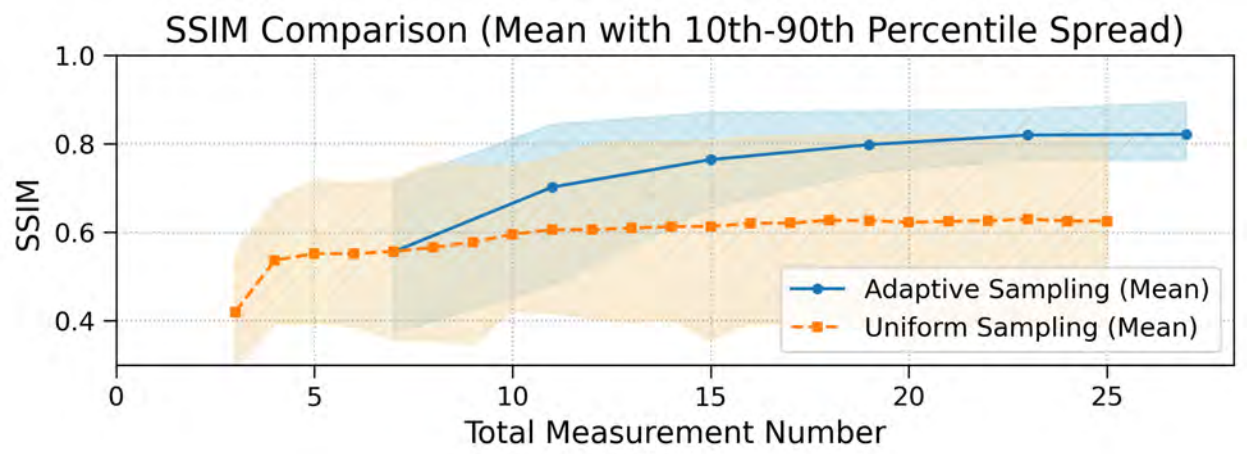


Figure 3.11 Convergence behavior for Gaussian beams in the $(x - x')$ plane with a nonlinear kick and beam loss, exploring beta functions $\beta \in [0.1, 1.0]$ m/rad and initial alpha functions $\alpha \in [0, 5]$. The RMS emittance is $\epsilon_{\text{rms}} = 5.5 \times 10^{-5}$ m $\cdot$ rad, and A $= 1.5 \times 10^{-4}$.

two methods. However, with nonlinear beam distributions, the initial three evenly spaced angles may not be optimally placed, leading to incorrect information for estimating the Twiss parameters required to accurately transform into normalized space. In these cases, NAPS's ability to iteratively self-correct during the reconstruction process yields superior or equivalent performance to NUPS.

3.4.4 2D Normalized Adaptive Projection Sampling With Limitations

(Compare results using adaptive sampling on a realistic lattice as the limitations progressively increase)

3.4.5 Discussion

This section has discussed a novel selection algorithm designed to iteratively select measurements, improving reconstruction accuracy and reducing data acquisition time; however, there are several other critical factors to evaluate when determining the optimal reconstruction algorithm for a given facility. These factors can be broadly categorized into three distinct areas: computational time, mathematical accuracy, and practical implementation.

Regarding operational time, the total duration of the tomographic process depends on the initialization time, the measurement acquisition time, the analysis time required to determine the next optimal measurement, and the algorithmic reconstruction time itself. Often, the time required to

select the subsequent measurement depends on the reconstruction time. Integrating ML techniques can significantly accelerate the reconstruction phase, while the proposed selection algorithm minimizes the total number of physical iterations required. Furthermore, different initialization can provide varying priors, which can affect the overall convergence rate of the algorithm.

In terms of accuracy, this remains somewhat under-explored. Key considerations include uncertainty quantification, validation methods, robustness to measurement noise and magnetic field errors, and sensitivity to model mismatches. Among these, uncertainty quantification emerges as the primary focus for future development, as it both depends on and can help explain these error sources.

Finally, practical implementation depends on the availability of beam diagnostics, optical flexibility, and the ability to accurately model the beamline. Customizing a framework for a new beamline introduces operational overhead. One potential solution is the development of a dedicated, software package capable of simulating various beamline configurations with modular optics and diagnostics. The primary challenge in such an approach lies in accurately modeling non-ideal conditions, such as misalignment and aperture-induced beam loss. Alternatively, data-driven methods offer the advantage of bypassing many of these effects at the cost of large datasets for training.

As such, the field of phase-space reconstruction presents a vast landscape for future research to fully realize its operational potential over all accelerators.

CHAPTER 4

TRANSVERSE PHASE SPACE DIAGNOSTICS AT ATLAS

Phase-space diagnostics provide the high-fidelity characterization required for precise beam control. The need to capture non-Gaussian features is essential in order to push the frontiers of accelerators. This chapter first details the development of a direct analysis code for pepper-pot emittance measurement. Operational constraints of the pepper-pot, primarily its restriction to low beam energies and limited spatial resolution, motivate the development of indirect diagnostic techniques. For the second part of this chapter, the standard quadrupole scan will be used to estimate the phase-space distribution before advancing to beam tomography, demonstrating its superior ability to reconstruct complex beam distributions.

4.1 Pepper-Pot Emittance Meter Software Development and Benchmarking

The pepper pot is a destructive diagnostic device that requires direct interaction with the beam. Its single-shot measurement capability allows for rapid phase-space analysis, typically in the low-energy sections of the beamline. This speed comes at the expense of position resolution, which is fundamentally limited by the spacing of the mask holes, whereas momentum resolution is limited by the imaging camera's pixel resolution. As the accelerator community advances toward the ultimate goal of single-particle measurement and control, it is worthwhile to study this device to understand both its capabilities and its fundamental limitations.

This section begins with an introduction to the pepper-pot device, covering hardware considerations, and underlying theoretical framework. While these physical and theoretical design principles are essential for constructing or upgrading a pepper pot, this chapter focuses primarily on the data processing pipeline. Specifically, the following subsections will detail the development of the analysis software, benchmarking, image processing techniques, and the practical challenges of operational implementation.

4.1.1 Introduction to Pepper Pot

The fundamental mechanics of the pepper-pot diagnostic are introduced in Sec. 2.2.1.3. A metallic mask filters the incoming beam into distinct beamlets at known transverse positions, x_{hole}.

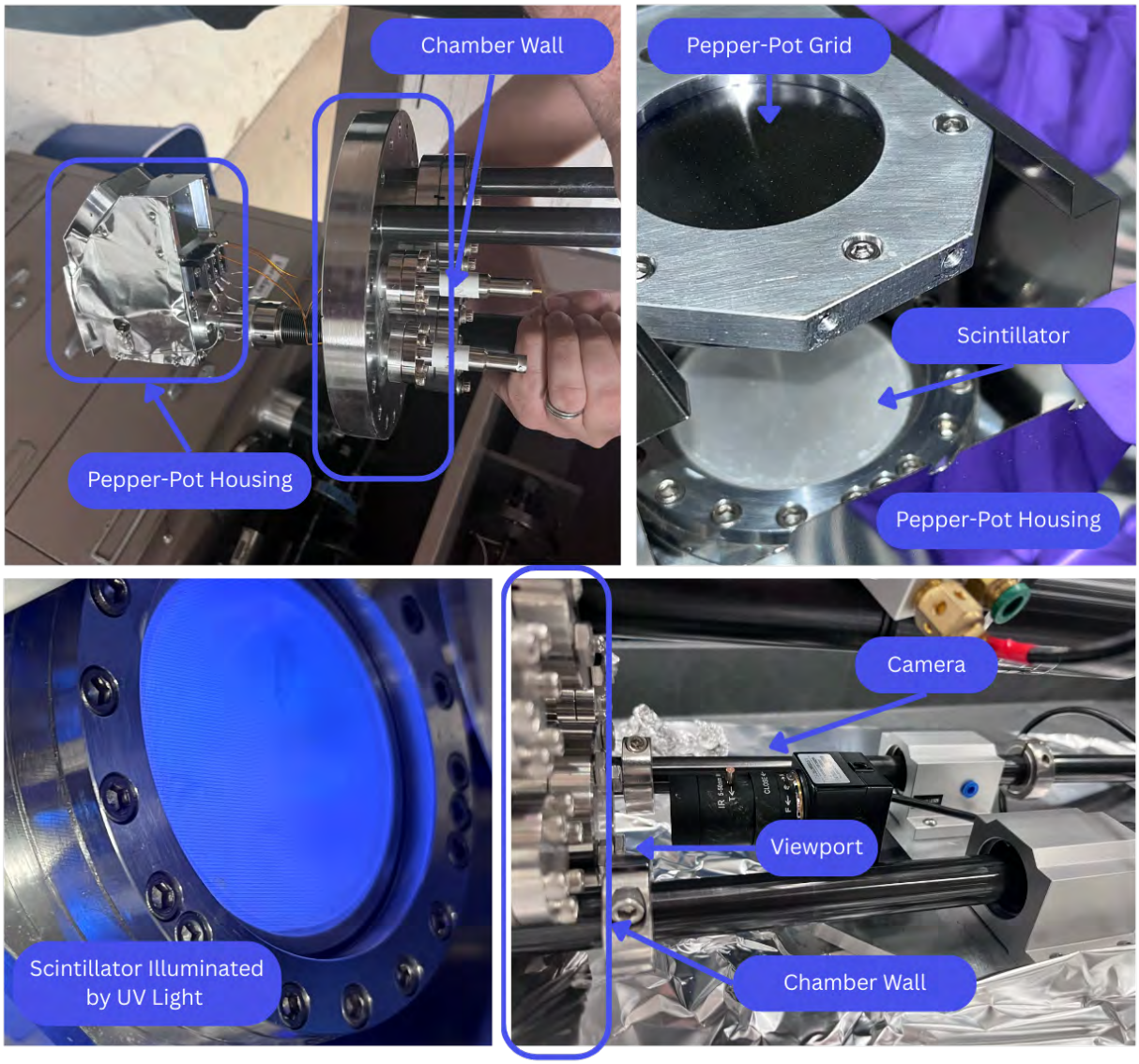


Figure 4.1 Setup showing Pepper-Pot device.

After propagating across a fixed drift length, L, these beamlets strike an imaging screen at position x_{screen}. Assuming the hole apertures are sufficiently small such that the particle's initial position is $x \approx x_{\text{hole}}$, the transverse divergence angle, x', of each beamlet is simply calculated as

$$x' = \frac{x_{\text{screen}} - x_{\text{hole}}}{L}. \tag{4.1}$$

The physical device is shown in Fig. 4.1.

The mask is a thin plate with precisely drilled holes. It must be thick enough to fully block the incident beam without sustaining damage, yet thin enough to act as an ideal 2D sheet and minimize particle scattering off the inner walls of the holes. The hole positions are accurately measured and calibrated. The apertures' sizes must also be carefully optimized. If the holes are too small, the transmitted beam intensity drops, leading to a decreased signal-to-noise ratio. If the holes are too large, neighboring beamlets may overlap on the imaging screen, complicating the analysis. Finally, the apertures can be arranged in various patterns, such as a grid, a single line, or a radial distribution. This thesis uses the grid configuration which offers the distinct advantage of capturing the full 4D transverse phase space, whereas other configurations can only capture the uncoupled 2D transverse phase spaces.

The beamlets pass through the apertures and travel a known drift distance, L, before striking the scintillator screen. The incident particles excite electrons within the scintillator material to a higher energy band. As these electrons de-excite and return to their ground state, they emit light in the visible spectrum. This light is then reflected by a mirror and captured by a camera. An assumption in this process is that the scintillator and camera system exhibit a linear response, meaning the measured pixel intensity is directly proportional to the number of incident particles. This linearity must be strictly verified. Scintillators can saturate under high beam intensities, and radiation damage can degrade their uniformity over time. Cameras may apply a gamma correction factor by default, which artificially alters the linearity of the recorded image intensity and must be accounted for or disabled.

Each beamlet illuminates a specific region on the scintillator, and each illuminated spot must be

accurately mapped back to its corresponding source aperture. Once the beamlets are correctly identified and assigned, their divergence is calculated using Eq. 4.1. The statistical analysis framework outlined in Sec. 2.1.4 is then applied to derive the beam moments, Twiss parameters, and transverse emittance. This also assumes that the statistical moments calculated from discrete beamlet sampling accurately represent the continuous parameters of the full beam.

The ATLAS pepper pot was developed in 2008 to measure the emittance of the low-energy DC beam extracted from the ECR ion sources [66]. While it is not currently used for routine operations, the recent initiative to implement automated beam tuning via ML/AI has renewed interest in this diagnostic. An accurate initial phase-space distribution into beamline simulations would aid the development of these advanced tuning algorithms.

4.1.2 Code Development

The original analysis software for the ATLAS pepper pot, initially written in FORTRAN, was later translated into Python. For this research, the code was advanced, benchmarked, and finalized for deployment in the control room. To demonstrate the core mechanics of this software, a simulated test image (Fig. 4.2) was generated using the TRACK beam dynamics simulation code, which included space-charge effects. In the simulation, an initial particle distribution was computationally filtered through a virtual mask featuring holes spaced 4 mm apart, center-to-center, in both the x and y directions. After filtering, the beamlets were tracked over a 100 mm drift length, expanding due to their inherent transverse momentum, before finally being recorded on a virtual screen with a spatial resolution of 0.1 mm/px.

The image can be analyzed using either a 1D or a 2D approach. In the 1D analysis, integrating the image intensity along the x- or y-axis generates a projection characterized by a series of peaks and valleys, as illustrated in Fig. 4.3. The position coordinates are defined by the known positions of the holes in the mask. To accurately assign the measured intensity to the correct source hole, boundaries are established between the projected peaks. It is assumed that all particles passing through a specific hole originate from the center of that hole. (This will be shown to cause a slight error, discussed in a later section.) Using Eq. 4.1, the difference between a pixel's position on the

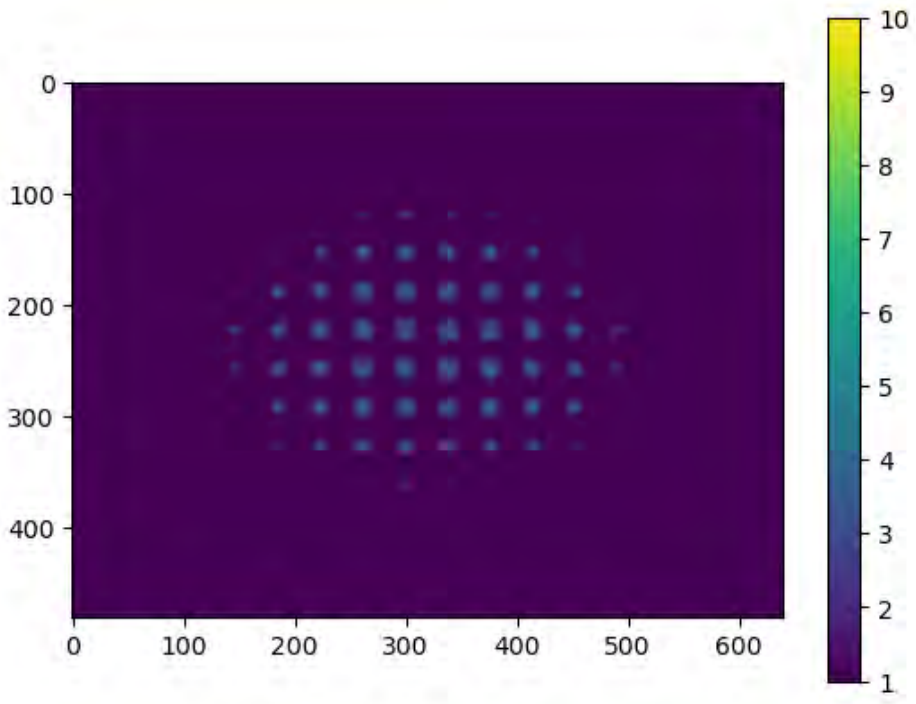


Figure 4.2 Test image of a pepper pot created in TRACK.

screen and its corresponding hole position determines the particle's divergence angle. The intensity at that pixel then serves as the statistical weight for the phase-space coordinate.

A practical consideration is whether the absolute transverse alignment of the mask relative to the imaging screen is strictly necessary. Because the mask forms a uniform grid, as long as the relative spacing between the holes is conserved, any global misalignment merely results in an artificial shift of the beam centroid. Since the Twiss parameters and transverse emittance are calculated relative to the beam centroid, they are fundamentally insensitive to these offsets.

In the 2D analysis, it is necessary to map each generated beamlet on the screen back to its corresponding source aperture. As with the 1D case, the uniformity of the grid means that identifying the absolute hole index is not strictly required, provided the relative spacing between the beamlets is accurately preserved.

To systematically identify the beamlets, the boundaries established during the 1D analysis are utilized. The intersections of these 1D dividing lines form a 2D bounding grid that encapsulates each individual beamlet. The center of each beamlet is then determined by calculating its intensity

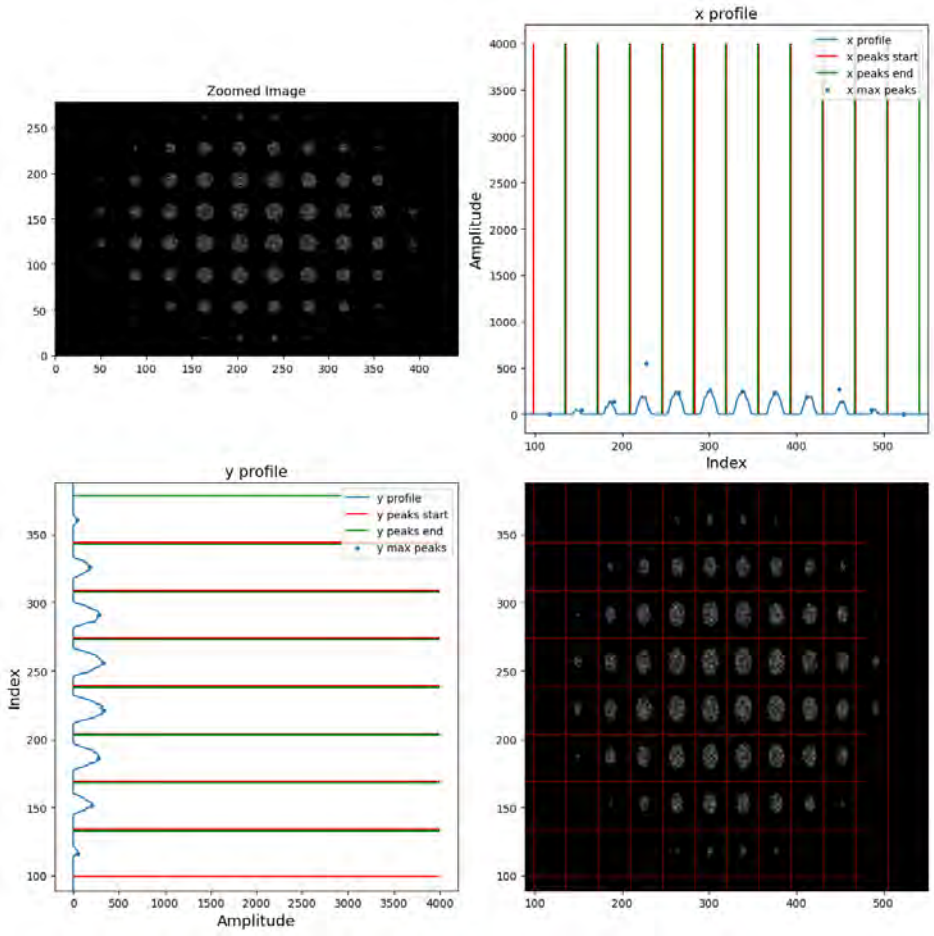


Figure 4.3 Beamlet center being identified using 1D analysis.

