

ION ACCELERATOR APPLICATIONS, COMPACT RF LINEAR ARCHITECTURES
AND MACHINE LEARNING FOR REAL-TIME TUNING AND OPTIMIZATION

By

Nicholas Anthony Valverde

A DISSERTATION

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

Physics — Doctor of Philosophy

2024

ABSTRACT

Linear ion accelerators are widely applicable tools for scientific research and high technology industry. The broad scope of applicability demands a wide range of beam and accelerator operating regimes that entail careful design for desired performance. Optimally meeting targeted beam properties rely on detailed analysis of the interplay between the charged particle beam and material structures that are responsible for producing the electric and magnetic fields used to guide, focus, and accelerate the beam. This interplay is captured in the study of beam dynamics where the transverse and longitudinal motion of beam of charged particles is studied, to the extent possible, separately under reduced models to gain intuition about the system and better form optimal strategies. The integrated system performance is then examined with first principle particle codes that simulate the full 3D dynamics with combined and space charge effects to verify performance. Ideally, this is carried out in source to target simulations.

This thesis mainly discusses the theoretical and computational work done to support the development of compact, cost effective, multi-beam linear ion accelerator systems. Part I provides a brief description of the theoretical concepts needed to better understand particle dynamics in radio frequency linear accelerators. Part II discusses efforts made to optimize conductor geometries used to focus and accelerate beam arrays, and reduced models developed to analyze and better optimize the longitudinal and transverse dynamics in isolation. These reduced models are used to formulate improved systems that are then self-consistently simulated using a particle-in-cell accelerator code called Warp to model full 3D effects including non-linear and space charge effects to better evaluate system performance. Ideas for improved future systems building off the developed computational tool kit are presented. Lastly, Part III details a limited scope project that applied a robust auto tuning algorithm to

an existing ion induction accelerator NDCX-II at Lawrence Berkeley National Laboratory. The NDCX-II system has complicated longitudinal pulse compressions that are difficult to optimize for the final pulse durations, spot size, and flux on target. Once found, optimal settings will drift in time as due to various non-static changes (lab conditions such as temperature, etc.). The algorithm was used both as a case study in finding optimum settings for the accelerator, and as a way to sample the accelerator parameter space in order to collect data and build a neural network to act as a surrogate model. With a tuning algorithm, the system can rapidly be brought back to the most-current design optimum and the surrogate model can be used to efficiently probe the parameter space for new optimum, compare performance and identify faults, and more. The tuning algorithm and data cycle is detailed. A neural network is created to predict accumulated charge on target with limited experimental data—approximately 400 experimental runs.

To the times I missed with family, friends, and acquaintances. It was not personal, I was busy with this work. Lastly, to whoever finds this useful.

ACKNOWLEDGMENTS

I could have not completed this work without the assistance of so many people. Unfortunately, I cannot name everyone as the names and reasons would fill volumes but I do my best to capture the immense support I have received from various persons. Firstly, my mother Connie Nikolaidis, father Robert Valverde, and brother Nathan Valverde for their endless support and for not questioning my path. Though, from an outside perspective my “path” probably looked like the unintelligible meandering of a dolt at best. I was strongly influenced, encouraged, and motivated to pursue higher education by the folks at Complex Adaptive Systems at Arizona State University (ASU), namely: Stephanie Calderone, April Johnson, Dr. Anna Barker, and Dr. George Post who acted as a mentor and an excellent academic to someday emulate. Dr. Sangeeta Malhotra at ASU was incredibly influential and helpful in my early undergraduate years. Before I was a graduate student at Michigan State University (MSU), I was lucky enough to land a Research Experience for Undergraduates (REU) at MSU and was mentored by Dr. Subhendra D. Mahanti who was an excellent mentor and gave me my first exposure to theoretical calculations and computer simulation work. This experience was instrumental to my pursuing of graduate work at MSU. While attending MSU I had nothing but good experiences despite the rigorous classwork and the ever imposing imposter syndrome. I want to acknowledge the graduate department secretary Kim Crosslan; Kim looks out for all of us graduate students and has always looked out for me. It is not currently known if the department can function without her. I’m not sure how she does it but she is one of a kind and I cannot thank her enough. Initially I joined the IMSRG group under Dr. Scott Bogner, Dr. Morten Hjorth-Jensen, and Dr. Heiko Herbert. While in this group I had to evaluate my skillset and design a self-study curriculum to improve my programming language. The members of this group were incredibly helpful.

They were eager to assist and Morten provided me with the best resource for programming in Python, to which I am thankful to him and the late Dr. Hans Petter Langtangen for his inimitable writings and textbooks. I am grateful to Heiko for providing me with additional computational resources when asked. I am also very grateful to Scott for having the difficult conversation with me about my interests and setting me on the path to accelerator physics. When searching for new a research group, Dr. Artemis Spyrou curated a list of interviews for me which I'm especially grateful. This landed me under the advisement of Dr. Steven Lund who I will save for last (best?). I also joined the Accelerator Science Engineering and Technology (ASET) group which has been a tremendous group that never ceases to offer support and opportunities. When I first joined, Dr. Robert Hipple along with Dr. Lund gave me an initial project. Robert was generous with his time and wisdom, and helped me avoid many pitfalls early on in my graduate career. My initial training, and much of my subsequent research, was done using the Warp software and Dr. Dave Grote was incredibly helpful in troubleshooting at all stages of my graduate work. I would like to specifically acknowledge and thank Dr. Peter Ostroumov who is the group lead at MSU and a great one at that. I would also like to thank my comrades in the ASET group who are as humble as they are talented. When forming my committee I was worried about collecting such remarkable scientists to judge my work. Years later I am grateful that this worry was misplaced as my committee consisting of Dr. Steven Lund (Chair), Dr. Yue Hao, Dr. Dean Lee, Dr. Steven Lidia, and Dr. Chong Yu Ruan have been supportive, encouraging, and helpful through my years. Although late to the accelerator world, I was rapidly brought up to speed by taking courses offered by the United States Particle Accelerator School (USPAS) and my sincerest thanks to this wonderful school and the many scientist who volunteer their time to keep it running. I'd especially like to thank Susan Winchester and Dr. Lund for handling

the logistics of the school and navigating the various blocks that arise each session. After completing this coursework, I was lucky to be placed into a research position at Lawrence Berkeley National Lab (LBNL) under Dr. Qing Ji and Dr. Arun Persaud who are both great scientists. I have learned much working with these two over the past few years and their patience and support know no bounds. Graduate school is a difficult journey (an understatement) and my journey would have stopped long before the finish line had it not been for the external support of various friends and communities. My volleyball group consisting of Dr. Zoe Hansen, Dr. Macy Pell, Dr. Adam Kawash, Kurt Walcheske, Steven Hemmerle, Meeshon Rogers, and Dr. Mike Pajkos were an endless source of support and jubilation. I'd like to separately acknowledge Mike for his friendship and indulging me in many discussions. The yoga community at Yoga Connect was an especially important place to me and one that I will not be able to easily replace, if at all. I want to thank the Tucson and East Lansing volleyball community where I can vent frustrations with a great group of people in a fantastic sport. I'd like to acknowledge Josh Canlas who has indulged me in many discussions and has been a kindred spirit that has given me inspiration. Captain Tyler Dolezal of the United States Air Force who has been my friend and fellow physicist since undergraduate. His help, counsel, and our various discussions have been invaluable. Hang Khuu deserves special thanks. Her support is unrivaled and I'm especially grateful to her ensuring that I ate when research interests held my focus from hunger.