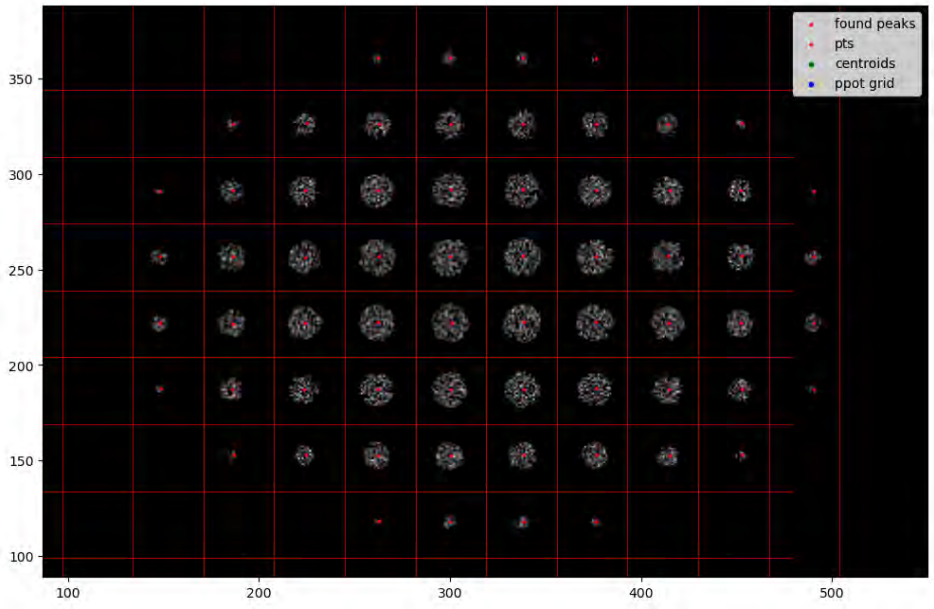


Figure 4.4 Beamlet center being identified using 2D analysis.

centroid. This bounding grid and the calculated centroids are illustrated in Fig. 4.4. Once every pixel within a bounded region is correlated to its specific source hole, the divergence angle can be calculated using Eq. 4.1, allowing the full 4D transverse beam parameters to be estimated.

The reconstructed parameters from both the 1D and 2D analyses are summarized in Table 4.1, demonstrating excellent agreement between the two methods.

	α		β [cm/rad]		ϵ_{rms} [cm·mrad]	
Method	x	y	x	y	x	y
1D	0.993	1.965	199.689	147.967	3.393	2.529
2D	1.004	1.983	202.517	149.606	3.312	2.478
Truth	1.000	2.000	200.000	150.000	3.330	2.500

Table 4.1 Comparison of Twiss parameters and emittance between 1D and 2D reconstruction methods against the ground truth.

4.1.3 Benchmarking

To rigorously benchmark the analysis code, hundreds of distinct Gaussian beam distributions were generated, propagated through a simulated pepper pot, and analyzed to compare the reconstructed parameters against the ground truth. The initial distributions were varied systematically with α_x and α_y from 0 to 0.4, β_x from 400 to 500, and β_y from 300 to 500, while the emittances were held constant at $\epsilon_x = 2$ and $\epsilon_y = 1$. These bounds were deliberately chosen to prevent both beam loss at the mask and beamlet overlap on the screen.

The simulation relied on an analytical calculation rather than a full tracking program for computational efficiency. After the mask, the transmitted beamlets carry sufficiently low current that space-charge forces during the drift are negligible. The simulated intensity images were then analyzed using the software to estimate α, β, and ϵ. The differences between these estimates and the true values were compiled into error histograms, presented in Fig. 4.5.

To evaluate the reconstruction accuracy, both absolute and percentage errors were calculated. The percentage error, Eq. 4.2, was applied to β and ϵ. Since α has values 0.0 to 0.4, the absolute error was used to prevent the numerical instability when $\alpha = 0$. All resulting estimation errors were contained within $\approx \pm 2\%$, which aligns well with the error tolerances reported in prior work [67].

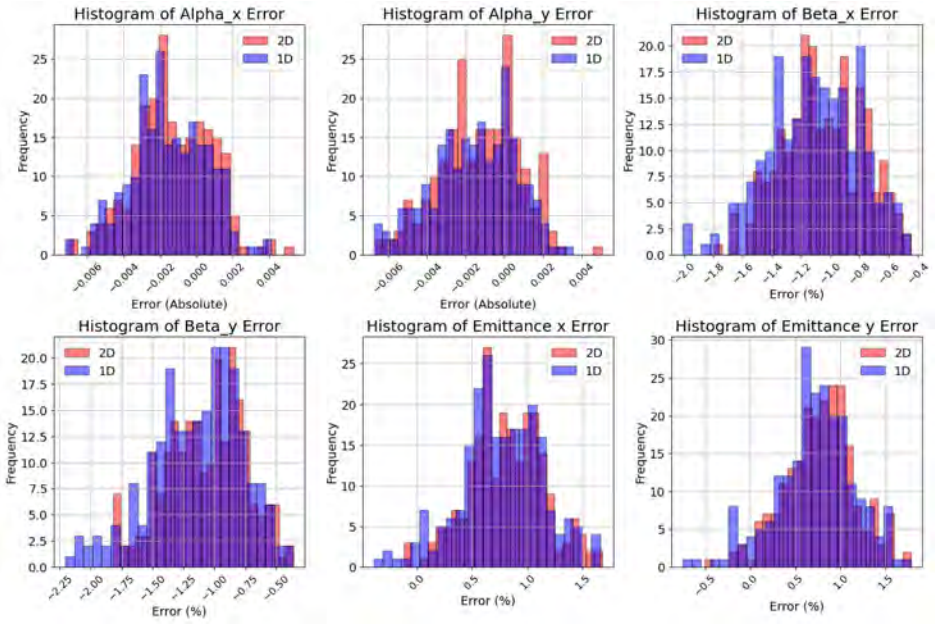


Figure 4.5 Histogram of Twiss parameter and emittance error from the Pepper Pot benchmark.

$$\% \text{ error } = \frac{\#\text{ experimental } - \#\text{ actual}}{\#\text{ actual}} \times 100 \tag{4.2}$$

Further inspection of the histograms reveals a systematic error, evidenced by the distributions' biases away from zero. This systematic bias is consistent with uncertainties reported in prior work [67].

To better understand the origins of this bias, an investigation was conducted to evaluate the effects of pepper-pot hole size and analysis grid resolution on reconstruction error. The core analysis assumes that every particle passes through the exact center of its assigned hole. Although this approximation holds for sufficiently small apertures, finite aperture dimensions inevitably introduce a systematic error [68]. As for the grid resolution, sometimes a finer resolution would reduce errors. These two variables were systematically adjusted in simulation, and their resulting effects on the analysis are illustrated in Fig. 4.6.

The figure shows that grid-resolution effects only become significant for values below 1500 px because, as resolution increases, the effect of hole size dominates. As the hole size decreases, the errors across all estimated parameters converge toward zero, confirming that finite aperture dimensions introduce a measurable systematic bias into the analysis. However, as previously noted,

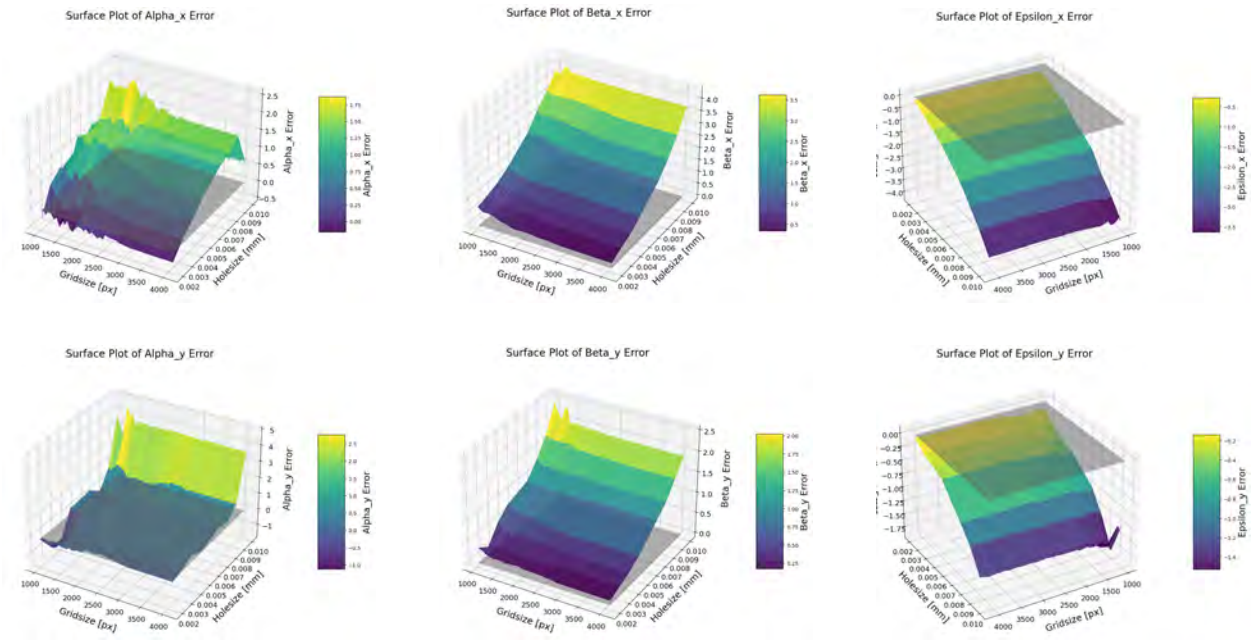


Figure 4.6 Analysis of error due to hole size and grid-size. The direction of decreasing hole size is not the same for each plot.

reducing the hole size simultaneously decreases the number of transmitted particles, thereby reducing the signal intensity on the scintillator. Future work could focus on developing a correction method to mitigate this inherent bias.

It is also possible that different distributions would give different results. Only Gaussian distributed beams were used for this analysis. Before concluding that hole size affects the error, various distributions should also be tested.

4.1.4 Image Processing

Experimental conditions rarely yield the idealized images depicted in Fig. 4.2. Bumps may cause the camera to rotate slightly. Noise due to image collection, secondary scattering, or an improperly light-sealed viewpoint may cause errors. This section evaluates various digital image processing techniques to address these issues. While not exhaustive, this analysis distinguishes strategies for mitigating or correcting artifacts under non-ideal operational conditions.

4.1.4.1 Rotation Correction

As seen in Fig. 4.1, the camera is out in the open and may be bumped or shaken, causing a slight tilt in the beamlet distribution, as seen in the left panel of Fig. 4.7. It is also possible that the

pepper pot grid itself is misaligned. Upon later inspection, the misaligned camera was found to be more likely. Thus, it is assumed that any observed tilt is due to the camera's alignment.

To correct this rotation, the image is computationally rotated within a bounded interval of $\pm 5^{\circ}$ using the golden-section search algorithm to progressively narrow the search interval until the optimal correction angle is reached. Proper alignment is achieved when the 1D intensity projection maximizes the peak amplitudes.

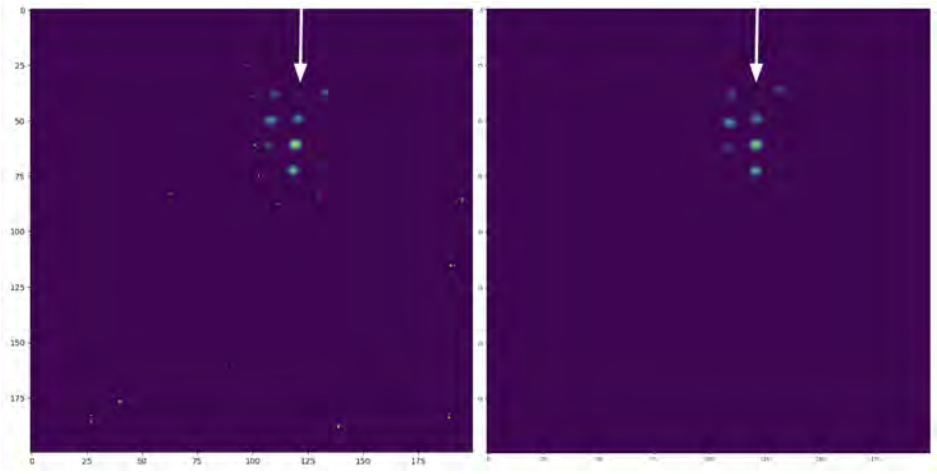


Figure 4.7 Comparison of the raw image (left) and the processed image (right) after rotation correction and noise filtering. The white arrow indicates a beamlet that remains unaffected by the noise filtering process.

4.1.4.2 Salt and Pepper Noise

Salt-and-pepper noise manifests as small, isolated pixels randomly distributed across the image. Because these artifacts do not represent the true beam distribution, they must be removed prior to analysis. While the exact causes of this noise have not been comprehensively studied in this specific setup, probable factors include background radiation from particle scattering and intrinsic thermal sensor noise. The diagnostic utilizes an ELP-USB8MP02G-SVF9(2.8-12) camera, which is not specifically designed for low-light or high-radiation environments. An example of this characteristic noise is visible in the left panel of Fig. 4.7. The goal of this processing step is to eliminate these artifacts while preserving the fidelity of the original beam signal.

To evaluate noise removal techniques, a comparative study was conducted using a simulated beam image. Artificial salt-and-pepper noise, generated by randomly setting high-intensity pix-

els across the image, was superimposed on the simulation, visually matching the noise profiles typically observed during experimental runs.

Initially, a standard Median Filter was employed. This technique replaces each pixel's value with the median intensity of its neighbors. Various kernel radii can be applied, but a standard radius of 1 pixel was used. This approach is highly effective at eliminating high-frequency noise. It introduces a smoothing effect across the entire image, potentially artificially increasing the measured beam size.

Consequently, an "Isolated Pixel" filter was developed to eliminate noise without smoothing the underlying macroscopic beam structure. Because salt-and-pepper noise typically appears as isolated pixels, this algorithm scans the image to identify any pixel surrounded by dark background pixels. Once identified, these artifacts are zeroed out. To account for small clusters of noise, this filter can be applied iteratively until all such artifacts are eliminated.

Parameter	**No Noise**	**Noise**	**Median Filt.**	**Iso. Pixel**	**Med + Iso**
$\Delta\alpha_x$ (abs.)	-2.5	3.5	-3.0	-8.5	**-2.5**
$\Delta\alpha_y$ (abs.)	-4.0	-35.0	-7.5	**-2.5**	-7.5
$\Delta\beta_x$ (%)	-1.0	-5.67	-2.0	**-1.67**	-2.0
$\Delta\beta_y$ (%)	-2.67	-8.67	-5.67	**-4.33**	-5.67
$\Delta\epsilon_x$ (%)	1.0	14.0	-4.0	**2.0**	-4.0
$\Delta\epsilon_y$ (%)	2.0	24.0	**-0.2**	3.4	**-0.2**

Table 4.2 Comparison of relative errors for Twiss parameters and emittance under various noise reduction techniques. Bold text indicates the best result.

Table 4.2 presents a comparison of the reconstruction errors resulting from the raw noise. The results are compared directly against the ground-truth values for the various filtering techniques. The data indicate that untreated salt-and-pepper noise introduces non-negligible errors. Overall, the Isolated Pixel filtering method outperforms standard median filtering across most beam parameters. The median filter's inherent smoothing appears to introduce a slight systematic error in the divergence profile. Based on these results, the Isolated Pixel filter will be the primary method employed for noise reduction whenever feasible.

4.1.4.3 Major Noise and Artifacts

It is also possible for significant noise to occur, for which the previous methods are insufficient. The first significant source of noise is ambient light entering the camera viewport. To mitigate this physically, optical shielding, such as aluminum foil, is applied around the camera and viewport assembly. However, other unknown sources of noise are observed as shown in Fig. 4.8.

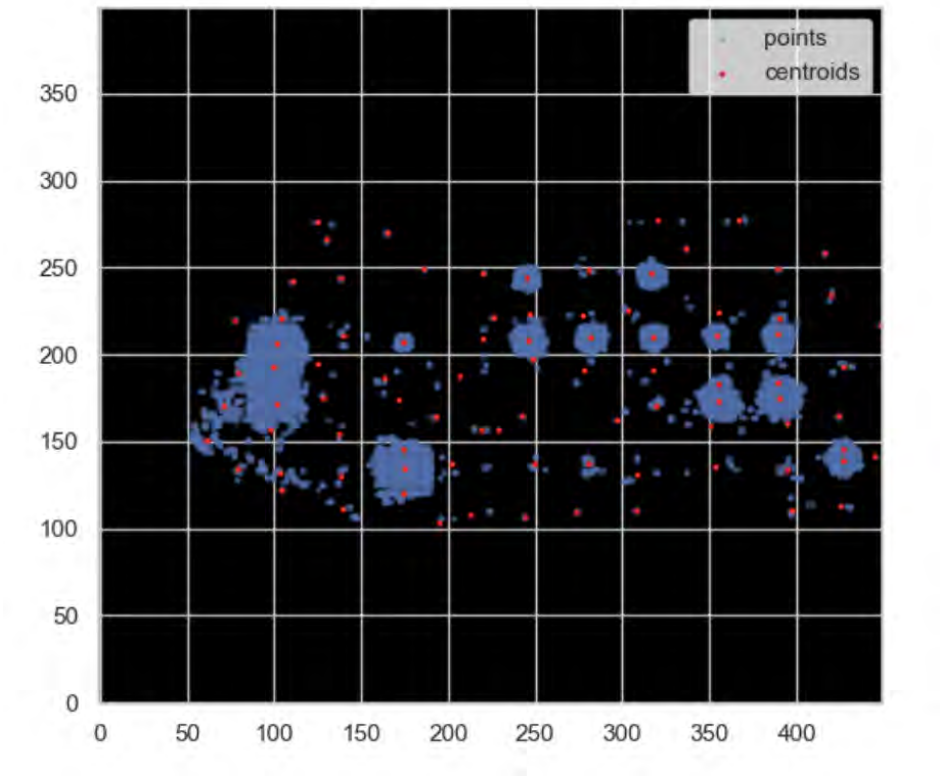


Figure 4.8 Example of an experimental pepper-pot image degraded by severe background noise and optical artifacts. Under these conditions, the algorithm fails to accurately identify the centroids of the individual beamlets.