Finally, I'd like to circle back to my adviser and, dare I say, friend Dr. Steven Lund. He has been with me through almost all the tribulations offering endless support, guidance, and wisdom. His mentorship is only matched by his incredible scientific mind. I would consider myself a lucky individual for having the opportunity to be guided by so brilliant an individual and to have shared so many discussions, thoughts, and time. I doubt I will ever

repay the debt I have accrued to this individual but I hope that I can succeed as a notable scientist and that would pay off some of it. Thank you Steve for these last few years and your help through all of it. I hope we can continue to work together and find time to chat.

TABLE OF CONTENTS

Part I Linear Ion Accelerators	1
Chapter 1. Introduction	2
Chapter 2. Longitudinal Dynamics	4
2.1 The Transit Time Factor	5
2.2 RF Synchronism for Particle Acceleration	7
2.3 Summary	17
Chapter 3. Transverse Dynamics	18
3.1 Single Particle Dynamics	19
3.1.1 Calculating the Self-Fields for an Elliptical Beam with Uniformly Dis-	
tributed Charge Density	23
3.1.2 Forces from Applied Electrostatic Quadrupole (ESQ) Fields	25
3.1.3 The Kapchinskij-Vladimirskij (KV) Envelope Equations	27
3.2 Summary	36
Chapter 4. Self Consistent Simulations	37
Part II Compact Multi-Beam RF Linear Ion Accelerators	42
Chapter 5. RF-MEMS System	43
5.1 Introduction	43
5.2 Overview of Compact Multi-Beam MEMS Linac Structures	45
5.3 Fields from RF Acceleration Arrays	49
5.3.1 Electric Field Profile of the RF-Gap	51
5.4 Optimized Electrostatic Quadrupole (ESQ) Focusing Array Fields	53
5.4.1 Numerical Methods	54
5.4.2 Optimizing ESQ Arrays to Minimize Leading Order Error in Field Ex-	
pansion	60
5.4.2.1 Benchmarking for a Single Axially Long ESQ	60
5.4.2.2 Finite Axial Length ESQ Optimization for Full and Chopped	
Rod Geometries	61
5.4.2.3 Voltage Breakdown Limits for ESQ Array Geometry	62
5.4.2.4 Optimization Considerations From Realistic Geometry	66
5.5 Summary and Concluding Remarks	70
Chapter 6. Longitudinal Dynamics in RF-MEMS	72
6.1 Numerical Methods	73
6.2 Benchmark of the 1D Model with Warp in a 4-Gap System	75
6.3 Lattice Design for Optimal Acceleration, Bunching, or Both	77
6.3.1 Initial Conditions and Parameter Settings	77

6.3.2	Comparing Lattice Designs	79
6.4	Summary and Concluding Remarks	85
Chapter 7.	Transverse Dynamics in RF-MEMS	87
7.1	Geometric and Beam Parameters	87
7.2	Numerical Integration of KV-Envelope Equations	88
7.3	Modeling the RMS-Edge Envelope Evolution Without Acceleration	90
7.3.1	Benchmarking with Warp	92
7.3.2	Finding Acceleration Lattice Matched Conditions Using the Nelder-Mead Search Algorithm	96
7.4	Modeling the RMS-Edge Envelope Evolution with Acceleration	100
7.4.1	Consideration for Creating a Benchmark with Warp	100
7.4.2	Thin-gap KV Envelope Model for Transverse Acceleration Effects from Acceleration	101
7.4.3	Thick-gap KV Envelope Model to Capture Acceleration Effects	105
7.5	Lattice Design For Compact RMS Edge Envelope	115
7.5.1	The Gap-only Acceleration Section	115
7.5.2	The Regular Acceleration and ESQ Doublet Focusing (Gap+ESQ) Section	119
7.5.3	The Matching Section	120
7.5.4	Gap+ESQ Lattice	122
7.6	Summary and Concluding Remarks	127
Chapter 8.	3D Self-Consistent Particle-In-Cell Simulations of RF-MEMS	131
8.1	Aside: Panel Plots and Simulation Snapshots	131
8.2	Beam Creation and Loading	133
8.3	Simulation Mesh	135
8.4	Simulation Control	137
8.5	Diagnostic and Data Management	138
8.6	Short Pulsed Beam Simulations	139
8.7	Continuous Beam Injection Simulations	152
8.8	Summary and Concluding Remarks	161
Part III	Machine Learning and Optimization of an Ion Induction Linac	163
Chapter 9.	Introduction and System Overview	164
Chapter 10.	Auto Tuning Algorithm for Safe Extremum Seeking in NDCX-II	168
Chapter 11.	Data Collection, Retrieval, and Preprocessing	170
Chapter 12.	Training a Neural Network to Predict Peak Current and Arrival Time Using Detector Waveform Data	173
12.1	Neural Network Architecture	173
12.2	Training and Validation	180

12.3 Summary and Concluding Remarks	189
BIBLIOGRAPHY	191

Part I

Linear Ion Accelerators

Chapter 1. Introduction

The application of accelerator technologies is manifold and classes of accelerator system are diverse [1, 2, 3]. Accelerators are powerful tools that play a pivotal role in advancing scientific research, medical diagnostics and treatments, industrial processes, and our understanding of the fundamental building blocks of matter. At the heart of accelerator technologies is the principle of imparting kinetic energy to a directed beam of charged particles using applied electric and magnetic fields. There are two primary types of accelerators: linear accelerators (linacs) and circular accelerators (synchrotrons and cyclotrons). Linear accelerators propel particles in a straight line and typically thread focusing and accelerating elements once per cycle. Circular accelerators use magnetic fields to guide particles along a circular path allowing the beam to go through each element many times. Particle accelerators find applications in various fields. In high-energy physics, enormous accelerators like the Large Hadron Collider (LHC) at CERN are used to explore the fundamental particles that make up the universe. In medicine, smaller accelerators are employed for cancer treatment through techniques like tumor therapy using the beam to kill cancer cells or alternatively by using the beam to generate radio isotopes for targeted killing of cancer cells. Industrial processes benefit from accelerators in areas such as materials science, where they are used for material analysis and modification. Additionally, accelerators have become integral in generating intense beams of light, known as synchrotron radiation, for advanced imaging and spectroscopy applications driving materials science, biology, and new drugs in medicine. The development of accelerator technologies has been marked by continuous innovation and collaboration among scientists and engineers worldwide. From the early days of particle

accelerators in the 20th century to the cutting-edge technologies of today, the field has witnessed remarkable advancements. These include the integration of superconducting magnets, precise beam control systems, and the exploration of novel acceleration techniques, such as plasma acceleration. As we delve into the intricacies of accelerators and accelerator technologies, we discover not only the immense impact they have on scientific discovery but also their vital role in shaping the technological landscape of the modern world. The ongoing pursuit of more efficient, cheaper, and powerful accelerators promises to unlock new frontiers in scientific exploration and technological applications, fueling progress across diverse fields.

Chapter 2. Longitudinal Dynamics

Longitudinal beam dynamics describes the evolution of axial degrees of freedom of charged particle beams in accelerators as the beam gains kinetic energy in the acceleration lattice. Control of the longitudinal motion of particles within an accelerator is essential for achieving precise energy and maintaining stability for optimal performance. The longitudinal dynamics of charged particle beams primarily centers on analysis of the gain in kinetic energy and its spread as the beam evolves in the lattice. As particles traverse an accelerator, they experience kinetic energy changes due to the applied accelerating fields as well as interactions between particles comprising the beam. Understanding and controlling the longitudinal dynamics of the beam is crucial for achieving desired beam characteristics such as beam quality, stability, and beam lifetime.

A simple acceleration gap can be a pair of electrodes with one of the pair held at ground and the other held at an applied voltage bias. This will establish a direct current (DC) or static electric field between the two and as particles enter through the center aperture they will experience an accelerative force. If these acceleration gaps are then distributed in a straight line then this accelerator system design is called a linear accelerator or linac which was first proposed by Ralph Wideröe in 1928 [4]. DC acceleration can be simple, but is limited by voltage breakdown which occurs at a few tens of megavolts. To circumvent the problem of voltage breakdown, time-varying fields can be applied which gives rise to the radio-frequency (RF) linac. The history and development of the RF linac is beyond the scope of this dissertation but can be found elsewhere [2]. This chapter is focused on the key equations in analyzing the longitudinal dynamics and the considerations involved in using

an RF linac.

2.1 The Transit Time Factor

The on axis electric field $E_z(r = 0, z, t)$ for an RF cavity is given by [4]

$$E_z(r = 0, z, t) = E_z(r = 0, z) \cos(\omega t + \phi), \quad (2.1)$$

with RF angular frequency $\omega = 2\pi f_{\text{rf}}$, and phase shift ϕ . The time the particle is at position z is given by

$$t(z) = \int_{-g/2}^{g/2} \frac{dz}{v_z(z)} + \text{const.} \quad (2.2)$$

We choose time $t = 0$ to be when the particle is at the center of the gap at $z = 0$ and the field relative to the crest is at ϕ . Using the paraxial approximation, the axial velocity $v_z \simeq v$ is much greater than the transverse velocities with $v_x \approx v_y$ and $v \gg v_x$. If the RF gap is centered at $z = 0$ and the field confined between $-L/2$ and $L/2$, then, denoting $E_z(r = 0, z) \equiv E_z(0, z)$, the gain in kinetic energy $\Delta\mathcal{E}$ of a charged particle moving on axis is

$$\Delta\mathcal{E} = q \int_{-L/2}^{L/2} E_z(0, z) \cos(\omega t(z) + \phi) dz, \quad (2.3)$$

with charge q . Using a common trig identity, Eq. 2.3 is given by

$$\Delta\mathcal{E} = q \int_{-L/2}^{L/2} E_z(0, z) [\cos \omega t(z) \cos \phi - \sin \omega t(z) \sin \phi] dz.$$

Is we define the RF voltage as

$$V_0 \equiv \int_{-L/2}^{L/2} E_z(0, z) dz, \quad (2.4)$$

then the energy gain can be written in a deceptively simple form known as Panofsky's equation [4]:

$$\Delta\mathcal{E} = qV_0T \cos \phi. \quad (2.5)$$

The transit time factor T in Eq. 2.5 can be found by first rewriting the cosine term in terms of cosine and sine using the trigonometric identities (see Refs. [4, 5]). Then, multiplying and dividing by V_0 gives

$$T \equiv \frac{\int_{-L/2}^{L/2} E_z(0, z) \cos(\omega t) dz}{V_0} - \tan \phi \frac{\int_{-L/2}^{L/2} E_z(0, z) \sin(\omega t) dz}{V_0}. \quad (2.6)$$

The time-variation in the electric field will always reduce the maximum energy gain that can be obtained from a static field in the gap. This can be seen directly by comparing the numerators and denominators in Eq. 2.6. If $T \approx 1$ then the particle will experience an approximately static field and a lower value of T implies more variation the particle experiences in the gap. In most cases the electric field profile $E_z(0, z)$ is symmetric about the gap center and the second term in Eq. 2.6 integrates to zero. Furthermore, if the energy gain in the gap is small compared to the particle's kinetic energy, then the velocity is roughly constant (perturbative energy gain) and the integral for $t(z)$ reduces to $t(z) \approx L/v_z$. With

these two assumptions Eq. 2.6 reduces to

$$T = \frac{\int_{-L/2}^{L/2} E_z(r=0, z) \cos\left(\frac{2\pi L}{\beta\lambda_{rf}}\right) dz}{V_0}, \quad (2.7)$$

where $\beta = v_z/c$ and $\lambda_{rf} = c/f$. Note that T is dependent on the width of the gap and the speed of the particle. This suggests that gaps be designed thin or use an applied voltage that is small compared to the particle's kinetic energy. A useful approximation is to assume the electric field is hard-edge so that it is of a constant value E_g within the gap and zero outside

$$E(r=0, z) = \begin{cases} E_g & -g/2 \leq z \leq g/2, \\ 0 & \text{otherwise.} \end{cases}$$

Computing the integral gives

$$T = \frac{\sin\left(\frac{\pi g}{\beta\lambda_{rf}}\right)}{\left(\frac{\pi g}{\beta\lambda_{rf}}\right)} = \text{sinc}\left(\frac{\pi g}{\beta\lambda_{rf}}\right). \quad (2.8)$$

With Eq. 2.8 one can quickly calculate and understand the interplay between the particle velocity and the acceleration gap geometry. This equation will be used in later sections to justify use of a static approximation to the electric field within the gap.

2.2 RF Synchronism for Particle Acceleration

If $T \approx 1$ then the energy gain of a particle progressing through the gap will only depend on the phase of arrival $\phi = \omega t$. The distance from the $(n-1)^{\text{th}}$ gap center to the n^{th} gap center will be d_n . Before arriving, the design particle will have a velocity $\beta_{s,n-1}c$ from the

previous gap at a phase of $\phi_{s,n-1}$. At the next gap center, the particle will arrive at $\phi_{s,n}$ based on the equation:

$$\phi_{s,n} = \phi_{s,n-1} + \frac{\omega d_n}{\beta_{s,n-1} c}. \quad (\text{design particle})$$

Similarly, for an arbitrary particle the equation is:

$$\phi_n = \phi_{n-1} + \frac{\omega d_n}{\beta_{n-1} c}. \quad (\text{arbitrary particle})$$

An RF gap can operate in two modes. If all RF gaps are in phase with each other then this is referred to as the O-mode (“oh”-mode). The RF gaps can also operate such that each successive gap is 180° out of phase relative to the previous gap and is called the π -mode (“pi”-mode) which will produce a more compact structure. The mode type can be incorporated into the previous equations by:

$$\phi_{s,n} = \phi_{s,n-1} + \frac{\omega d_{n-1}}{\beta_{s,n-1} c} + \begin{cases} 0, & \text{O-Mode} \\ \pi, & \pi - \text{mode} \end{cases}, \quad (2.9a)$$

$$\phi_n = \phi_{n-1} + \frac{\omega d_{n-1}}{\beta_{n-1} c} + \begin{cases} 0, & \text{O-Mode} \\ \pi, & \pi - \text{mode} \end{cases}. \quad (2.9b)$$