To computationally mitigate this issue, two distinct thresholding methods are employed. The first approach is grid-based thresholding. As established during the 2D analysis, the 1D projection algorithm identifies the boundaries of each intensity peak along both the x and y axes, creating a bounding grid that encapsulates each individual beamlet. A user-defined parameter, *threshold_detect*, is then introduced. The algorithm calculates the pixel sum within each grid cell

and registers a valid beamlet only if the local sum exceeds the *threshold_detect* value. This parameter is manually adjusted to ensure actual beamlets are preserved while rejecting localized noise clusters. The default value is 0, but can be adjusted by eye until the noise disappears. It would be beneficial to automatic this selection in the future.

When localized grid-based thresholding is insufficient, global direct thresholding is used. In this method, the original image is processed by setting all pixel intensity values below a specified global threshold to zero. While this technique serves as an effective fallback for highly contaminated images, it inherently induces analytical inaccuracies. Specifically, global thresholding truncates the low-intensity tails of the beamlet distributions, artificially reducing the calculated statistical moments and, consequently, underestimating the beam emittance.

As an initial compromise, the default software settings are designed to automatically correct rotational misalignment and remove isolated noise while rigorously preserving the essential features required for accurate phase-space analysis. For fine adjustments in non-ideal experimental conditions, grid-based and direct thresholding provide necessary manual controls.

4.1.5 Implementation and Difficulties

Following the verification of the analysis algorithms, the software was integrated into the ATLAS control system in collaboration with the controls group. This integration process required several software modifications, including removing incompatible Python packages. To evaluate the robustness of the image processing code, previously acquired experimental images were analyzed. Although the absolute accuracy of the pepper-pot diagnostic cannot currently be independently verified, a quadrupole scan is proposed as a method for cross-validation of the emittance measurements.

To evaluate beam stability, a sequence of five images was captured at intervals of 4 to 5 seconds, as depicted in Fig. 4.9. The analysis of these images, presented in Table 4.3, give an indication that the beam parameters remained stable throughout the observation period. Further analysis to understand how uncertainty of the algorithm is needed before a judgement on the stablity can be made.

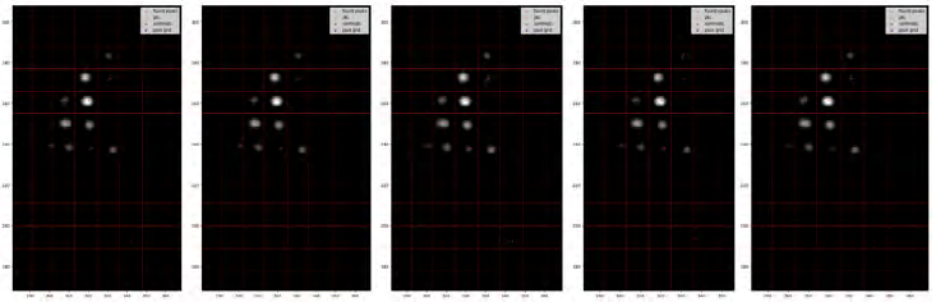


Figure 4.9 Testing beam stability. Images are taken 4-5 seconds apart.

Method	α_x	α_y	$\epsilon_{rms,x}$ [cm·mrad]	$\epsilon_{rms,y}$ [cm·mrad]	β_x [cm/rad]	β_y [cm/rad]
1	7.82	23.13	0.22	0.36	22.24	67.32
2	7.62	16.98	0.20	0.26	21.95	49.51
1	9.58	23.41	0.31	0.39	27.91	68.12
2	8.99	21.23	0.29	0.35	26.38	62.11
1	9.62	23.43	0.30	0.38	28.04	68.21
2	8.46	20.21	0.25	0.32	23.86	58.81
1	8.66	21.67	0.24	0.33	24.40	63.05
2	7.79	17.01	0.20	0.26	22.54	49.50
1	8.27	23.03	0.24	0.37	23.27	67.02
2	7.80	19.64	0.22	0.31	22.16	57.13
Mean	**8.46**	**20.97**	**0.25**	**0.33**	**24.27**	**61.08**
Std Dev	0.74	2.48	0.04	0.05	2.36	7.22

Table 4.3 Measured beam parameters using 1D and 2D analysis under the same initial condition. Each group correspond to an image.

Subsequent attempts to acquire data were unsuccessful due to a degradation of the imaging system. During live adjustments, the scintillator's light yield decreased rapidly, fading completely within a few minutes. Although increasing the beam current momentarily restored the signal intensity, it decayed shortly thereafter.

A review of the literature [69, 70, 71, 72] suggests that radiation damage is the primary cause of this phenomenon. Scintillation occurs when incident particles excite electrons into the conduction band, which subsequently emit photons upon de-excitation. The scintillator material is assumed to be thallium-doped cesium iodide [CsI(Tl)]. Intense ion bombardment can induce radiation damage by creating a color center that renders the crystal opaque. Consequently, prolonged beam exposure

likely caused the scintillator to lose its transparency, preventing further detectable light emission.

This hypothesis was confirmed upon physical inspection. After the pepper-pot assembly was removed and illuminated with an ultraviolet (UV) source, the scintillator failed to fluoresce. This lack of response indicated permanent material damage, necessitating a replacement. To prevent similar degradation in future experiments, it is recommended to implement a fast beam shutter to strictly limit the scintillator's exposure time to the incident beam.

4.1.6 Future Outlook

An analysis framework was developed and verified to process images acquired from the pepper-pot device for the estimation of Twiss parameters and beam emittance. The software is written in Python and adheres to standard development practices, including version control via GitLab, a verification test suite, and comprehensive documentation.

To enhance operational robustness, the 2D-based algorithm was modified to identify individual beamlets and correlate them with specific mask holes using a 2D grid from the 1D analysis algorithm. Additionally, various noise-filtering methods were investigated and implemented to improve data fidelity.

Future work will focus on replacing the degraded scintillator and integrating the diagnostic system into the facility's control infrastructure. This integration requires automating various parameters for image acquisition and noise-filtering routines to enable real-time monitoring. While an error analysis module based on [58] has been implemented, it requires measurements of the current uncertainties.

To further advance pepper pot theory, subsequent studies should investigate the systematic effects of finite hole sizes and develop corresponding correction algorithms. Moreover, emerging techniques, such as the virtual pepper-pot method [73], are being explored to mitigate the beamlet overlap effect.

4.2 Phase Space Reconstruction in the PII-to-Booster Line

While the previous section detailed the implementation, challenges, and limitations of a direct measurement device, this section shifts focus to indirect methods for phase-space reconstruction.

It begins by describing the front section of the ATLAS beamline and providing an overview of the existing instrumentation and diagnostics. Next, the methodologies for data collection and processing are outlined. The analysis results from the quadrupole scan and MENT tomography are then presented to highlight the advantages of using a non-Gaussian-reliant diagnostic. Finally, various assumptions and limitations are evaluated—including the effects of space charge, finite apertures, and beam stability—before the chapter concludes with potential directions for future research.

4.2.1 PII-to-Booster Transport Line Overview

The experiment was conducted at the ATLAS PII-to-Booster Transport Line (see Fig. 4.10) concurrently with a study validating Bayesian optimization algorithms for automated beamline tuning.

The ATLAS beam originates in the Electron Cyclotron Resonance (ECR) source, is transported through the Low Energy Beam Transport (LEBT), and finally accelerates through the Radio-Frequency Quadrupole (RFQ). This experiment focuses on the section between the PII and the Booster. The relevant beam parameters are listed in Table 4.4.

Atomic Mass	40
Charge state	8
Energy per amu	1.3 MeV/u
Current	200 nA

Table 4.4 Beam state of ATLAS during the PII-to-Booster transport line.

Selecting the optimal tomography location involves balancing several competing factors. Primarily, tomography requires sufficient manipulation of the phase space using magnetic elements. While the precise threshold for sufficient manipulation is not fully understood, a general guideline is to maintain enough flexibility to rotate the phase space by at least 180°. On the other hand, as the beamline length increases, realistic effects such as component misalignment, component lengths, calibration, and magnetic field errors must be modeled with increasing precision. To ensure reliability, it is desirable to minimize these error sources by shortening the transport section. Finally, because ATLAS is a user facility, performing experiments at lower energies is preferred to avoid

disruptions caused by beamline damage.

Balancing these considerations, manipulability versus potential error accumulation and operational stability, the front end was selected. This section features a quadrupole magnet located approximately 1.85m upstream of a Beam Profile Monitor (BPM), with six additional quadrupoles upstream to facilitate the necessary phase-space manipulations. They are denoted as “QDP" for doublets or “QSP" for singlets.

Since the beam is typically off-axis, dipole steering magnets (denoted as “STP") are employed to correct orbit deviations and feed-down effects of quadrupoles.

This beamline imposes two primary limitations. First, there is only one usable beam profile monitor, “PMP303", which is located near the end of the beamline. Second, the absence of coupling beam elements or coupling diagnostics limits the analysis. This means that dynamics in the x and y planes occur independently, so only 2D reconstruction is possible. 4D reconstruction is not possible, as noted in Sec. 2.3.2.

4.2.2 Beamline Instrumentation and Diagnostics

To characterize the beam dynamics required for this analysis, two primary diagnostics were employed. Beam Profile Monitors (BPM) for projection measurements, and Faraday cups for beam transmission.

The transverse phase space is reconstructed primarily from data provided by the BPM, specifically the Model BPM81 shown in Fig. 4.11. This device operates by intercepting the beam with a single wire in the form of a 45° helix rotating within the vacuum chamber. As the wire scans through the beam, secondary electron emission generates a current proportional to the local beam density. This helical geometry allows a single device to produce two orthogonal profiles, horizontal and vertical, integrating the beam distribution into 1D projections. While several BPMs appear on the beamline schematic (labeled “PMP” in Fig. 4.10), only PMP303 was utilized to obtain the profile measurements for this experiment.

Faraday cups are employed to measure the total beam current. An example is shown in Fig. 4.12. This model is designed to maximize clearance within the beam pipe, enabling insertion or

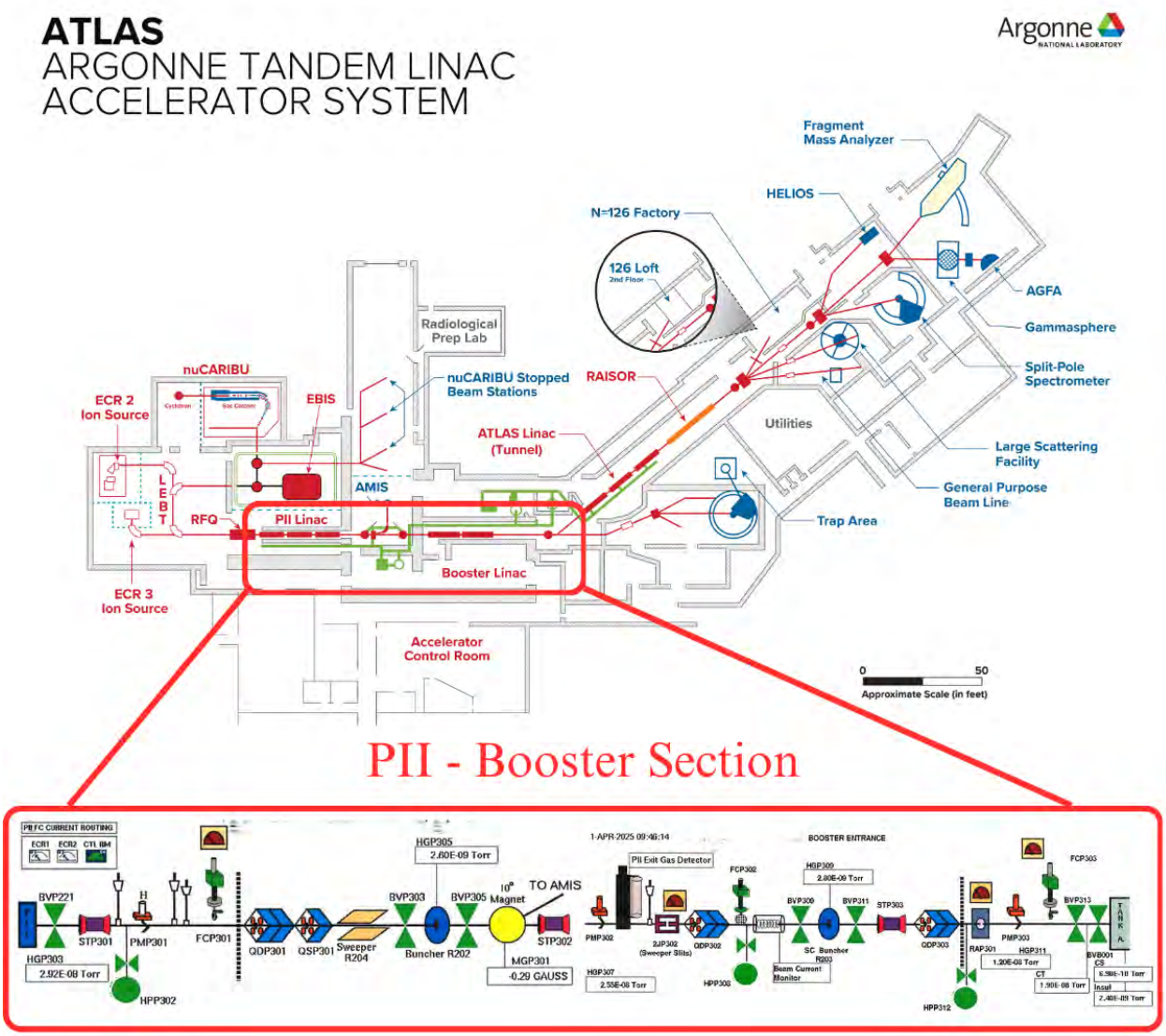


Figure 4.10 ATLAS Facility and the PII-Booster section. Quadrupole doublets are denoted as “QDP" while singlets are “QSP". Faraday cups are “FCP", steers are “STP", and beam profile montiors are “PMP".

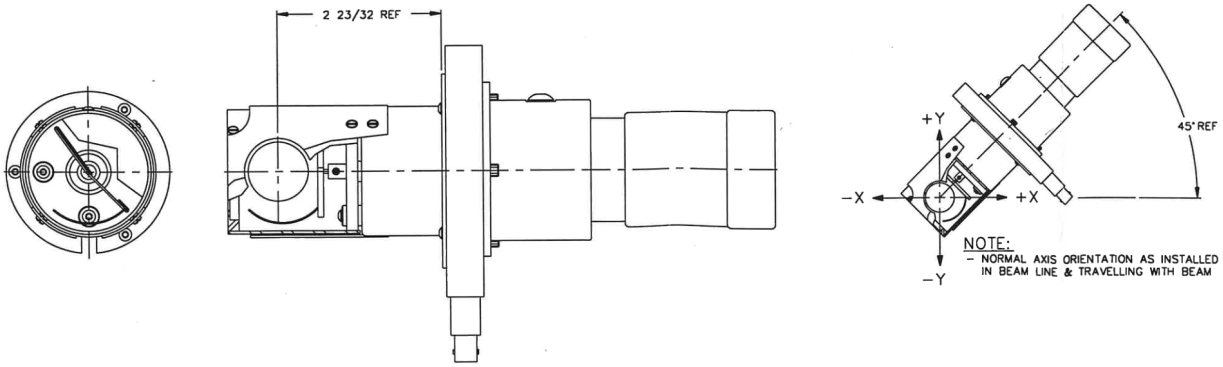


Figure 4.11 Schematic of Beam Profile Monitor model BMP81 (Integral Collector) [74].

Figure 4.12 Image of FC50 Faraday Cup [75].

retraction as needed. Beam transmission and loss are calculated by comparing current readings between two Faraday cups. The upstream cup establishes the reference current, while the downstream cup measures the remaining current. This section contains three Faraday cups, labeled “FCP”. FCP301 and FCP303 cups are used to determine beam loss for each measurement.

The accuracy of the beam transmission is affected by the geometric acceptance and secondary electron emission. The aperture of the Faraday cup is smaller than the beam pipe diameter. The beam envelope may exceed the cup’s clear aperture. This "grazing" effect leads to measurement errors. For instance, if the beam grazes the upstream reference cup (FCP301) but is fully captured downstream, the calculated transmission will exceed 100%. These data points were identified as nonphysical and excluded from the analysis. Conversely, if the beam grazes the downstream cup (FCP303), the measured current will underestimate the true transmission.

Second, the impact of secondary electron emission, which can increase current readings, is mitigated by the hardware design. A negative bias voltage is applied to the Faraday cup to repel secondary electrons back into the collector.

These effects were not considered for this thesis. Regimes in which the beam grazes the Faraday cups result in significant beam loss, which is marked as "unusable" in the subsequent analysis.

4.2.3 Data Acquisition and Processing

Initially, beamline configurations were simulated and tested; however, the initial tests yielded insufficient beam transmission. Consequently, various alternative strategies were employed to obtain usable data. They are explained here and summarized in Table 4.5.

1–3. The first few datasets resulted in high beam loss, rendering them unusable for accurate analysis. Although a simulation model of the ideal lattice was used to generate configuration points intended to maximize transmission and angular spread, these configurations resulted in negligible transmission when applied experimentally. Factors such as inaccurate modeling, quadrupole misalignments, and beam offsets likely contributed to this significant loss.

   In subsequent trials, a stable beam configuration was identified, and the quadrupoles were adjusted within a narrow range relative to this initial baseline. These trials helped establish the safe operational bounds for datasets 6 and 7.

4. For the fourth dataset, the strategy was simplified. A quadrupole scan was performed to estimate the beam phase space. This method is straightforward and serves as a benchmark for the MENT analysis. The scan was performed on the 6th quadrupole, with all other elements held constant. This approach yielded a sizable set of measurements with beam transmission exceeding 80%.

5. A second quadrupole scan was performed on the 7th quadrupole, keeping all other elements constant. This provided an additional independent dataset for analysis and comparison.

6. During an extended operational window, a dataset of approximately 2000 points was collected. Based on the previous datasets, a "safe" operational range was derived to minimize beam loss. Accordingly, adjustments were limited to the last four quadrupoles and two steerers.

7. Using the data obtained from the 6th run, an Autoencoder-decoder analysis was utilized to compress the data into a 2D latent space. This effectively separated the high- and low-

transmission cases, as illustrated in Fig. 4.13. Values in the high-transmission region were sampled along the latent dimension, allowing the decoder to generate a new set of configuration points with improved overall transmission across the extended operational window.

Dataset	Strategy	Quad Used	Notes
1–3	Initalization	All Quads	High beam loss due to modeling inaccuracies and misalignments. Helped establish a stable baseline.
4	Quad scan	Quad 6	Achieved measurements with beam transmission exceeding 80%.
5	Quad scan	Quad 7	Provided an additional, independent dataset for analysis and comparison. Less modeling error due to shorter length.
6	Safe scan	Quad 4-7	Samples were taken inside bounded ranges to minimize beam loss.
7	Autoencoder-decoder generation	Quad 4-7	Utilized Dataset 6 and autoencoder-decoder latent space to generate high-transmission sample. Improved transmission.

Table 4.5 Summary of experimental datasets, configuration strategies, and resulting beam transmission outcomes.

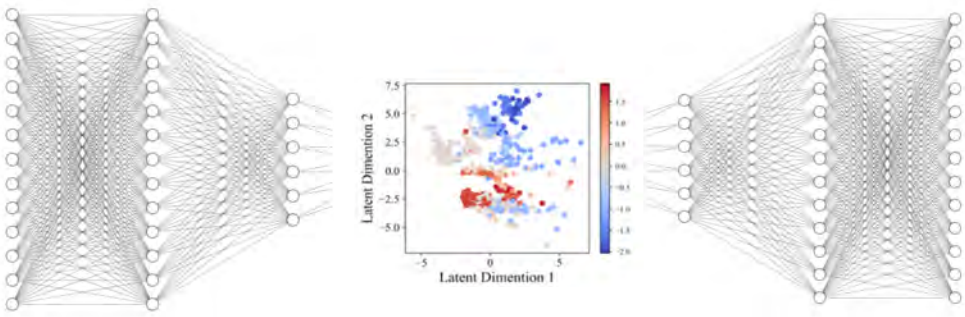


Figure 4.13 Cartoon of latent dimension.

The preceding analysis will utilize datasets 5 and 7 to keep the report concise. Set 5 was chosen because it has the least error, as its analysis uses the shortest beam length. Set 7 represents the set with the most beam manipulation while maximizing beam transmission. These two set have the best chance of producing an accurate reconstruction.

4.2.3.1 Data Cleaning and Post-processing

Each measurement comprises a set of monitored quadrupole and steerer values, Faraday cup readings, beam profile monitor calibration data, and timestamps. Derived quantities, such as beam transmission, are calculated by comparing the currents from the upstream and downstream Faraday cups.

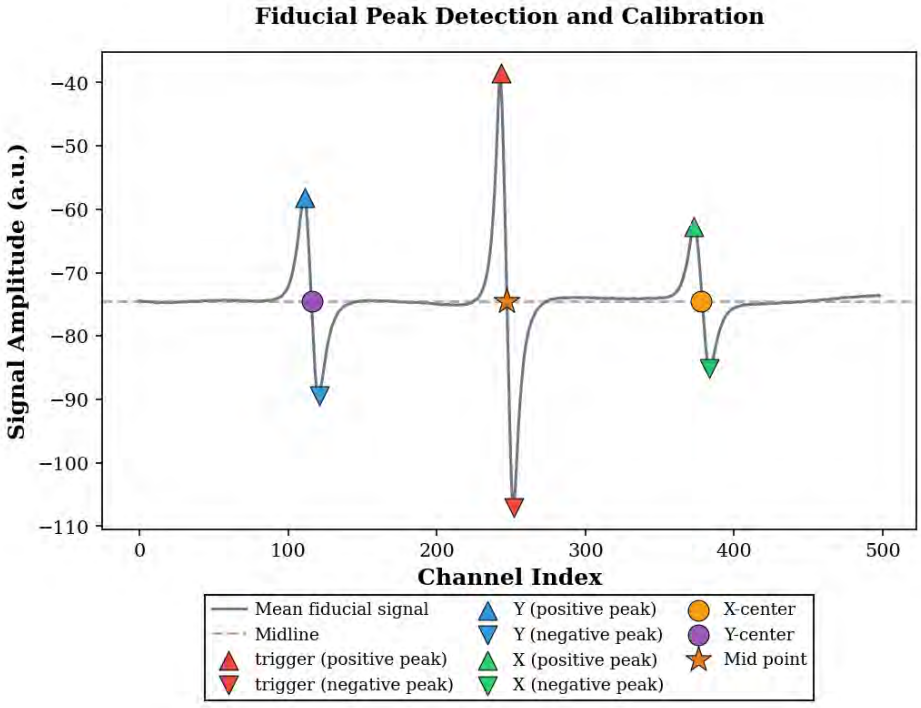


Figure 4.14 The algorithm identifies the highest-magnitude trigger, followed by the second highest peaks for the vertical (y) and the third highest for the horizontal (x) beam profiles. The identified zero-crossing centers are used to establish the 6 cm calibration distance between the x and y profiles

The beam profiles are not provided directly as spatial distributions but as time-domain signals containing fiducial markers. The signal exhibits distinct peaks. The highest-magnitude peak serves as the trigger; the second-highest corresponds to the vertical (y) profile; and the third-highest corresponds to the horizontal (x) profile. The center of each profile is defined by the horizontal crossing point, and the distance between the x and y horizontal crossings corresponds to the physical calibration distance of 6 cm. Figure 4.14 illustrates the analysis code used to identify these relevant peaks and centers. An example of an extracted beam profile is shown in Fig. 4.15.

Following extraction, the profiles were normalized and mean-centered. To mitigate noise while

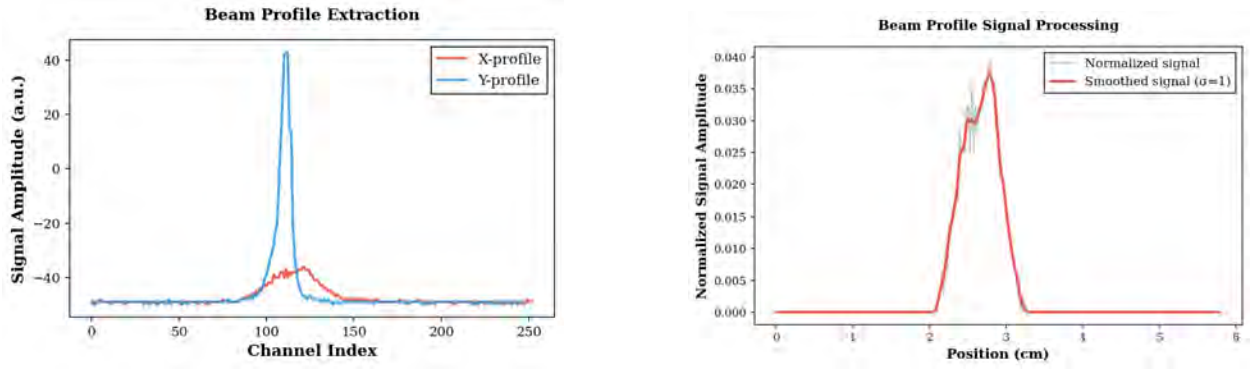


Figure 4.15 Example of beam profile extraction (left), and an example of beam smoothing (right).

retaining the structural details required for reconstruction, a Gaussian filter with a standard deviation of 1σ was applied. This value was chosen to balance the smoothing of high-frequency noise with the preservation of the unique profile shape. The resulting smoothed profile is presented in Fig. 4.15.

Transfer matrices were obtained for later analysis by simulating the lattice using codes such as TRACK [27] or Accelerator Toolbox [76]. The analysis utilizes the monitored magnet readbacks rather than the setpoints. Due to settling times, these values may differ, and the monitored values more accurately represent the machine's actual state at the time the measurement was taken. In addition, the magnet values were the raw current readings in Amps and must be converted into quadrupole gradients or steering angles before being input into the simulation. The conversion formulas are provided in Eq. 4.3, and the corresponding calibration factors are listed in Table 4.6. Here, $B\rho$ is the magnetic rigidity, B_p is the pole-tip field, and r_p is the beam pipe radius:

$$K = \frac{B_p}{(B\rho)r_p}, \quad \theta_k = \frac{\int B_k \, ds}{B\rho} \approx \frac{B_k L}{B\rho} \tag{4.3}$$

Table 4.6 Magnet properties for PII-Booster.

Parameter	**Value**	**Unit**
Amps to Field (B_p)	1.457×10^{-2}	T/A
Amps to BL	1.112×10^{-3}	T·m/A
Pipe radius (r_p)	0.03175	m

4.2.4 Quadrupole Scan Analysis

A quadrupole scan was performed first to obtain an initial estimate of the Twiss parameters and beam emittance. This estimate serves as a baseline to help guide the subsequent tomographic reconstruction algorithms.

Initially, the data were processed by fitting each profile to a Gaussian curve; however, it was observed that the y profiles shown in Fig. **??** sometimes exhibited a small tail, making the distribution asymmetric. This led to a slight inaccuracy, so the data was instead fitted to a smoothed curve, which better retains the original characteristics of the measured profiles.

Furthermore, the dataset was filtered to include only measurements with beam transmission greater than 80%. This limit was established as the minimum threshold required for an accurate analysis, and its justification is detailed in Sec. 4.2.6.2.

Additionally, a noise floor is present, as seen in Fig. **??**, which artificially increases the calculated emittance. A noise threshold was implemented to remove this background signal. The threshold value was carefully chosen to minimize statistical inaccuracy while retaining the bulk of the beam profile. Because of the relatively high noise floor, this filtering resulted in a sharp cutoff at the tails of the distribution shown in Fig. **??**. Finally, the parameters were extracted from the remaining filtered data using the Linear Least Squares (LLSQ) method.

To evaluate the fit, two visualization methods are employed. The first shows a fitting of a Gaussian approximated phase space, while the second compares the measured and reconstructed spatial profiles.

Figure 4.16 show the phase space fitting, the calculated Twiss parameters, and the RMS fitted emittance. Vertical lines at the measurement location bound the RMS beam size at $x_1 = \pm\sigma_x$. To map these boundaries back upstream, the inverse of the linear transfer matrix, M^{-1}, is utilized:

$$\begin{pmatrix} x_0 \\ x_0' \end{pmatrix} = M^{-1} \begin{pmatrix} x_1 \\ x_1' \end{pmatrix} \tag{4.4}$$

Computationally, this back-projection is performed by selecting two points along the vertical boundary line at the measurement location. These points are multiplied by the inverse matrix M^{-1}

to determine their corresponding coordinates in the initial phase space. A new linear equation is then derived from these back-propagated points to construct the bounding lines at the initial location. By overlaying the back-projected bounds from all quadrupole settings, a void is created that constrains the rms initial phase space distribution. A red ellipse shows the rms fit. This provides insight into the beam size and is an effective way to visualize any errors.

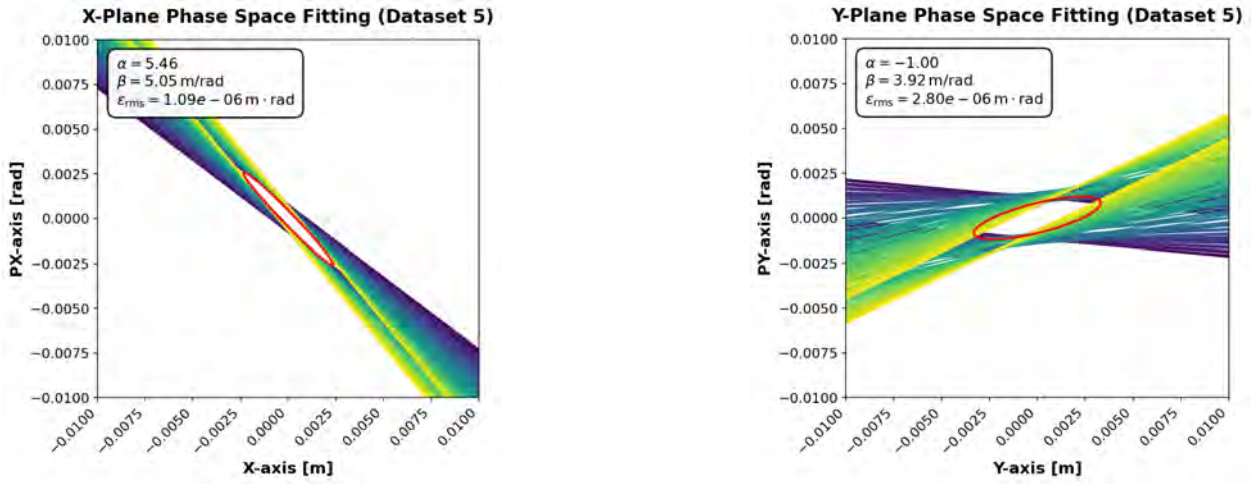


Figure 4.16 Phase space fitting for Dataset 5 for the X-Plane (left) and Y-Plane (right).

Figure 4.17 compare the measured and reconstructed profiles. To generate the reconstructed profiles, a Gaussian beam approximation was initialized using the estimated Twiss parameters and propagated through a lattice corresponding to each measurement. The fit accuracy is quantified using the Mean Squared Error (MSE). These results highlight the inherent limitations of the quadrupole scan method. Since the reconstruction relies solely on second-order moments and assumes a Gaussian distribution, it fails to capture subtle non-Gaussian features in the measured profiles. This motivates beam tomography as an alternative method to increase reconstruction fidelity.

4.2.5 MENT Tomography Analysis

Tomographic reconstruction will be used to recover the phase-space fidelity typically lost in standard quadrupole-scan analysis, which relies only on beam moments rather than full-profile details. The MENT algorithm (Sec. 2.3.5) is our chosen method for this experiment. Because our measurements are non-uniform, traditional FBP yields unwanted artifacts. While both SART and

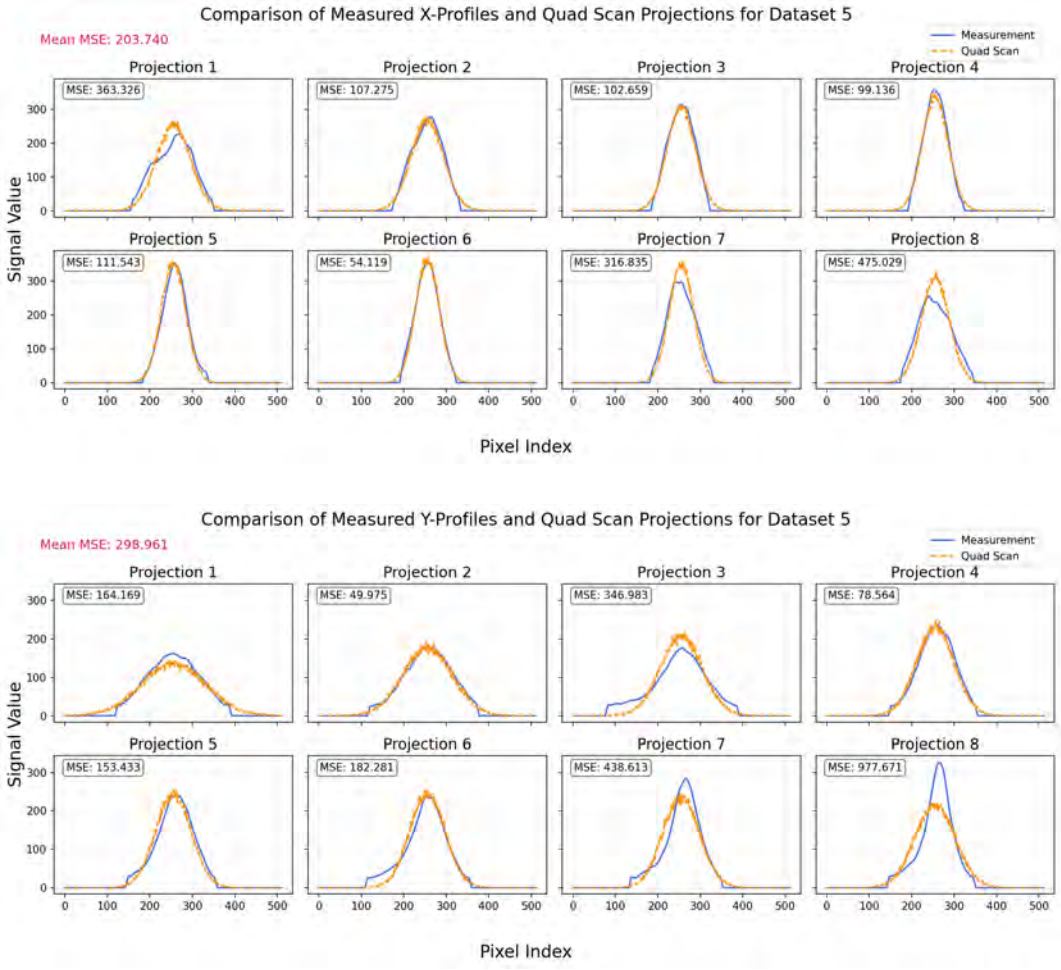


Figure 4.17 Comparison of measured profiles and reconstructed profiles for quadrupole scan for Dataset 5 in the X-plane (top) and Y-plane (bottom).

MENT successfully mitigate these issues for non-uniform projections, MENT is favored to avoid the curse of dimensionality inherent to SART. Also, from experience, MENT has been shown to be more robust to artifacts than SART. A publicly available software for MENT [77] was used for this analysis.

Dataset 5, the quadrupole scan set, will first be analysed and compared against the quadrupole scan analysis to show the advantages of MENT. Then, Dataset 7 will be used to showcase the limits of this method.

Initial attempts to apply MENT to the experimental dataset prove more difficult than expected. Techniques such as effective visualization and an initial estimate using the quadpole scan method helped to troubleshoot errors. Reconstruction in the normalized phase space has also shown im-

proved accuracy, and a selection algorithm was utilized to enable the first tomographic phase-space reconstruction at ATLAS (Fig. 4.18).