The system this dissertation is concerned with uses the π -mode operation for a more compact structure. Therefore, from here on the π -mode convention will be used and case notation dropped. The synchronous particle, by definition, arrives at each gap at the synchronous phase and thus the phase difference between each pair of gaps should remain π . This condi-

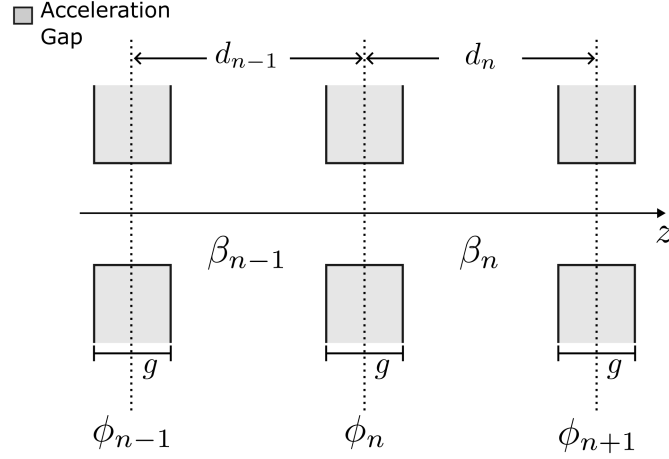


Figure 2.1: Schematic for accelerating cells in linac structure. The particle arrives at each gap center at a phase ϕ_{n-1} and is accelerated which changes the particle velocity from β_{n-1} to β_n . The particle will then progress to the next acceleration center and arrive at a phase of ϕ_n .

tion can be used to rewrite d_{n-1} in Eq. 2.9 for the arbitrary particle in terms of the design particle parameters. First, using the definition of the synchronous particle:

$$d_{n-1} = \frac{\beta_{s,n-1} c \pi}{\omega}$$

Substituting for d_{n-1} in Eq. 2.9 gives:

$$\phi_n = \phi_{n-1} + \pi \frac{\beta_{s,n-1}}{\beta_{n-1}} + \pi.$$

It is useful to find the relative phase difference as particles travel from gap to gap. This can be done by subtracting the synchronous phase equation from the arbitrary phase condition yielding

$$(\phi_n - \phi_{s,n}) - (\phi_{n-1} - \phi_{s,n-1}) = \pi \beta_{s,n-1} \left(\frac{1}{\beta_{n-1}} - \frac{1}{\beta_{s,n-1}} \right).$$

Next, we calculate the particle energy from the velocity. First, the β -terms on the right can be manipulated as follows:

$$\begin{aligned}\beta_{s,n-1} \left(\frac{1}{\beta_{n-1}} - \frac{1}{\beta_{s,n-1}} \right) &= - \left(1 - \frac{\beta_{s,n-1}}{\beta_{n-1}} \right) \\ &= - \left(\frac{\beta_{n-1} - \beta_{s,n-1}}{\beta_{n-1}} \right) \\ &= \frac{\Delta\beta_{n-1}}{\beta_{s,n-1} + \Delta\beta_{n-1}}.\end{aligned}$$

Here, $\Delta\xi = \xi - \xi_s$ denotes the relative difference between some measure ξ of an arbitrary particle and the design particle. Generally, the difference $\Delta\beta_{n-1}$ will be small compared to $\beta_{s,n-1}$ and the denominator in the last line of the equation above can be approximated as $\beta_{s,n-1} + \Delta\beta_{n-1} \approx \beta_{s,n-1}$. Using this approximation and substituting into Eq. 2.9 gives:

$$\Delta\phi_n - \Delta\phi_{n-1} \approx \pi \frac{\Delta\beta_{n-1}}{\beta_{s,n-1}}.$$

The β -terms can now be written in terms of the kinetic energy $\mathcal{E} = (\gamma - 1)mc^2$ with $\gamma = (1 - \beta^2)^{-1/2}$. Using these identities and some algebraic manipulations, one can show that $\Delta\mathcal{E} = \gamma_s^3 \beta_s mc^2 \Delta\beta$. The relative change in the gap-to-gap phase can be written in terms of the relative change in kinetic energy as:

$$\Delta\phi_n - \Delta\phi_{n-1} = \frac{-\pi}{\gamma_{s,n-1}^3 \beta_{s,n-1}^2} \left(\frac{\Delta\mathcal{E}_{n-1}}{mc^2} \right) \quad (2.10)$$

A similar equation can be derived for the gap-to-gap kinetic energy difference $\Delta\mathcal{E}_n - \Delta\mathcal{E}_{n-1}$. To do this, we use the Panofsky equation given by Eq. 2.5 and perform an analogous differ-

encing procedure to obtain

$$\Delta\mathcal{E}_n - \Delta\mathcal{E}_{n-1} = qV_{0,n}T_n(\beta_{s,n})[\cos\phi_n - \cos\phi_{s,n}]. \quad (2.11)$$

Together, Eq. 2.10 and Eq. 2.11 form a pair of coupled finite difference equations that can be used to model the relative change in phase, which can be converted to time with correct bookkeeping of the design particle parameters and energy.

The difference equations can be converted into coupled continuous differential equations under the assumption that the gap-to-gap changes are small. The steps to do so can be found in multiple places (in particularly see file 09 in Ref. [6]). Making the conversion to differential form results in

$$(\gamma_s\beta_s)^3 \frac{d\Delta\phi}{dz} = \frac{-2\pi}{\lambda_{\text{rf}}} \left(\frac{\Delta\mathcal{E}}{mc^2} \right), \quad (2.12a)$$

$$\frac{d\Delta\mathcal{E}}{dz} = qE_0T(\cos\phi - \cos\phi_s), \quad (2.12b)$$

with the average electric field over the gap defined as $E_0 = V_0/g$ being used in the second line. Taking another derivative of Eq. 2.12a and substituting $\Delta\mathcal{E}/dz$, results in a second order nonlinear differential equation for $\Delta\phi$

$$\frac{d}{dz} \left[(\gamma_s\beta_s)^3 \frac{d\Delta\phi}{dz} \right] = \frac{-2\pi}{\lambda_{\text{rf}}} \frac{qE_0T}{mc^2} [\cos(\phi_s + \Delta\phi) - \cos\phi_s]. \quad (2.13)$$

Further analysis can be done by making two approximations. Assume that $\Delta\phi$ is small ($\Delta\phi \ll 1$) and the energy change is small $d(\gamma_s\beta_s)/dz \approx 0$. Making these approximations

allows one to derive a conserved Hamiltonian H_ϕ :

$$H_\phi = \frac{1}{2}(\Delta\phi')^2 + \frac{k_s^2}{2}(\Delta\phi)^2, \quad (2.14)$$

where $' \equiv d/dz$ and the synchrotron wave number k_s is given by

$$k_s = \sqrt{\frac{2\pi}{\lambda_{rf}} \frac{qE_0T \sin(-\phi_s)}{mc^2(\gamma_s\beta_s)^3}}. \quad (2.15)$$