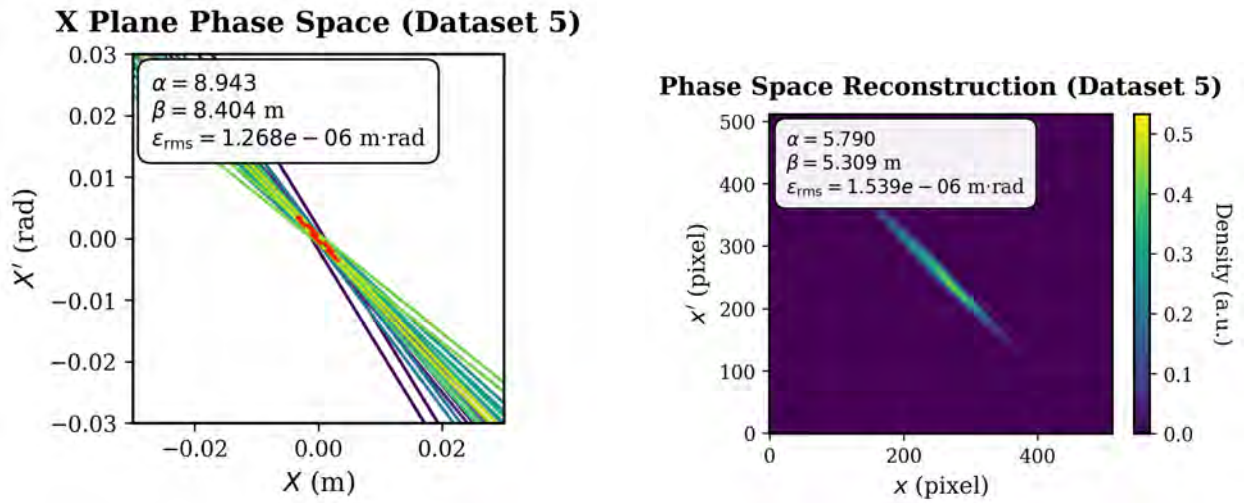


Figure 4.18 Comparison of Quad Scan estimate (left) and MENT reconstruction (right) for Dataset 5.

Eventually, a workflow was developed to effectively process the measurements and produce an accurate reconstruction.

1. The initial analysis utilizes a quadrupole scan, similar to the one described in the previous section, but relies on only 10 measurements in total. Because the goal is merely to obtain an initial estimate, using the full dataset is unnecessary. The measurement to choose should be high-quality data, date with high transmission and angular spread in real space. Additionally, if implemented for real-time analysis, this approach significantly reduces the number of measurements required for a scan, as shown in Fig. 4.18.

2. The previous initial estimate was required to compute the reconstruction in normalized phase space. For this step, a new 10-measurement subset was selected, specifically optimized for high beam transmission and a large normalized angular spread. Figure 4.19 visualizes this spread. Ideal data would be evenly distributed around the circumference of the circle.

3. These selected measurements are then processed using MENT to reconstruct the 2D phase space seen in Fig. 4.18.

4. Step 3 yields a new set of Twiss parameters, which may differ from those obtained in Step 1. Steps 2 and 3 are then repeated iteratively until convergence is reached for the Twiss parameters. While convergence typically occurs within two iterations, this can vary depending on the dataset.

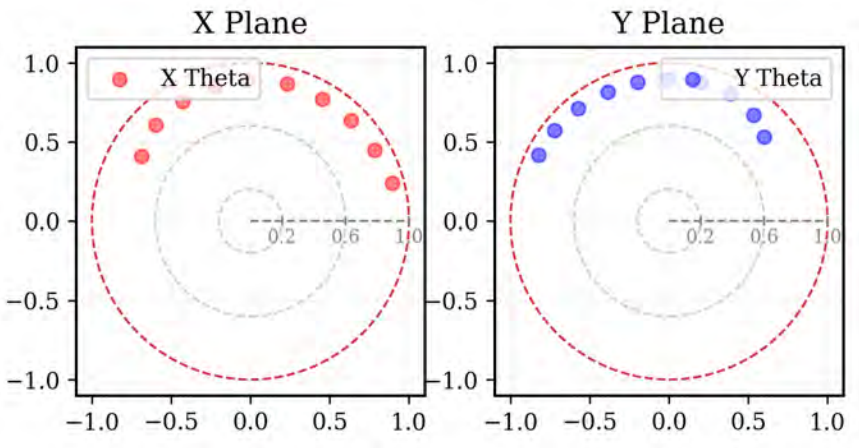


Figure 4.19 Distribution of angles in phase advance. The radial distance denotes the percentage beam transmission.

The measured and predicted profiles can be compared to assess the reconstruction's accuracy. Unlike the quadrupole scan, which assumes a Gaussian beam distribution, the actual reconstructed distribution can be used here. By setting this as the initial distribution and propagating it downstream, the resulting simulated profiles should match the measured ones. As seen in Figs. 4.20 and 4.21, the profiles show excellent agreement. Although different profiles were utilized, they clearly reconstruct the non-Gaussian features of the beam. A direct comparison is provided in the Appendix in Fig. B.1. Additionally, the error metric was changed from Mean Squared Error (MSE) to Mean Absolute Error (MAE) to reduce the magnitude of the values, though the relative differences remain consistent.

To test the robustness of the results, the reconstructed distribution was also compared against the remaining measurements in the dataset. Because these measurements were excluded from the MENT algorithm, they serve as an independent testing set. Using the initial distribution and the corresponding lattice settings, the predicted profiles were generated, and the Mean Absolute Error

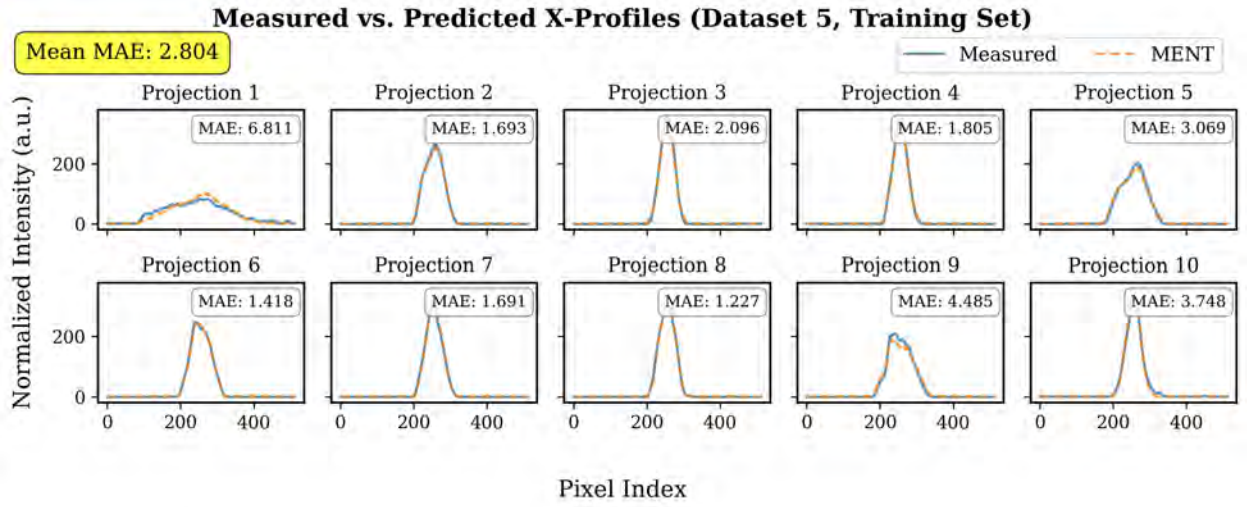


Figure 4.20 Comparison of measured and predicted profiles for dataset 5 in the X-plane.

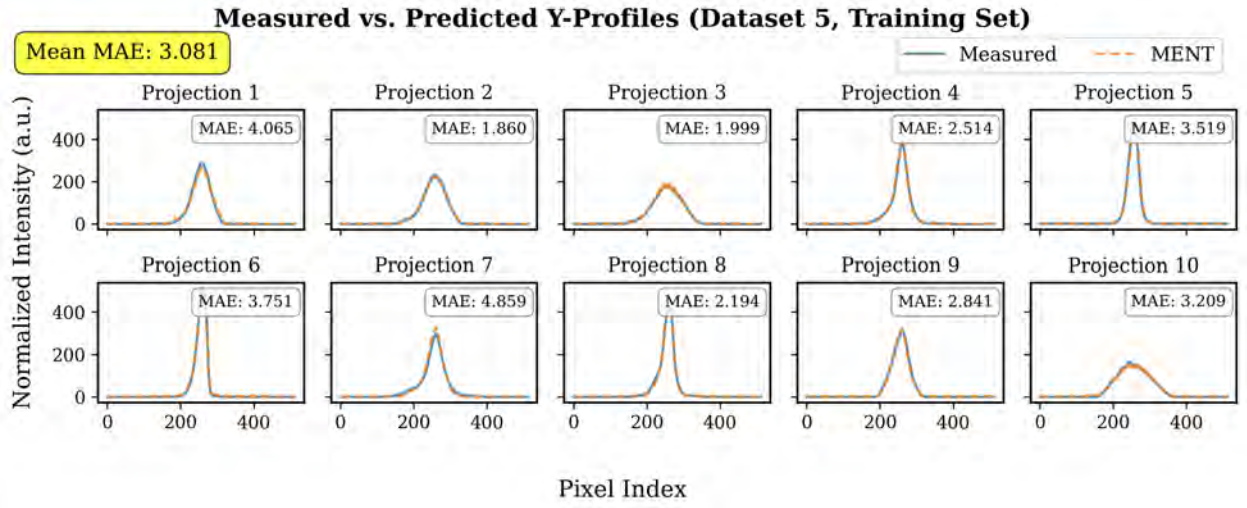


Figure 4.21 Comparison of measured and predicted profiles for dataset 5 in the Y-plane.

(MAE) was calculated. The results, arranged by beam transmission, are presented in Figs. 4.22 and 4.23. Generally, the MAE remains under 10, with a mean value of approximately 4 to 5. An intuitive sense of this fit quality can be seen in Fig. 4.26, which shows profile agreement with an MAE of up to 8.249.

While these results show an excellent fit, they do not guarantee that the reconstructed distribution perfectly matches the actual beam distribution, as this specific set of projections does not yield a mathematically unique solution. An ideal way to verify this would be to compare the results against a ground-truth measurement from a diagnostic device, such as a pepper pot or an Allison

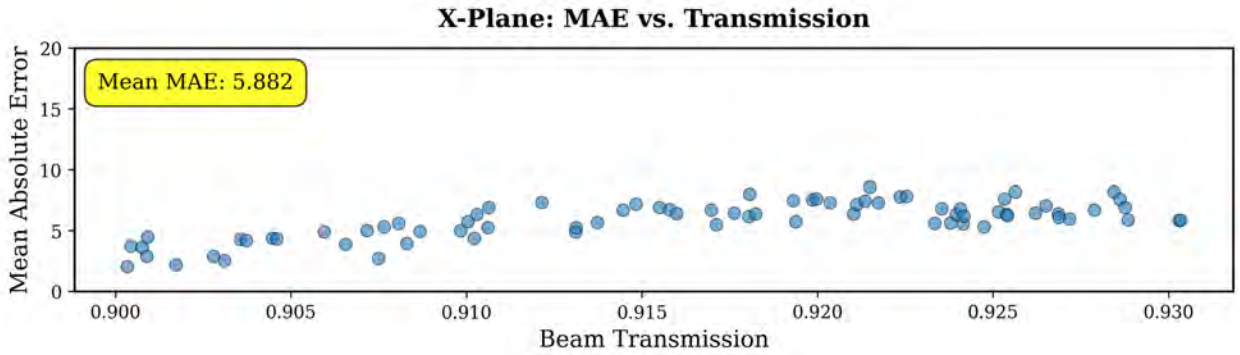


Figure 4.22 Comparison of measured and reconstructed profiles on testing data for dataset 5 on the X-plane.

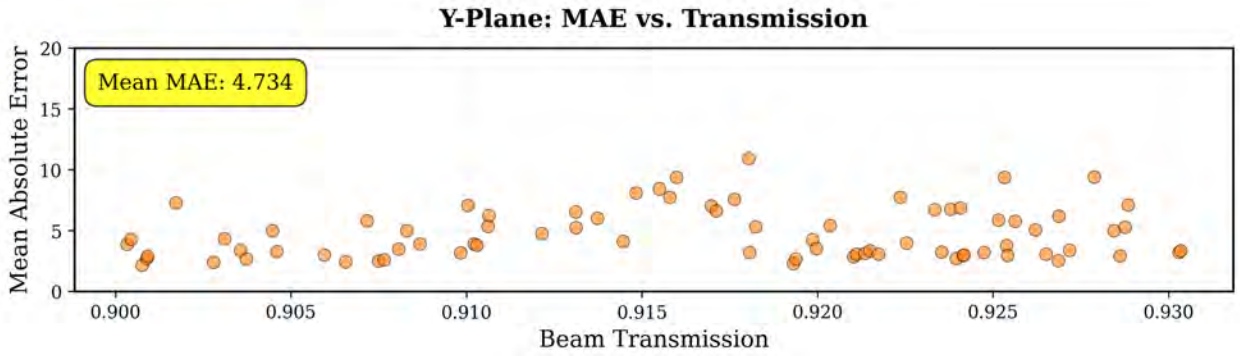


Figure 4.23 Comparison of measured and reconstructed profiles on testing data for dataset 5 on the Y-plane.

scanner [78]. A pepper pot device does exist on the ATLAS beamline, but it was unavailable for use during this experiment.

The same analysis was performed on dataset 7, utilizing four quadrupoles instead of one. The primary advantage of using multiple quadrupoles is the increased flexibility to achieve a larger angular range while maintaining high beam transmission. Conversely, the disadvantages include increased susceptibility to errors in the beamline modeling and a higher potential for beam loss. The results of this analysis are presented in Figures 4.24–4.29.

The final comparison demonstrates a better mean MAE than the single-quadrupole scan dataset. This suggests that, for this specific beamline, using four quadrupoles can yield a more accurate reconstruction. However, as seen in Fig. 4.24, the angular spread for the X plane is shorter, and for the Y plane, it is significantly narrower. While reconstructed projections may perfectly match experimental data, insufficient angular sampling also leads to uncertainties in the accuracy

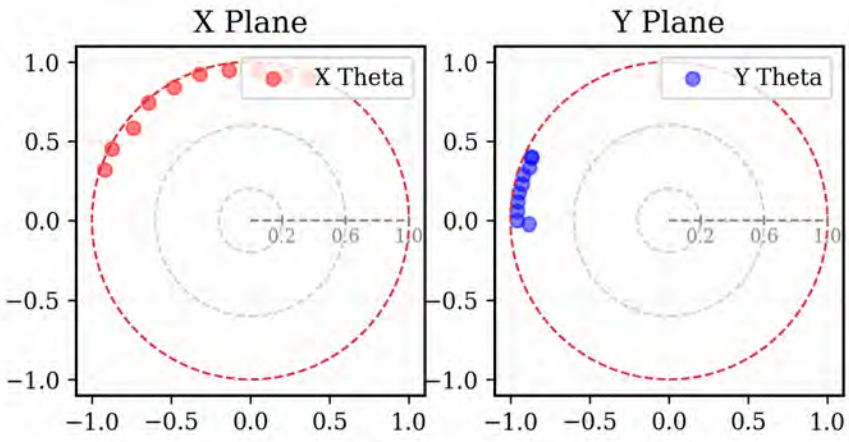


Figure 4.24 Distribution of angles in phase advance

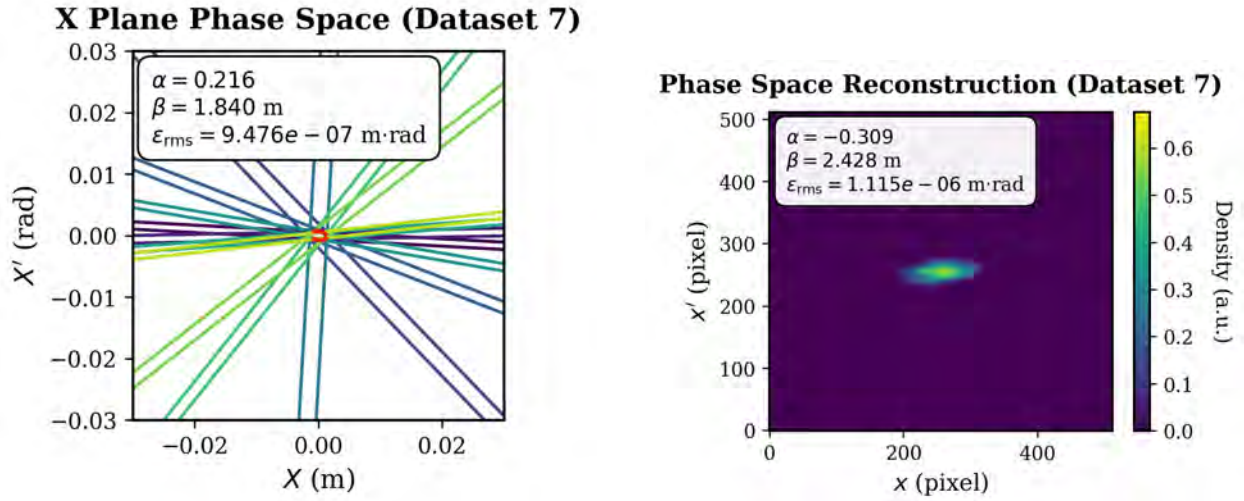


Figure 4.25 Comparison of Quad Scan estimate (left) and MENT reconstruction (right) for Dataset 7.

of the phase-space reconstruction. Quantifying the impact of this undersampling on reconstruction fidelity remains an area for future research.As such, even though the MAE is lower, the resulting reconstruction may still be inaccurate.

4.2.6 Assumptions and Limitations

Several assumptions are required for accurate phase-space reconstruction. The assumption of negligible space charge is necessary to ensure purely linear beam dynamics. Furthermore, assuming zero beam loss ensures that the measured distribution accurately reflects the total beam, while assuming a stable initial beam distribution guarantees that the exact same phase space is

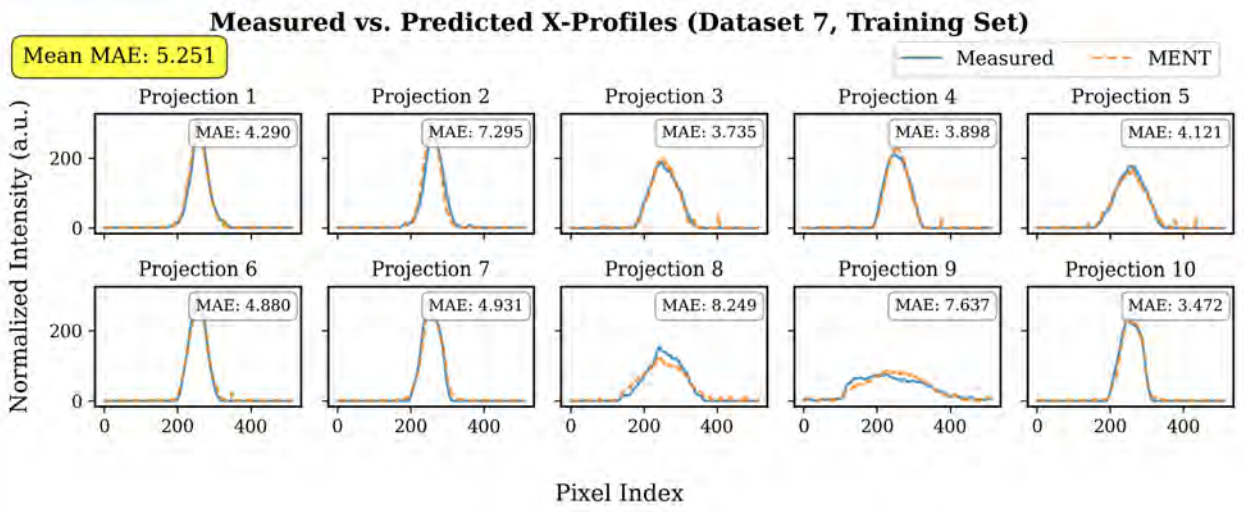


Figure 4.26 Comparison of measured and predicted profiles for dataset 7 in the X-plane.