Note the sine argument under the square-root is $-\phi_s$. Here, the convention that the RF-accelerating field in the gap ranges from $-\pi \leq \phi \leq \pi$ and that from $-\pi \leq \phi < 0$ the electric field is rising. At $\phi_s = 0$ the field is a maximum and then decreases to the minimum value at $\phi_s = \pi$. Eq. 2.14 immediately lends intuition about the dynamical behavior of a particle subject to the assumptions made (small $\Delta\phi$ and small energy gain). Additionally, regions of stable oscillations can be identified by applying those same approximations to Eq. 2.13 in which the equation is reduced to:

$$\frac{d^2\Delta\phi}{dz^2} + k_s^2\Delta\phi = 0.$$

This is a simple harmonic oscillator equation and one can immediately see that amplitude oscillations around the design particle occur for $k_s^2 > 0$, or equivalently $-\pi < \phi_s < 0$ while the oscillations are unstable for $0 < \phi_s < \pi$. Armed with these insights, additional analysis can be done to include acceleration and the impact on the longitudinal beam dynamics. Reverting back to Eq. 2.13 and only assuming what is needed for the continuous limit, the

derivatives can be taken to give:

$$\frac{d^2\Delta\phi}{dz^2} + 3\frac{(\gamma_s\beta_s)'}{(\gamma_s\beta_s)}\frac{d\Delta\phi}{dz} = \frac{-2\pi}{\lambda_{rf}}\frac{qE_0T}{mc^2\gamma_s^3\beta_s^3}\left[\cos(\phi_s + \Delta\phi) - \cos(\phi_s)\right] \quad (2.16)$$

The synchrotron Hamiltonian H_ϕ for this nonlinear equation can be derived when the acceleration rate is small $(\gamma_s\beta_s)' \approx 0$ as

$$H_\phi = \frac{A}{2}\left(\frac{\Delta\mathcal{E}}{mc^2}\right)^2 + B(\sin\phi - \phi\cos\phi_s). \quad (2.17)$$

where:

$$A \equiv \frac{2\pi}{\lambda_{rf}(\gamma_s\beta_s)^3}, \quad B \equiv \frac{qE_0T}{mc^2}.$$

Examining H_ϕ , the first term represents the “kinetic energy” while the second term is the “potential” $V_\phi = B(\sin\phi - \phi\cos\phi_s)$. This identification allows one to find stable regions for the potential energy by examining the first and second derivatives of V_ϕ :

$$\begin{aligned} \frac{\partial V_\phi}{\partial\phi} &= B(\cos\phi - \cos\phi_s), \\ \frac{\partial^2 V_\phi}{\partial\phi^2} &= -B\sin\phi. \end{aligned}$$

The following operating regions for ϕ_s are apparent:

1. $-\pi < \phi_s < -\pi/2$: deceleration and unstable longitudinal defocusing,
2. $-\pi/2 < \phi_s < 0$: acceleration and stable longitudinal focusing,
3. $0 < \phi_s < \pi/2$: acceleration and unstable longitudinal defocusing ,
4. $\pi/2 \leq \phi_s \leq \pi$: deceleration and unstable longitudinal defocusing.

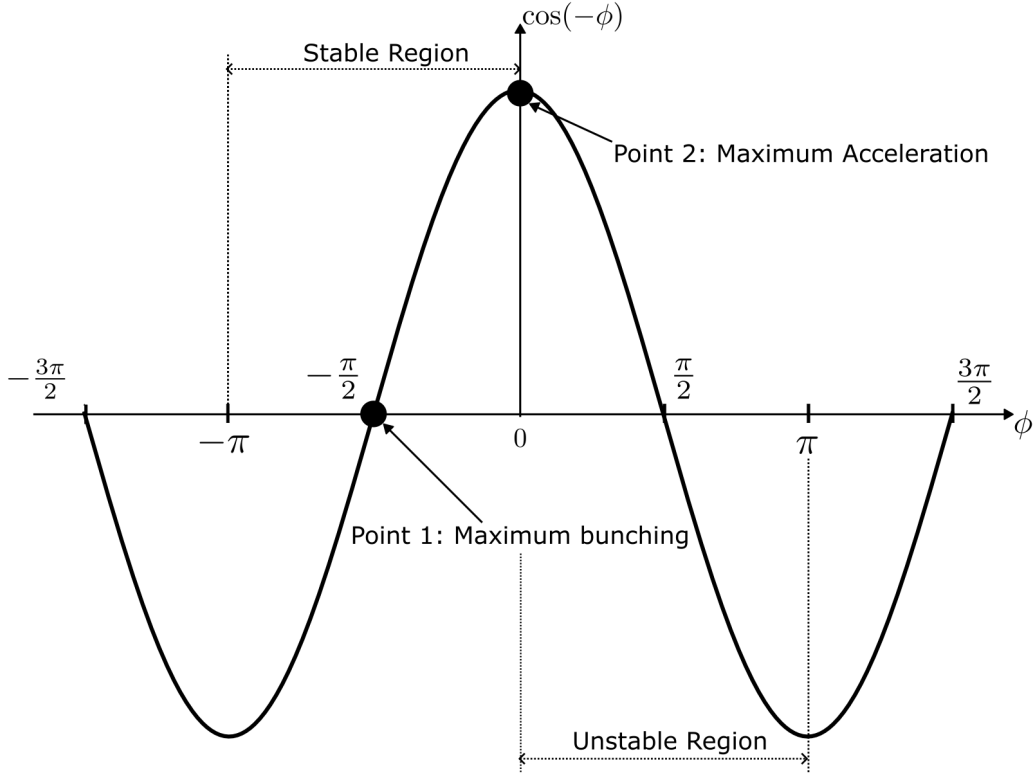


Figure 2.2: Operating phase convention typically used for the cosine oscillating electric field in the gap. The design phase is chosen to be between $-\pi \leq \phi_s \leq \pi$. Examination of the Hamiltonian H_ϕ shows that the region between $-\pi/2 < \phi < 0$ is a region of stability. Points 1 and 2 represent operation points for maximum bunching and maximum acceleration respectively. When operating at Point 1 with $\phi_s = -\pi/2$, the maximum amount of particles will be captured in the RF-bucket and there will be no overall acceleration of the beam. At Point 2 with $\phi_s = 0$ the design particle is maximally accelerated but the RF-bucket will be zero width and no particles will be captured. Between these two points there will be both acceleration and focusing (bunching).

Two operating points for ϕ_s are useful landmarks for navigating the operating region. The first point (Point 1 in Fig. 2.2) is when $\phi_s = -\pi/2$ and is used for maximum focusing or bunching of the beam. To see this, consider three particles. The first particle will have slightly less energy than the design particle ($\Delta\mathcal{E} < 0$) and will be trailing behind the design particle as the particles progress to the next gap ($\Delta\phi > 0$). Call this slower particle particle-S. The second particle will have the opposite conditions and will be leading the design particle, this fast particle will be labeled particle-F. Lastly, there is the design particle.

At $\phi_s = -\pi/2$, the design particle will not be accelerated. However, particle-F will arrive before the design particle and see a decelerating field. This will cause the design particle to pass particle-F. On the flip-side, particle-S will arrive late and see an accelerating field. Particle-S will gain energy and catch up with the design particle. Extrapolating this logic to a multitude of particles, this operating phase will prevent particle losses by bunching the slow and fast particles around the design particle. For $\phi_s = -\pi/2$, focusing occurs for the full RF period around the design particle. The other landmark is when the design particle is maximally accelerated at $\phi_s = 0$ (Point 2 in Fig. 2.2). Applying similar logic, both particle-F and particle-S will see a lower magnitude electric field than the design particle and there is no focusing for max acceleration. If there are many gaps with $\phi_s = 0$, the design particle will surpass all particles and $\Delta\phi_s$ will tend to negative infinity for all particles.

The conservation of H_ϕ implies that the phase-dynamics of particles will evolve on nested closed-contours within a bounding contour that partitions stable phase-space evolution from unstable. This bounding contour is called the separatrix. The separatrix extent can be found by solving for the stable and unstable fixed points in phase-space, ϕ_1 and ϕ_2 , where $\Delta\mathcal{E} = 0$. First, evaluating the point where $\partial V_\phi/\partial\phi = 0$ gives the maximum potential where $\Delta\mathcal{E} = 0$. This first point at $\phi_1 = -\phi_s$, can then be used to find the constant value of H_ϕ . Then, upon substitution into Eq. 2.17, one has the following equation for ϕ_2 :

$$\frac{A}{2} \left(\frac{\Delta\mathcal{E}}{mc^2} \right)^2 + B(\sin\phi_2 - \phi_2 \cos\phi_s) = -B(\sin\phi_s - \phi_s \cos\phi_s).$$

With $\phi_1 = -\phi_s$ and ϕ_2 as constrained above, the separatrix is defined and the stable region within the separatrix is referred to as the RF-bucket. The stable region defined by the separatrix will form a fish-shape structure as depicted in the central plot on the left panel in

Fig. 2.3. The two stable points defining the separatrix are shown as dashed-lines intersecting the ϕ -axis at ϕ_2 and $-\phi_s$. The panel to the right is the separatrix when $\phi_s = -\pi/2$, corresponding to maximum bunching as discussed before. Here, the contours make an eye-shape and the maximum width bucket is the full RF period phase-width of 2π . Lastly, if the acceleration term in Eq. 2.16 is not neglected then there is damping in the equation and

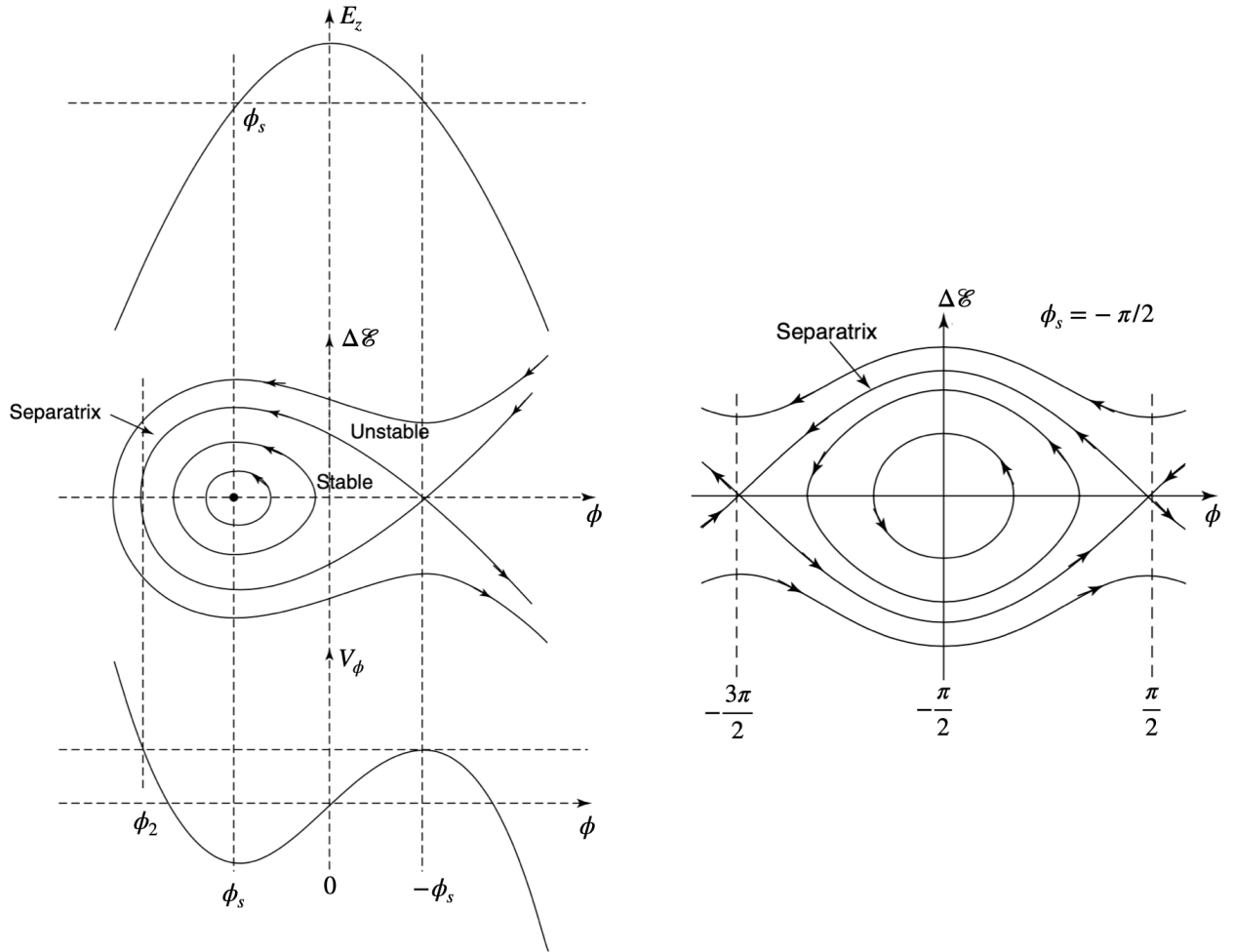


Figure 2.3: Image borrowed from Wangler [4]. The left panel shows three plots as a function of ϕ . The top plot is the electric field in the gap. The middle plot is the $\Delta\mathcal{E}$ - ϕ phase-space contours of $H_\phi = \text{constant}$ surfaces. Lastly, the bottom plot is the potential term V_ϕ in the Hamiltonian H_ϕ . The phase-space contour plot labels the region of stable orbits, or RF-bucket, that reside within the outer most stable contour called the separatrix. If operating at maximum bunching $\phi_s = -\pi/2$ then the phase-space contours form an eye shape as depicted by the right plot. Operating phase for stability covers the full RF-bucket.

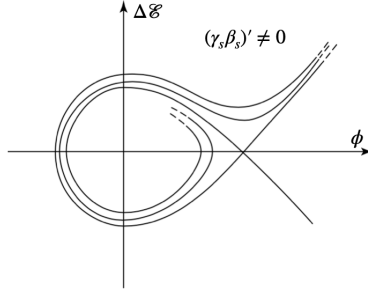


Figure 2.4: Image borrowed from Wangler [4]. Acceleration in the system will introduce damping to the phase-space orbits and the fish structure in Fig. 2.4 morphs into a “golf club” structure.

H_ϕ is no longer conserved. When this term is present, the phase-space plot looks more like a golf club structure as shown in Fig. 2.4 [4]. Additionally, when H_ϕ is no longer conserved the separatrix can no longer be identified and though the RF-bucket is no longer defined, the analysis done for the non-acceleration case can be employed to give a prescription for optimizing transport.