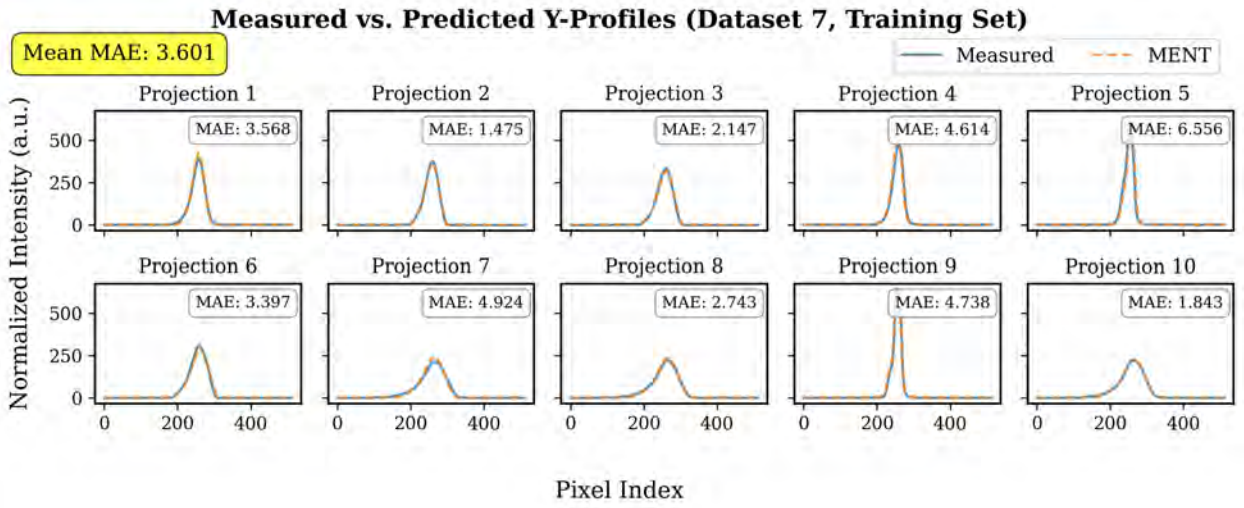


Figure 4.27 Comparison of measured and predicted profiles for dataset 7 in the Y-plane.

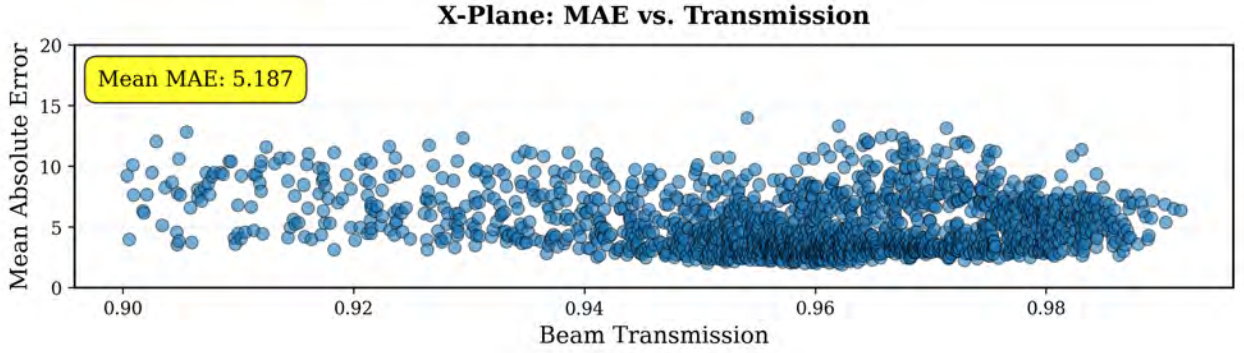


Figure 4.28 Comparison of measured and reconstructed profiles on testing data for dataset 7 on the X-plane.

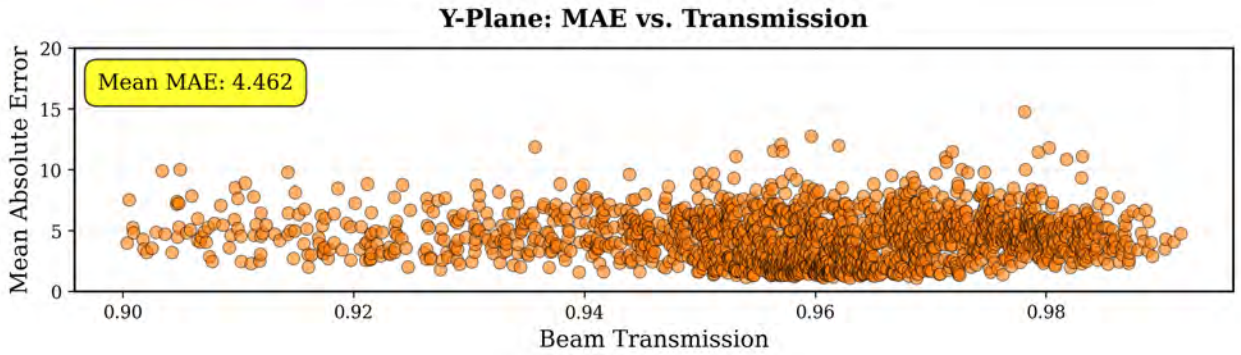


Figure 4.29 Comparison of measured and reconstructed profiles on testing data for dataset 7 on the Y-plane.

sampled across all measurements. These will be covered in the following sections, along with the limitations of tomography.

Other effects that induce nonlinearity and degrade the reconstruction include magnetic fringe fields, chromatic aberrations due to beam energy spread, and higher-order multipole errors in the magnets. Additionally, instrumental limitations such as the finite spatial resolution of the profile monitors, power supply jitter, and unmodeled transverse coupling can further compromise the fidelity of the reconstructed phase space. While not studied, these are assumed to be negligible due to the short beamline used.

4.2.6.1 Effect of Space Charge

The space-charge effect is a common term in the accelerator community referring to the influence of a beam's own electromagnetic self-fields, distinct from the contributions of external focusing fields. Phase-space reconstruction assumes linear beam transport between the reconstruction and measurement locations; however, strong space-charge forces induce nonlinearities that violate this assumption.

Assuming a uniform-density beam with $r = r_b$, and given a line charge density, the current is $I = \lambda\beta_b c$, and this leads to the volume charge density being $\rho = \frac{\lambda}{\pi r_b^2}$ for $r < r_b$, and zero elsewhere. The dimensionless perveance, Q, can be calculated to estimate the magnitude of the space-charge effect:

$$Q \equiv \frac{q\lambda}{2\pi\epsilon_0 m\gamma_r^3\beta_r^2 c^2} \tag{4.5}$$

For reference, Q reaches relatively high magnitudes of 10^{-3} to 10^{-4} near injectors, compared to a typical high-energy transport range of 10^{-4} to 10^{-6}.

It can be shown that the equations of motion in the transverse plane become [79]:

$$\frac{d^2x}{ds^2} + K_{\beta 0}x - \frac{Q}{r_b^2}x = 0 \tag{4.6}$$

$$\frac{d^2y}{ds^2} + K_{\beta 0}y - \frac{Q}{r_b^2}y = 0 \tag{4.7}$$

By rearranging the focusing strength, this can be written as:

$$\frac{d^2x}{ds^2} + K_{\beta}x = 0 \tag{4.8}$$

where

$$K_{\beta} = K_{\beta 0} - \frac{Q}{r_b^2} \tag{4.9}$$

Here, the term $\frac{Q}{r_b^2}$ acts as an effective defocusing force. By calculating $\frac{Q}{r_b^2}$ and comparing it to a typical quadrupole focusing strength, one can determine whether space-charge effects are negligible. Using the experimental parameters from Table 4.4, $Q \approx 2 \times 10^{-11}$, the beam radius is $r_b \approx 0.003$ m leading to the defocusing force being $\frac{Q}{r_b^2} \approx 2 \times 10^{-6}$ m^{-2}. Typical values for the magnetic focusing strength $K_{\beta 0}^2$ are on the order of 1 m^{-2}; therefore, the space-charge effect due to the self-field is negligible.

The conducting beam pipe may also affect beam motion via the image-charge effect. Assuming a cylindrical beam pipe and treating the beam as a line charge at its center of mass, the resulting force can be approximated as arising from an image charge located a distance $\frac{r_p^2}{|r|}$ away from the center, where r_p is the beam pipe radius. The electric field due to this image charge is:

$$E_{image} \approx \frac{\lambda}{2\pi\epsilon_0(r_p^2/r - r)} \tag{4.10}$$

Assuming the beam is near the axis such that $\frac{r}{r_p} \ll 1$, this simplifies to:

$$E_{image} \approx \frac{\lambda r}{2\pi\epsilon_0 r_p^2} \tag{4.11}$$

This introduces a defocusing term analogous to the free-space self-field, but with r_b replaced by r_p. Given $r_p \approx 0.03$ m, the defocusing force due to the beam pipe is approximately $\frac{Q}{r_p^2} \approx 2\times10^{-8}$ m^{-2}.

Thus, the effect of image charges is negligible as long as the beam remains relatively centered and small. While it is possible for the beam to travel near the edge of the pipe, such trajectories would likely induce significant beam loss and are therefore excluded from this analysis.

4.2.6.2 Effect of Finite Aperture

Beam loss is a common occurrence on the ATLAS beamline, rendering phase-space tomography difficult since it directly affects the reconstruction. To provide more experimental flexibility, we investigated whether a correction factor could be developed to account for minor beam loss. A study was conducted to evaluate the effect of beam loss due to finite apertures in the beamline using the Accelerator Toolbox simulation package. Various measurements were simulated at specified beam-loss levels to identify a systematic correction.

The simulation utilizes 10,000 particles with a kinetic energy of 1.3 MeV, a charge state of 8, and an atomic mass of 40 amu. The reconstruction focuses on the x-plane with an unnormalized RMS emittance of 3.0×10^{-6} m-rad, with $\alpha_x = 1.0$ and $\beta_x = 5.00$ m/rad. In the y-plane, the unnormalized RMS emittance is 5.0×10^{-9} m-rad, with $\alpha_y = -0.35$ and $\beta_y = 1.99$ m/rad. The x-plane represents typical beam parameters from the real beamline while the y plane emittance is set 3 orders of magnitude smaller in order to sole study the effect of aperture on the X-plane. The same settings were used for all simulated measurements for this study.

The simulation workflow for evaluating the impact of beam loss on tomographic reconstruction proceeded as follows:

1. Each data point utilizes six sets of seven quadrupole settings, generated to produce phase advances uniformly distributed over a full 180° range. An optimizer was used to determine the settings for all seven quadrupoles that yielded the requisite phase advances while maintaining perfect transmission in an ideal lattice. Steering elements were disabled during this step to maximize the optimization's efficiency in achieving the full angular spread required for accurate tomography.

2. Experimentally derived misalignment values [80] were introduced into the lattice. The steer-

ing elements were then adjusted to introduce beam loss and to set the transmission to a targeted value between 70% and 100%. This variable transmission range was selected to systematically study the effects of beam loss on reconstruction fidelity.

3. The resulting dataset consists of six measurements spanning the full angular range with the specified transmission levels. MENT was then used to calculate the transverse emittance.

4. Steps 1 through 3 were iterated over a defined range of beam loss values. Each iteration was performed multiple times to generate a statistical spread of results, as the optimizer converges on different steer settings each time.

Fig. 4.30 shows the results of this study. The error bars represent the spread in values using a 20% to 80% cut, and a clear trend emerges for transmission values of 80% and above.

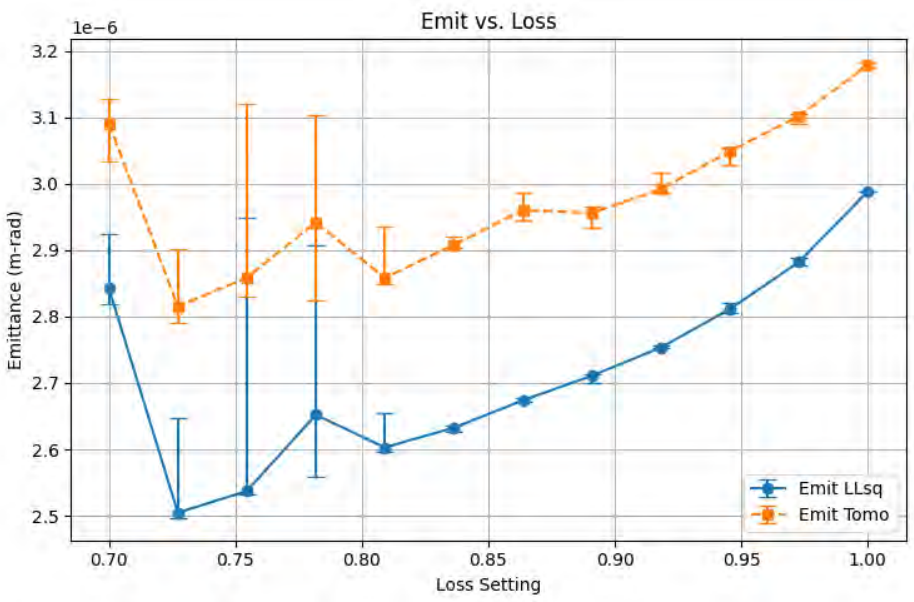


Figure 4.30 Simulation study evaluating the effect of beam loss on reconstructed emittance.

Following the encouraging simulation results, a comparable analysis was performed on the experimental data. Due to the limited size of the experimental dataset, measurements were binned into transmission intervals of 0.10, ranging from 0.6 to 1.0, except for Dataset 4, which lacked measurements below 0.7.

The MENT algorithm was then applied to these binned datasets to estimate the emittances, with the results shown in Fig. 4.31.

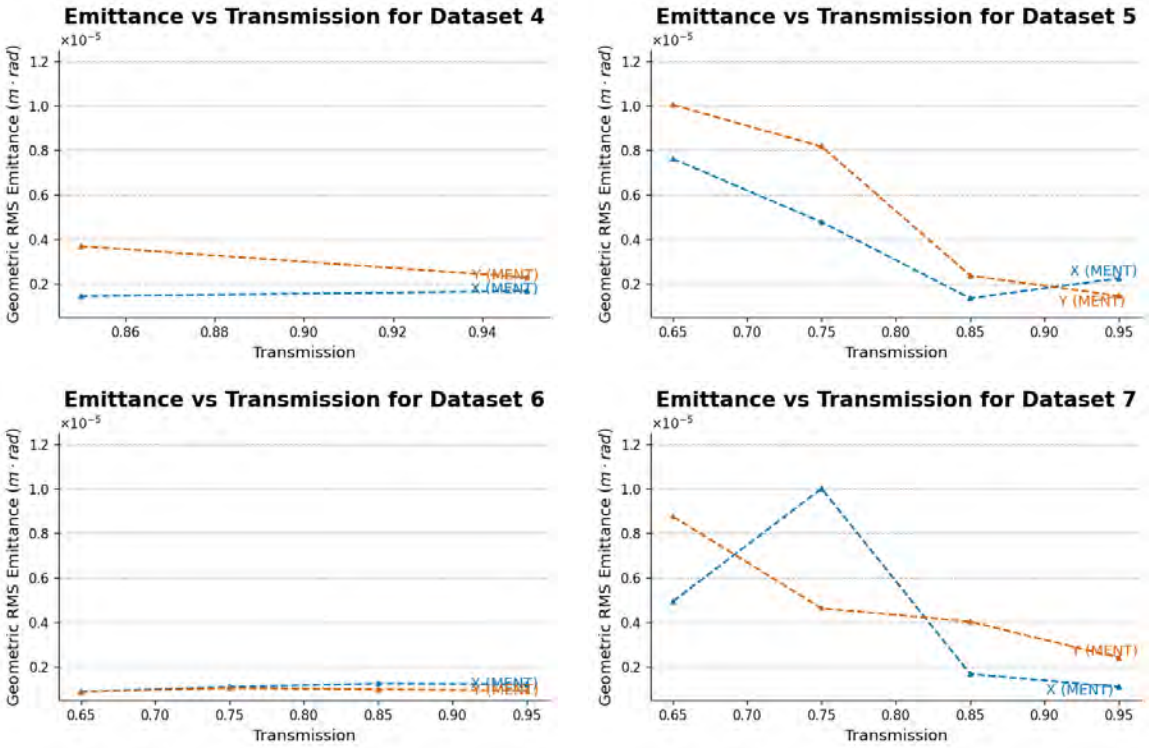


Figure 4.31 Experimental evaluation of beam loss effects on reconstructed emittance.

Unlike the simulated study, the experimental data reveal no clear trend to facilitate an emittance correction. Because some datasets yielded overestimations while others yielded underestimations, a simple correction formula is unfeasible for this beamline. This unpredictable behavior likely stems from the fact that beam losses can occur at varying locations throughout the lattice, and the resulting measurements are highly dependent on the initial beam distribution.

Despite this, one conclusion is clear. The emittance across all datasets in Fig. 4.31 converges to approximately 0.2×10^{-6} m-rad. This convergence begins at roughly 80% transmission, consistent with the simulation results. Therefore, the minimum beam transmission of 80% is derived.

While predicting and correcting for beam loss is not strictly impossible, this study highlights the inherent difficulty of the task, suggesting that any functional correction would require a highly complex model capable of pinpointing exact loss locations.

4.2.6.3 Effect of Beam Stability

The final assumption to address is that the initial beam distribution remains constant throughout the entire data collection session. Intuitively, this could be verified by comparing measurements taken with identical lattice settings at different time intervals; however, such data were not collected during this experiment. Instead, an alternative analysis of the beam's stability was conducted by evaluating variations in the reconstructed Twiss parameters and emittance over time.