2.3 Summary

This chapter summarized the conceptual ideas for longitudinal dynamics in a linac system. In doing so, important parameters were defined. Additionally, the conventions used that will be used in this thesis were reviewed. The concepts described here provide a useful guide for navigating the challenging operational landscape we are to face (see Ch. 6) for the RF MEMS systems with strong acceleration. Lastly, this material is by no means complete. More thorough discussion can be found in many texts with Refs. [4, 6, 7] being particularly good.

Chapter 3. Transverse Dynamics

Transverse particle dynamics describes the evolution of charged particles in the degrees of freedom transverse to the longitudinal axis. It encompasses the study of particle trajectories in phase-space in response to the focusing and defocusing forces acting on particle beams in response to both applied and self fields. In the realm of particle accelerators, transverse particle dynamics plays a crucial role in the design, operation, and optimization of these complex machines. Understanding the transverse dynamics of particles is essential for controlling and manipulating the bundle of particle trajectories within accelerators to avoid losses and degradations in beam quality. Defocusing effects arise from interactions with various electromagnetic fields in the system such as the inter-particle (space charge) that tend to expand the beam. forces. These forces beam expansion resulting in a degraded beam and possible system damage from *charged* particles impinging on the system walls and being lost and inducing damage. It is essential to comprehend these effects to achieve control over the beam's characteristics such as its intensity, size, and shape. Transverse focusing is critical to ensuring that particles remain confined within the system. Focusing elements, such as quadrupole field optics and potentially other electromagnetic lenses, are employed to control the beam divergence thus allowing for the concentration of particles to maintain a tight and collimated beam. The primary focusing elements the system in this thesis are the electrostatic quadruples (ESQs) and RF focusing provided by the acceleration gaps.

This chapter will review the theory of transverse beam dynamics that is necessary for understanding the models and results in later sections. It is by no means a comprehensive review. For a more thorough treatment of the subject matter, there are many fine textbooks

[4, 7] as well as open access course notes provided by the United States Particle Accelerator School (USPAS) [8] and the CERN document server [9]. The chapter is organized as follows. We start by examining the equations of motion in the transverse plane for a single particle in applied and self-electric and self-magnetic fields. Assumptions are made to approximate both the applied fields and the self-fields in ideal cases. The single-particle case will yield a set of second order non-linear differential equations that can be used to model the particle motion in the presence of applied fields self-field defocusing forces of the beam. To describe a distribution of particles, a statistical approach is built from the single particle approach.

3.1 Single Particle Dynamics

It is convenient to first consider a single charged particle of mass m and charge q . This particle is moving in the presence of electric and magnetic fields for which there are two sources. The first source comes from the applied fields that arise from the various conducting objects in the system used to accelerate, focus, and steer the beam. These fields are denoted \vec{E}^a and \vec{B}^a for the applied electric and magnetic fields respectively. The second source comes from the surrounding beam that the single particle under consideration is evolving in the presence of. This beam is comprised of other charged particles that each produce an electric and magnetic field. Solving each individual particle's field is intractable. However, various reasonable assumptions can be made and the influence of neighboring particles can be modified by a continuous distribution. This distribution of charges can be analyzed and the self-field can be calculated. These self-fields are denoted \vec{E}^s and \vec{B}^s for the self-electric and self-magnetic field respectively. The two sources can be added together so that the individual particle moves in a total electric field $\vec{E} = \vec{E}^a + \vec{E}^s$ and a total magnetic field

$\vec{B} = \vec{B}^a + \vec{B}^s$. Denoting the particle's momentum as \vec{p} and the speed of light as c , the single particle's evolution is described by the Lorentz force equation:

$$\begin{aligned} \frac{d\vec{p}}{dt} &= q[\vec{E}(\vec{r}, t) + c\vec{\beta} \times \vec{B}(\vec{r}, t)], \\ &= q[\vec{E}^a(\vec{r}, t) + c\vec{\beta} \times \vec{B}^a(\vec{r}, t)] + q[\vec{E}^s(\vec{r}, t) + c\vec{\beta} \times \vec{B}^s(\vec{r}, t)], \\ &= \vec{F}^a(\vec{r}, t) + \vec{F}^s(\vec{r}, t), \end{aligned}$$

where the applied and self-field forces are given by $\vec{F}^a(\vec{r}, t)$ and $\vec{F}^s(\vec{r}, t)$ respectively. The particle's velocity is expressed in terms of β where $\vec{v} = \vec{\beta}c$. The particle momentum can be written as

$$\vec{p} = \gamma mc\vec{\beta},$$

where $\gamma = (1 - 1/\beta^2)^{-1/2}$. Thus far, no approximations have been made outside of many beam particles to be modeled by a continuous distribution. The first approximation will be the so called paraxial approximation. In this approximation the beam, and therefore the particle being considered, is assumed to have small transverse velocities relative to the axial velocity—axial meaning along the accelerator system. Symbolically this means $v_{\perp} \ll v_z$ where the \perp symbol stands for the perpendicular components, which in this case are the x and y components. Additionally, it is more informative to transform the time derivative into a derivative along the beam line also known as Frenet-Serret coordinates. In the case of a Linac system, the longitudinal beam coordinate is the axial coordinate z . Denoting $' \equiv d/dz$ and analyzing the transverse components only, the momentum can be rewritten as:

$$\vec{p}_{\perp} = \gamma m v_z \vec{x}'_{\perp}.$$

The time derivative becomes $d/dt \approx v_z d/dz$ due to the paraxial approximation and we have:

$$mv_z \frac{d(\gamma v_z x'_\perp)}{dz} = \gamma m v_z^2 x''_\perp + m v_z (\gamma v_z)' x'_\perp.$$

Rewriting in terms of $\beta = v_z/c$ and including the force terms gives:

$$x''_\perp + \frac{(\gamma\beta)'}{(\gamma\beta)} x'_\perp = \frac{1}{\gamma m \beta^2 c^2} (F_\perp^a + F_\perp^s). \quad (3.1)$$

To progress, information is needed about both the applied and self fields. A reasonable assumption is to assume that the beam is comprised of a uniform distribution of particles that is cylindrically symmetric. To calculate the self-field, Gauss' law can be applied. For a cylinder of uniform charge with radius r_b , beam current I_b , and beam velocity v_b , Gauss' law gives ¹:

$$E_r^2 = \begin{cases} \frac{I_b r}{2\pi\epsilon_0 r_b^2 v_b}, & r \leq r_b \\ \frac{I_b}{2\pi\epsilon_0 v_b r}, & r > r_b, \end{cases} \quad (3.2)$$

where ϵ_0 is the vacuum permittivity. For a cold beam the particles' longitudinal velocity spread is small and thus $v_b \approx v_z$. Under the assumption that the charge is uniformly distributed, it follows that the current density is of the form $\vec{J}^s \approx \beta_b c \rho(x_\perp, t) \hat{z}$ with $\beta_b = v_b/c$.