Datasets 6 and 7, each spanning an 8-hour period, were divided into 1-hour intervals. The MENT algorithm was then applied to each subset, enforcing a minimum transmission cut of 80% to calculate the beam parameters over time. These results are illustrated in Figs. 4.32 and 4.33 and summarized in Table 4.7.

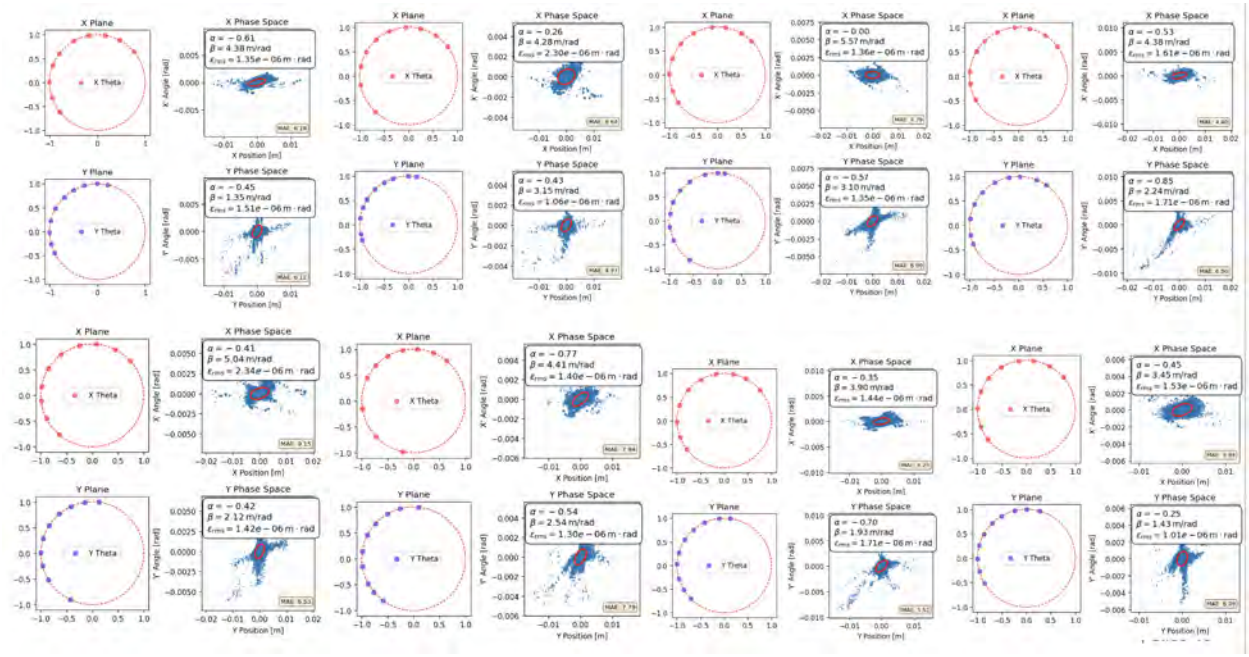


Figure 4.32 Beam evolution study on dataset 6.

	Dataset 6						Dataset 7					
Statistic	α_x	α_y	β_x	β_y	ϵ_x	ϵ_y	α_x	α_y	β_x	β_y	ϵ_x	ϵ_y
Mean	-0.4	-0.5	4.4	2.2	1.7	1.4	-0.6	-1.7	5.2	1.9	1.3	2.7
STD	0.2	0.2	0.6	0.7	0.4	0.3	0.1	0.2	0.6	0.1	0.3	0.3

Table 4.7 Summary statistics for Dataset 6 and Dataset 7. Emittance (ϵ) is expressed in units of 10^{-6} m*rad, Beta (β) in m/rad.

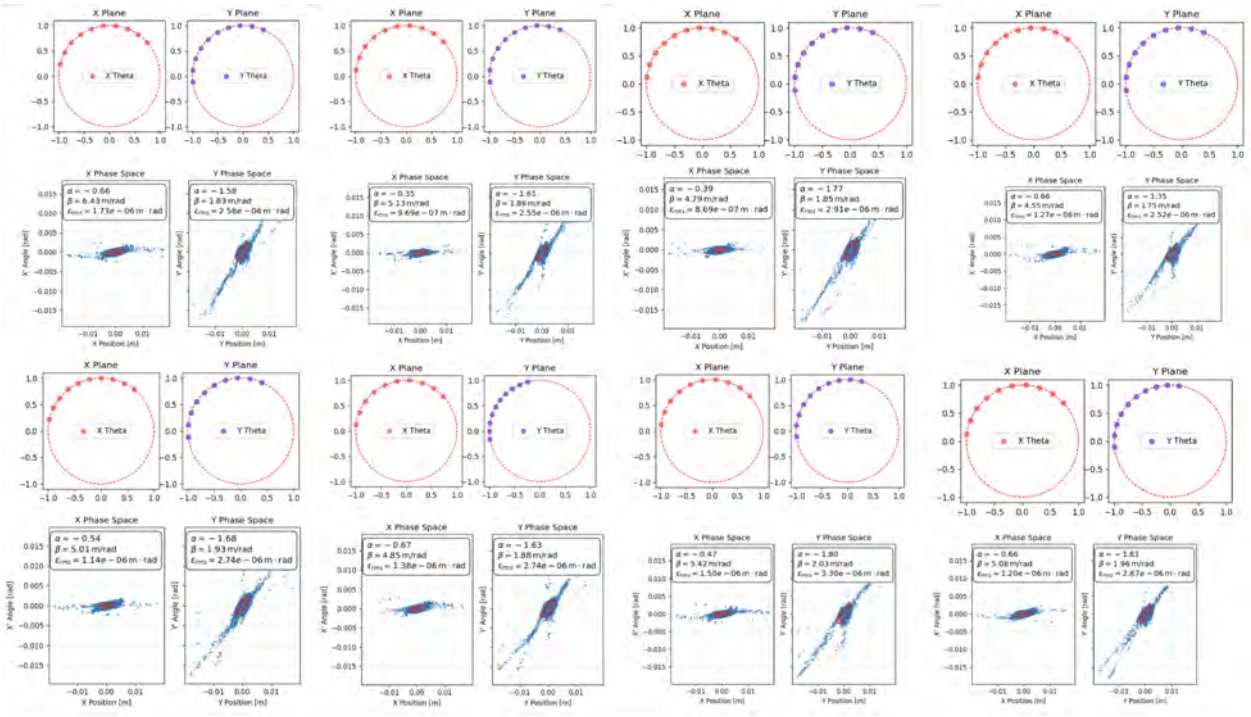


Figure 4.33 Beam evolution study on dataset 7.

In the end, while this study offers insight into the beam's stability, the results are inconclusive. Other factors that affect parameter estimation include uneven angular spreads and beam loss. Because the MENT algorithm's convergence is highly sensitive to these specific factors, distinguishing the error due to beam instability from that due to the algorithm is difficult. While the beam is assumed to be stable for the current reconstruction, achieving higher accuracy will require dedicated future research into diagnosing temporal drift.

4.2.6.4 Limitations

From a practical standpoint, the barrier to implementing phase-space tomography is high. Unlike a standard quadrupole scan, it relies on a complex algorithm. Various challenges, including precise beamline modeling, measurement limitations, complex data-acquisition procedures, beam loss, general machine feasibility, and experimental verification, hinder its widespread adoption at other accelerator facilities. To overcome this, a systematic framework is first needed to assess the feasibility of tomography on any given accelerator. Furthermore, the development of a generalized software package capable of importing a beamline model, automating the data sampling, and

performing the complete tomographic analysis would significantly enhance the utility and accessibility of this diagnostic technique.

Tomography also relies on a precise linear model of the beamline, and any deviations from this ideal degrade the accuracy of the results. These degrading factors include unmodeled beam dynamics, space-charge forces, non-linear magnetic fields, transverse coupling, magnet calibration errors, and time-dependent system drifts. To address some of these limitations, methods incorporating direct multi-particle tracking [81] have been developed, successfully demonstrating high-fidelity reconstructions even in regimes dominated by space-charge effects.

Another limitation is the distribution's inherent non-uniqueness. For applications where a simple estimate is needed, this level of accuracy is sufficient, but it can already be achieved with a simple quadrupole scan, without utilizing the full potential of a higher-fidelity reconstruction. To eventually reach particle-level measurement and control, this uncertainty needs to be better understood. There is currently no standard method to quantify the uncertainty introduced by this non-uniqueness.

4.2.7 Future Outlook

Future advancements in algorithmic efficiency could make phase-space tomography considerably faster and more practical. For instance, while the experimental results presented here required an overnight data-collection period of over 8 hours, our beam-stability analysis suggests that data collected within an hour, comprising approximately 200-300 individual measurements, can yield an accurate reconstruction. For machines with a sufficiently high chance of beam loss, future development of tomography could be developed as an automated refinement step immediately following an initial quadrupole scan, as illustrated in Fig. 4.34.

In this proposed workflow, after an initial estimate of the Twiss parameters is calculated, a dedicated algorithm selects optimal measurement points by maximizing the angular spread while minimizing beam loss. The MENT algorithm is then used to reconstruct the distribution, iteratively repeating this process until the Twiss parameters and emittance converge. The primary difficulty in implementing this lies in the sampling algorithm. In a physical beamline, inherent quadrupole

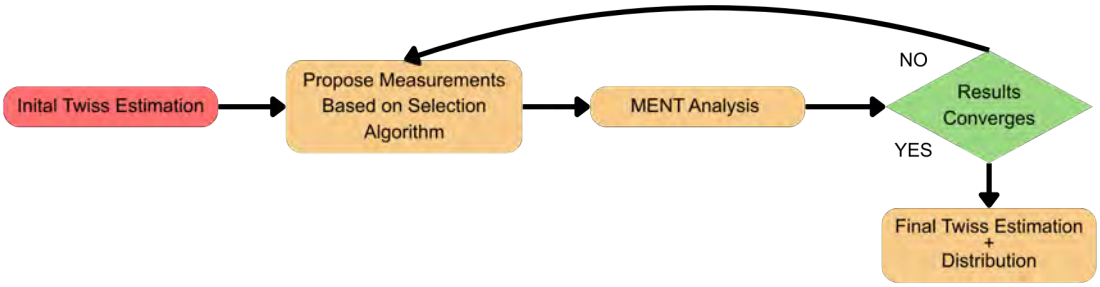


Figure 4.34 Proposed pipeline for the MENT analysis workflow.

limitations and alignment errors make beam loss more prominent, complicating the acquisition of ideal measurements.

Finally, uncertainty quantification (UQ) remains a critical area for future research. Establishing a robust framework to understand these uncertainties would enable the quantitative evaluation of new methods to further improve tomographic fidelity. The core challenge lies in the computational ability to fully sample the space of all possible mathematical reconstructions, while simultaneously accounting for the compounded effects of modeling and measurement errors.

To achieve UQ for phase-space reconstruction at the level of single-particle measurement and manipulation, research is needed to theoretically guarantee the accuracy of UQ methods. A current struggle with this is an inherent problem of the inverse problem. There is no unique distribution for a given set of projections. Thus, the solution space accounts for part of the uncertainty, but most current tomography methods provide a single optimal answer. It is possible that adding more projections will reduce this solution space, but the extension of such reduction is uncertain and needs further investigation.

To illustrate the fundamental nature of the reconstruction problem, consider a thought experiment in which we discretize a continuous distribution with compact support $f(\mathbf{r})$ onto a spatial grid of size $N \times N$. We treat the intensity at each grid point as a quantized variable, restricted to M distinct values within the normalized range $[0, 1]$.

This formulation constructs a finite configuration space Ω containing M^{N^2} unique distributions. In the continuum limit where $N, M \to \infty$, this discrete set approaches the infinite-dimensional

Hilbert space of all possible bounded, continuous functions. This represents the “set" of all potential solutions prior to any observation.

This is motivated by the ideas of SART in Sec. 2.3.4 and GPSR in Sec. 2.4.2. Measurement data at a given angle θ serves as a filter on this set. The projection operation creates a system of linear equations. Any distribution within the configuration space that does not satisfy these projection values is discarded.

Each subsequent measurement from a new angle acts as a geometric constraint, effectively slicing the high-dimensional solution space. The reconstruction process is equivalent to finding the intersection of the distributions permissible under each projection. Let Ω represent the total configuration space, P_θ the projection operator at angle θ, and $\mathbf{y}_\theta$ the measured data. The set of valid solutions S_k after measurements at k angles is

$$S_k = \{\mathbf{x} \in \Omega \mid P_{\theta_i}\mathbf{x} = \mathbf{y}_{\theta_i} \quad \forall i \in \{1, \ldots, k\}\} \tag{4.12}$$

As the number of angular views k increases, the solution set $|S_k|$ decreases and converges to the true distribution. In practice, finding this set is assumed to be computationally difficult, and research would be done to optimize this process or find reasonable approximations.

CHAPTER 5

CONCLUSIONS AND OUTLOOK

This thesis contributes efforts to push the frontiers of accelerator beam manipulation and control from RMS diagnostics and Gaussian assumptions to the single-particle level. Advancements in this direction will become necessary in the future. While this work does not yet reach single-particle resolution, it represents a significant step forward by developing and validating a range of diagnostic and computational tools.

The research explored three distinct pathways: data-driven optimization, direct phase-space measurement, and indirect phase-space measurement.

Data-driven optimization: A machine-learning model was built using TRACK simulations of the ATLAS low-energy beam transport line. By encoding phase-space projection images into compact latent variables, the model predicted downstream transmission and projected distributions. Tests showed good performance on in-distribution data and reasonable robustness under nonlinear perturbations; however, the analysis requires a critical piece of information, the initial distribution to initialize the model. This limitation motivated an investigation into direct and indirect methods for measuring the initial phase-space distribution.

Direct measurement: This thesis focuses on one specific direct phase-space mesurement method, the "one-shot" pepper-pot. A robust Python analysis pipeline was developed, benchmarked, and integrated for the ATLAS pepper-pot emittance meter. Simulation benchmarks quantified a finite hole-size bias and demonstrated agreement under controlled conditions within 2%. Experimental operation highlighted practical constraints and spurred the development of noise-processing algorithms. While direct methods are invaluable to beamline operation, inherent limitations in resolution and usable energy range motivate the development of indirect methods.

Indirect measurement: While direct measurement devices can be limited in resolution, an even simpler diagnostic with higher physical fidelity is the wire scanner, though it provides only a 1D integrated projection of the distribution, but by leveraging mathematical algorithms, indirect methods can outperform direct measurements in overall accuracy. A generalized framework

was developed to extend phase-space reconstruction algorithms to higher dimensions, and an algorithm was proposed to optimally sample the beamline, enabling efficient construction of these reconstructions. Experimentally, 2D transverse reconstructions in the ATLAS PII-to-Booster line showed that MENT reproduces measured non-Gaussian profile features more accurately than traditional quadrupole scans, provided that beam transmission remains high and the optics model is sufficiently accurate. Simulation and data studies indicated that these reconstructions become unreliable when significant beam loss occurs (e.g., when transmission drops to roughly 80%).

Achieving single-particle-level precision remains a significant challenge; however, this work demonstrates measurable progress toward that goal. Among the future directions identified in this thesis, the most consequential is the development of practical and accurate UQ to assess reconstruction reliability. Such capability would not only strengthen the interpretation of current results, but also enable systematic evaluation of complementary improvements to reconstruction algorithms, including approaches to the measurement selection problem.

With these advances, phase-space reconstruction could transition from a specialized research capability to a routine operational tool, enabling higher power, higher brightness, and more efficient tuning for next-generation accelerator facilities.

BIBLIOGRAPHY

[1] Linh Nguyen. Adiabatic capture of neutrons for high-brightness beams and storage of low- and high-field-seeking spin states. Master's thesis, Indiana University, December 2023. Indiana University/USPAS Master of Science Degree in Beam Physics and Technology.

[2] W. D. Coolidge. A powerful röntgen ray tube with a pure electron discharge. *Phys. Rev.*, 2:409–430, Dec 1913.

[3] Ernest O. Lawrence and M. Stanley Livingston. The production of high speed light ions without the use of high voltages. *Phys. Rev.*, 40:19–35, Apr 1932.

[4] J. D. Cockcroft and E. T. S. Walton. Disintegration of lithium by swift protons. *Nature*, 129(3261):649–649, 1932.

[5] D. W. Kerst. The acceleration of electrons by magnetic induction. *Phys. Rev.*, 60:47–53, Jul 1941.

[6] V I Veksler. A new method of acceleration of relativistic particles. *J. Phys.*, 9:153–158, 1945.

[7] Edwin M. McMillan. The synchrotron—a proposed high energy particle accelerator. *Phys. Rev.*, 68:143–144, Sep 1945.

[8] F. R. Elder, A. M. Gurewitsch, R. V. Langmuir, and H. C. Pollock. Radiation from electrons in a synchrotron. *Phys. Rev.*, 71:829–830, Jun 1947.

[9] Ernest D. Courant, M. Stanley Livingston, and Hartland S. Snyder. The strong-focusing synchroton—a new high energy accelerator. *Phys. Rev.*, 88:1190–1196, Dec 1952.

[10] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Phys. Lett. B*, 716(1):1–29, 2012.

[11] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett. B*, 716(1):30–61, 2012.

[12] Gerald Blazey, David Bruhwiler, John Byrd, Sarah Cousineau, Zhirong Huang, Sergei Nagaitsev, Mark Palmer, Philippe Piot, John Power, James Rosenzweig, Vladimir Shiltsev, Linda Spentzouris, and Jean-Luc Vay. Accelerator and Beam Physics Roadmap. Technical report, U.S. Department of Energy, Office of Science, Office of High Energy Physics, February 2023. Accessed: 2025-12-16.

[13] Brandon Cathey, Sarah Cousineau, Alexander Aleksandrov, and Alexander Zhukov. First six dimensional phase space measurement of an accelerator beam. *Phys. Rev. Lett.*, 121:064804, Aug 2018.

[14] V. Danilov, S. Cousineau, S. Henderson, and J. Holmes. Self-consistent time dependent two dimensional and three dimensional space charge distributions with linear force. *Phys. Rev. ST Accel. Beams*, 6:094202, Sep 2003.

[15] Alessandro G. Ruggiero. Crystalline beams. Technical Report BNL-66946, Brookhaven National Laboratory, November 1999. Invited Talk to the 2nd ICFA Workshop on The Physics of High Brightness Beams, UCLA, Los Angeles, CA.

[16] Yue Hao. Collider: Fundamental accelerator physics notes. Lecture notes, Michigan State University (FRIB), 2025. Available at `https://people.frib.msu.edu/~haoy/teaching/fundamental_AP/collider/collider.pdf` (accessed 2025-08-25).

[17] European Synchrotron Radiation Facility. Applications and case studies, 2025. Accessed: 2025-09-23.

[18] Timur Shaftan. Synchrotron radiation: Lecture 2. Lecture notes, Brookhaven National Laboratory, NSLS-II User Guide, 2021. Lecture slides prepared for the NSLS-II User Guide series.

[19] Philip H. Bucksbaum and Nora Berrah. Brighter and faster: The promise and challenge of the x-ray free-electron laser. *Physics Today*, 68(7):26–32, 2015.

[20] A. Aleksandrov, S. Cousineau, K. Ruisard, and A. Zhukov. Characterization of high dynamic range beam emittance. In *Proceedings of the 14th International Particle Accelerator Conference (IPAC'23)*, IPAC'23, pages 2724–2726, Venice, Italy, 2023. JACoW Publishing.

[21] Tim Broome. High power targets for spallation sources. In *Proceedings of the 5th European Particle Accelerator Conference (EPAC 96)*, pages 267–271, Sitges, Spain, 1996. Institute of Physics Publishing. Rutherford Appleton Laboratory, Chilton, Didcot, Oxon., UK.

[22] S. Nagaitsev, Z. Huang, J. Power, J.-L. Vay, P. Piot, L. Spentzouris, and J. Rosenzweig. Accelerator and beam physics research goals and opportunities, January 2021. Report of the Accelerator and Beam Physics Working Group.

[23] Brandon Cathey, Sarah Cousineau, Alexander Aleksandrov, and Alexander Zhukov. First six dimensional phase space measurement of an accelerator beam. *Phys. Rev. Lett.*, 121:064804, Aug 2018.

[24] W. H. Tan, S. Antipov, D. S. Doran, G. Ha, C. Jing, E. Knight, S. Kuzikov, W. Liu, X. Lu, P. Piot, J. G. Power, J. Shao, C. Whiteford, and E. E. Wisniewski. Demonstration of sub-GV/m accelerating field in a photoemission electron gun powered by nanosecond X-band radio-frequency pulses. *Physical Review Accelerators and Beams*, 25(8):083402, Aug 2022.

[25] Chunguang Jing, John Power, Jiahang Shao, Gwanghui Ha, Philippe Piot, Xueying Lu, Alexander Zholents, Alexei Kanareykin, Sergey Kuzikov, James B. Rosenzweig, Gerard Andonian, Evgenya Ivanovna Simakov, Janardan Upadhyay, Chuanxiang Tang, Richard J. Temkin, Emilio Alessandro Nanni, and John Lewellen. Continuous and coordinated efforts of structure wakefield acceleration (SWFA) development for an energy frontier machine, 2022.

[26] T. Kanemura et al. Liquid lithium charge stripping technology: Achievement and lessons learned. In *Proc. IPAC'25*, number 16 in IPAC'25 - 16th International Particle Accelerator Conference, pages 19–23. JACoW Publishing, Geneva, Switzerland, 06 2025.

[27] V. N. Aseev, P. N. Ostroumov, E. S. Lessner, and B. Mustapha. TRACK: The new beam dynamics code. In *Proceedings of the 2005 Particle Accelerator Conference*, pages 2053–2055, Knoxville, TN, USA, 2005. IEEE.

[28] A. Ravichandran and B. Mustapha. Machine-Learning-Assisted Rapid Beam Energy Change at the ATLAS Heavy Ion Linac. In *Proceedings of the 16th International Conference on Heavy Ion Accelerator Technology*, HIAT'25, pages 248–250, East Lansing, MI, USA, 2025. JACoW Publishing.

[29] R. R. Wilson. Radiological use of fast proton. *Radiology*, 47:487–491, 1946.

[30] Andrew Brown and Herman Suit. The centenary of the discovery of the bragg peak. *Radiotherapy and Oncology*, 73:265–268, 2004.

[31] M. G. Vincini, M. Zaffaroni, M. Schwarz, G. Marvaso, F. Mastroleo, S. Volpe, L. Bergamaschi, G. C. Mazzola, G. Corrao, R. Orecchia, B. A. Jereczek-Fossa, and D. Alterio. More than five decades of proton therapy: A bibliometric overview of the scientific literature. *Cancers (Basel)*, 15(23):5545, 2023.

[32] I. M. Kapchinskij and V. V. Vladimirskij. Limitations of proton beam current in a strong focusing linear accelerator associated with the beam space charge. In L. Kowarski, editor, *Proceedings of the International Conference on High Energy Accelerators and Instrumentation*, page 274, Geneva, 1959. CERN Scientific Information Service.

[33] Ryan Roussel, Juan Pablo Gonzalez-Aguilera, Eric Wisniewski, Alexander Ody, Wanming Liu, John Power, Young-Kee Kim, and Auralee Edelen. Efficient six-dimensional phase space reconstructions from experimental measurements using generative machine learning. *Phys. Rev. Accel. Beams*, 27:094601, Sep 2024.

[34] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. *Dive into Deep Learning*. Cambridge University Press, 2023. `https://D2L.ai`.

[35] Laurent Deniau, Hans Grote, Ghislain Roy, and Frank Schmidt. *The MAD-X Program (Methodical Accelerator Design) Version 5.08.01 User's Reference Manual*. European Laboratory for Particle Physics (CERN), Geneva, Switzerland, December 2023. CERN program library entry T5001.

[36] S. Y. Lee. *Accelerator Physics*. World Scientific, Singapore, 4th edition, 2018.

[37] F. Marti et al. Magnetic multipoles and field expansion, 2026. Accessed: January 29, 2026.

[38] P. Bryant. Beam transfer lines. In *CERN Accelerator School (CAS)*, 1992.

[39] Steven M. Lund, Yue Hao, Yichao Jing, and Jonathan Wong. 09 solenoid focusing, June 2021.

[40] Wikimedia Commons. Courant-Snyder parameters and the phase space ellipse, 2024. Accessed: 2026-01-29; Licensed under Creative Commons.

[41] M. P. Stockli, W. Blokland, T. Gorlov, B. Han, C. Long, T. Pennisi, and S. Assadi. Low-energy emittance studies with the new SNS Allison emittance scanner. In *Proceedings of PAC09*, pages 3974–3976, Vancouver, BC, Canada, 2009.

[42] Aurelia Laxdal. Allison scanner emittance diagnostic development at TRIUMF. In *Proceedings of the 27th Linear Accelerator Conference (LINAC14)*, Geneva, Switzerland, September 2014. TRIUMF, JACoW. Paper THIOC02.

[43] I. Knight and B. Mustapha. Beam emittance and twiss parameters from pepper-pot images using physically informed neural nets. In *Proc. LINAC2024*, number 32 in International Linear Accelerator Conference, pages 299–301. JACoW Publishing, Geneva, Switzerland, 08 2024.

[44] G. Z. Georgiev and M. Krasilnikov. Virtual pepper-pot technique for 4d phase space measurements. In *Proc. 8th Int. Beam Instrum. Conf. (IBIC'19)*, pages 586–590, Geneva, Switzerland, Sep. 2019. JACoW Publishing.

[45] Peter Forck. Lecture notes on beam instrumentation and diagnostics. Joint University Accelerator School (JUAS), January–March 2022. Section 3.3: Wire scanner, p. 45. Affiliation: Gesellschaft für Schwerionenforschung (GSI).

[46] Ippei Yamada, Motoi Wada, Katsuhiro Moriya, Junichiro Kamiya, Pranab Kumar Saha, and Michikazu Kinsho. High-intensity beam profile measurement using a gas sheet monitor by beam induced fluorescence detection. *Phys. Rev. Accel. Beams*, 24:042801, Apr 2021.

[47] C. Behrens and C. Gerth. On the limitations of longitudinal phase space measurements using a transverse deflecting structure. In *Proceedings of DIPAC09*, pages 269–271, Basel, Switzerland, 2009.

[48] Owen Convery, Lewis Smith, Yarin Gal, and Adi Hanuka. Uncertainty quantification for virtual diagnostic of particle accelerators. *Phys. Rev. Accel. Beams*, 24:074602, Jul 2021.

[49] K. M. Hock et al. Beam tomography research at daresbury laboratory. *Nuclear Instruments and Methods in Physics Research Section A*, 753:38–55, July 2014.

[50] K. M. Hock, M. G. Ibison, D. J. Holder, A. Wolski, and B. D. Muratori. Beam tomography in transverse normalised phase space. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 642(1):36–44, June 2011.

[51] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. Society of Industrial and Applied Mathematics, 2001.

[52] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014.

[53] Wikimedia Commons contributors. Kaczmarz iteration in $\mathbb{R}^2$. Wikimedia Commons, 2025. Accessed: 2025-10-17.

[54] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.

[55] Gerald N. Minerbo. Ment: A maximum entropy algorithm for reconstructing a source from projection data. *Computer Graphics and Image Processing*, pages 48–68, 1979. Received July 10, 1978; accepted July 14, 1978.

[56] A. Wolski, M. A. Johnson, M. King, B. L. Militsyn, and P. H. Williams. Transverse phase space tomography in an accelerator test facility using image compression and machine learning. *Phys. Rev. Accel. Beams*, 25:122803, Dec 2022.

[57] Alexander Scheinker. cdvae: Vae-guided diffusion for particle accelerator beam 6d phase space projection diagnostics. *Scientific Reports*, 14(29571), Nov 2024.

[58] S. Lidia and K. Murphy. Error analysis for pepperpot emittance measurements redux: Correlated phase spaces. In *Proceedings of the 3rd International Beam Instrumentation Conference (IBIC'14)*, pages 579–583, Monterey, CA, USA, 2014. JACoW Publishing.

[59] Dor Bank, Noam Koenigstein, and Raja Giryes. Autoencoders. *arXiv preprint arXiv:2003.05991*, 2020.

[60] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[61] Hervé Abdi and Lynne J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.

[62] Bin Ding, Huimin Qian, and Jun Zhou. Activation functions and their characteristics in deep neural networks. In *2018 Chinese control and decision conference (CCDC)*, pages 1836–1841. IEEE, 2018.

[63] Austin Hoover. n-dimensional maximum-entropy tomography via particle sampling. *Phys. Rev. Accel. Beams*, 28:L084601, Aug 2025.

[64] Austin Hoover. Four-dimensional phase space tomography from one-dimensional measurements of a hadron beam. *Phys. Rev. Accel. Beams*, 27:122802, Dec 2024.

[65] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[66] S. Kondrashev, A. Barcikowski, B. Mustapha, P. N. Ostroumov, and N. Vinogradov. FOUR-DIMENSIONAL EMITTANCE METER FOR DC ION BEAMS EXTRACTED FROM AN ECR ION SOURCE. In *Proceedings of LINAC08*, pages 597–599, Victoria, BC, Canada, 2008.

[67] R. Cee, C. Dorn, E. Feldmeier, T. Haberer, A. Peters, J. Schreiner, and T. Winkelmann. Development of a pepper pot emittance measurement device for the HIT-LEBT. In *Proceedings of the 10th International Beam Instrumentation Conference (IBIC'21)*, pages 214–217, Pohang, Rep. of Korea, 2021. JACoW Publishing.

[68] Klaus Flöttmann. Pepper pot design for space charge dominated high brightness beams. Technical Report TESLA-FEL 96-09, Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany, May 1996.

[69] M Ablikim et al. Design and construction of the besiii detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 614(3):345–399, 2010.

[70] Ren-Yuan Zhu. Radiation damage in scintillating crystals. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 413(2-3):297–311, 1998.

[71] M.M. Hamada, F.E. Costa, S. Shimizu, and S. Kubota. Radiation damage of CsI(Tl) scintillators: blocking of energy transfer process of Vk centers to Tl+ activators. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 486(1-2):330–335, 2002.

[72] Luxium Solutions. *CsI(Tl), CsI(Na) Cesium Iodide Scintillation Material*.

[73] Emre Cosgun, SeokHo Moon, DongHwan Kim, Seung Hyun Lee, HanSung Kim, HyeokJung Kwon, and Moses Chung. Characterization and benchmarking of complete 4D transverse phase space for high-intensity ion beams using novel virtual pepper-pot technique. *Scientific Reports*, 15(1):40761, 2025.

[74] National Electrostatics Corporation. *Instruction Manual for Operation and Service of Beam Profile Monitor Model BMP81 (Integral Collector)*, 1995. Instruction manual.

[75] National Electrostatics Corp. Faraday Cups, 2026. Accessed: 2026-01-14.

[76] B. Nash, N. Carmignani, L. Farvacque, S. M. Liuzzo, T. P. Perron, P. Raimondi, R. Versteegen, and S. M. White. New Functionality for Beam Dynamics in Accelerator Toolbox (AT). In *Proc. IPAC'15*, pages 113–116, Richmond, VA, USA, 2015. JACoW Publishing.

[77] Austin Hoover. N-dimensional maximum-entropy tomography via particle sampling, May 2025.

[78] Dong-O Jeon, Ji-Ho Jang, Jean-Baptiste Lallement, and Veliko Dimov. Transverse phase space tomography and comparison with allison scanner data in the raon heavy ion accelerator. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1082:171005, 2026.

[79] Steven M. Lund and Yue Hao. Transverse space-charge effects. US Particle Accelerator School "Accelerator Physics", June 2018.

[80] A. Ravichandran and B. Mustapha. Extracting Critical Beamline Element Misalignments from Data Using a Beam Simulation Model. In *Proc. LINAC'24*, number 32 in Linear Accelerator Conference, pages 53–55, Chicago, IL, USA, 2024. JACoW Publishing.

[81] V. A. Dimov, J. B. Lallement, A. M. Lombardi, and R. Gaur. Emittance Reconstruction Techniques in Presence of Space Charge Applied During the Linac4 Beam Commissioning. In *Proc. HB'16*, number 57 in ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams, pages 433–438, Malmö, Sweden, 2016. JACoW Publishing.

APPENDIX A

LAGRANGE MULTIPLIER METHOD

The method of Lagrange multiplier is a method to find the maximum and minimum of a function $f(x, y)$ subject to a constraint $g(x, y) = 0$.

First, a new function called the Lagrange is constructed.

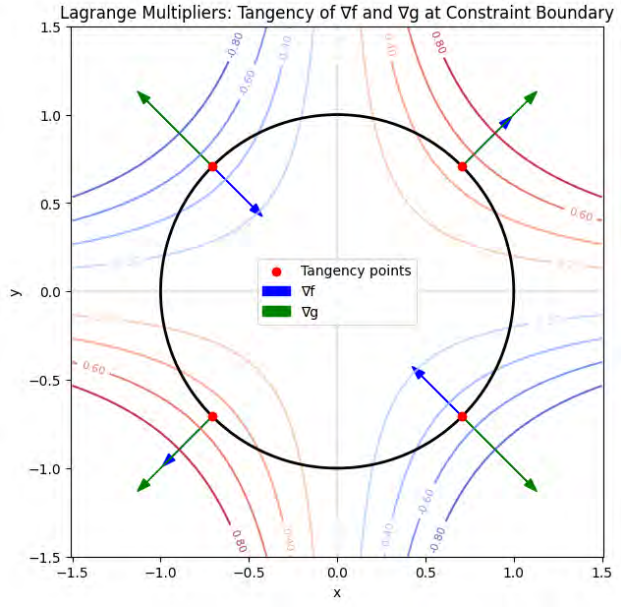
$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda \cdot g(x, y) \tag{A.1}$$

This is not to be confused with the Lagrangian from Lagrangian mechanics. λ is known as the Lagrange multiplier. The constraint $g(x, y) = 0$ is enforced when all particle derivatives of $\mathcal{L}$ equal zero.

In this case,

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial \mathcal{L}}{\partial y} = \frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

Note that $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ singles out the $g(x, y) = 0$ condition.



!

Figure A.1 Geometric interpretation of lagrange multiplier method.

A geometric example is given in Fig. A.1 for

$$f(x,y) = xy, \quad g(x,y) = x^2 + y^2 - 1 = 0$$

$$\mathcal{L}(x,y,\lambda = xy + \lambda(x^2 + y^2 - 1)$$

When $\nabla f(x,y) = \lambda \nabla g(x,y)$, this correspond for our conditions, and geometrically, it correspond to the point where the gradient align. Additionally, the gradient will differ by a factor, and that is taken into account by the lambda factor.

APPENDIX B

QUADUPOLE SCAN VS MENT

A direct comparison of the quadrupole scan reconstruction verse the MENT reconstruction. The same profiles were used for both methods and both uses MSE as the figure of merits.

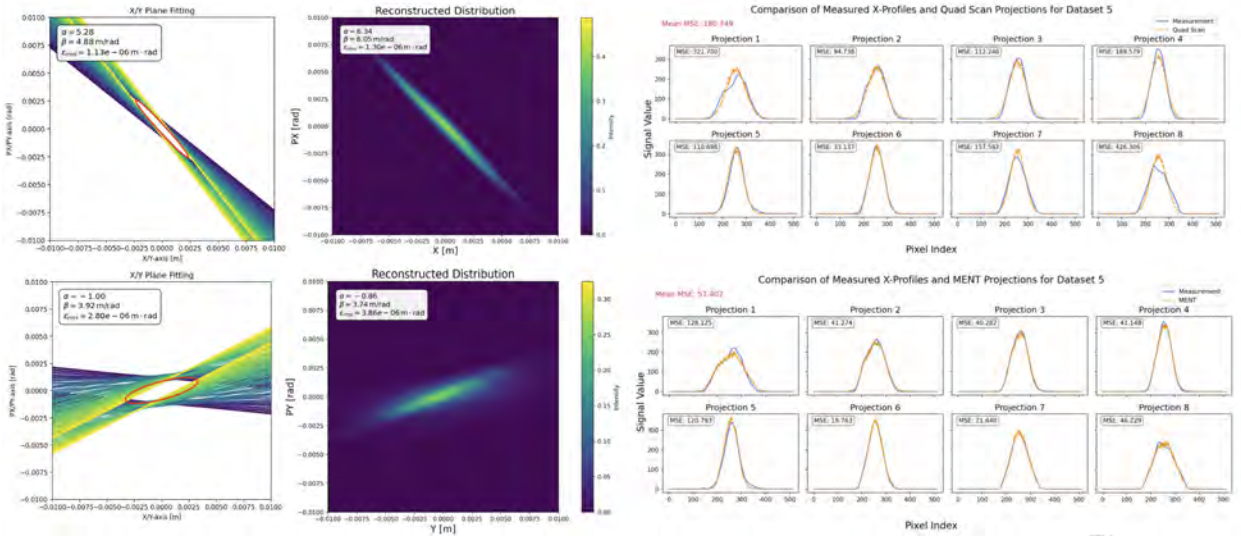


Figure B.1 Quadupole scan vs MENT tomography comparison.

MENT is clearly shown to be superior, able to capture the non-gaussian features of the profiles.