The self-magnetic field is obtained through applying Ampère's law:

$$B_\theta^s = \begin{cases} \frac{\mu_0 I_b r}{2\pi r_b^2}, & r \leq r_b \\ \frac{\mu_0 I_b}{2\pi r}, & r > r_b, \end{cases} \quad (3.3)$$

¹The derivation here closely follows steps taken in other sources such as those found in Refs. [7, 10, 11, 12]

where μ_0 is the vacuum permeability. At this point, it is convenient to convert the equations of motion into cylindrical coordinates where r is the radius and θ is the angle in x - y plane. Denoting the time derivative $\dot{\chi} = d\chi/dt$, the radial component of the Lorentz force equation gives:

$$\frac{d(\gamma_b m \dot{r})}{dt} - \gamma_b m r \dot{\theta}^2 = q(E_r^s + r \dot{\theta} B_z^s - \dot{z} B_\theta^s),$$

where $\gamma \approx \gamma_b$ is the gamma factor calculated for $v = v_b \approx v_z$. A few assumptions can be made. First, since the applied fields are not being considered then there is no acceleration and $\gamma \approx \text{const}$. Second, we take the beam to be slowly rotating such that $r \dot{\theta}$ can be assumed negligible. These assumptions reduce the above equation to

$$\gamma m \ddot{r} = q E_r^s - q \dot{z} B_\theta^s.$$

Using $c^2 = 1/(\epsilon_0 \mu_0)$ the self-fields together give

$$E_r^s - \beta_b c B_\theta^s = \frac{I_b r}{2\pi \epsilon_0 \gamma_b^2 v_b} \frac{r}{r_b^2}.$$

Additionally, transforming the time derivative to a spatial one gives $\ddot{r} = \beta_b^2 c^2 r''$. Combining and simplifying gives

$$r'' = \frac{q I_b}{2\pi \epsilon_0 m \gamma_b^3 \beta_b^3 c^3} \frac{r}{r_b^2} \quad (3.4)$$

Factors in the right-hand-side of Eq. 3.4 can be lumped into a dimensionless parameter called the generalized perveance Q :

$$Q = \frac{q I_b}{2\pi \epsilon_0 m \gamma_b^3 \beta_b^3 c^3}. \quad (3.5)$$

Rewriting Eq. 3.4 in terms of Q gives:

$$r'' = \frac{Q}{r_b^2} r.$$

Since $Q > 0$ for ions, this shows that self-fields act to expand an ion beam. Because it is only dependent on a handful of constants, the beam current, and beam energy, Q is a useful parameter in describing the space charge forces present. The beam edge radius r_b^2 will depend on the radial beam extent and will change depending on the size of the beam, thereby modulating the space charge force.

In addition to the paraxial approximation, it was assumed that the particles were uniformly distributed and cylindrically symmetric. It was further assumed that the beam is slowly-rotating so that the term $r\dot{\theta}$ was negligible. We will review a uniform density elliptical beam next but will have to reanalyze since we no longer have cylindrical symmetry.

3.1.1 Calculating the Self-Fields for an Elliptical Beam with Uniformly Distributed Charge Density

The previous section provides a recipe for calculating the self-field of the beam. Fig. 3.1 gives a schematic of the a uniform density elliptical beam. The beam has edge lengths of r_x and r_y with a uniform number density of particles given by $n(z)$ within the ellipse. The area of the ellipse is $\pi r_x(z)r_y(z)$, the line-charge density is a constant given by $\lambda = qn(z)\pi r_x(z)r_y(z) = \text{const.}$ There is constant current density $\vec{J} = \rho\beta_b c\hat{z}$ within the elliptical beam. A derivation of the self-field potential is given in [14] shows that within the

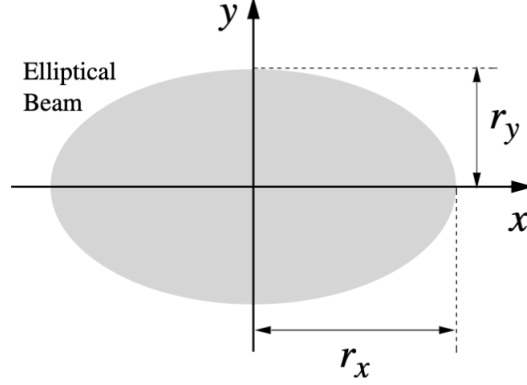


Figure 3.1: An elliptical beam of uniformly distributed particles within the ellipse. There is a constant current density $\vec{J} = \rho\beta_b c \hat{z}$ within the elliptical beam. The beam radii are given by r_x and r_y as shown. Image borrowed from Ref. [13].

elliptical beam ²:

$$\phi^s = -\frac{\lambda}{2\pi\epsilon_0} \left[\frac{x^2}{(r_x + r_y)r_x} + \frac{y^2}{(r_x + r_y)r_y} \right]. \quad (3.6)$$

Taking the negative gradient gives the transverse components of the electric field for the self-field:

$$E_x^s = \frac{\lambda}{\pi\epsilon_0} \frac{x}{(r_x + r_y)r_x}, \quad (3.7a)$$

$$E_y^s = \frac{\lambda}{\pi\epsilon_0} \frac{y}{(r_x + r_y)r_y}. \quad (3.7b)$$

Repeating the steps used for the cylindrical beam, the self-fields can be plugged into the Eq. 3.1 resulting in:

$$x''(z) = \frac{q\lambda}{\pi\epsilon_0 m \gamma_b^3 \beta_b^2 c^2} \frac{x}{(r_x + r_y)r_x},$$

$$y''(z) = \frac{q\lambda}{\pi\epsilon_0 m \gamma_b^3 \beta_b^2 c^2} \frac{y}{(r_x + r_y)r_y}.$$

²The URL provided in the Ref.[14] also has a video of the USPAS lecture given by Dr. Lund. This video is helpful in explaining how the non-trivial potential is derived.

In the cylindrical case, Q was defined in Eq. 3.5 and its usefulness comes from only depending on the beam parameters I_b and \mathcal{E}_b . The same Q factor is in the elliptical case. Noting that $\lambda = I_b/(\beta_b c)$, the self-field expressions can be rewritten in terms of Q as:

$$x''(z) = \frac{2Q}{(r_x + r_y)r_x}x, \quad (3.8a)$$

$$y''(z) = \frac{2Q}{(r_x + r_y)r_y}y. \quad (3.8b)$$

The right-hand-side of Eq. 3.8 is the contribution from the self-fields of an elliptical beam and will be used in following sections.

3.1.2 Forces from Applied Electrostatic Quadrupole (ESQ) Fields

We begin by considering a purely transverse electric field in vacuum. Following the steps in Ref. [12], the purely transverse potential can be expanded as a complex series since it satisfies the Cauchy-Riemann conditions. Using $\underline{z} = (x + iy)$ with $i = \sqrt{-1}$ and denoting the complex potential $\underline{P}(\underline{z})$ the series is written as:

$$\underline{P} = \sum_{n=0}^{\infty} \underline{C}_n \underline{z}^n,$$

where the coefficients \underline{C}_n are, in general, complex numbers such that $\underline{C}_n = (a_n + ib_n)$. The complex conjugate of the electric field $\underline{E}^* \equiv E_x - iE_y$ is compliant with the static vacuum Maxwell's equations. Taking the derivative of the potential gives the electric field:

$$\underline{E}^* = E_x - iE_y = i \frac{\partial \underline{P}}{\partial \underline{z}}.$$

Taking derivative of the series gives:

$$i\frac{\partial P}{\partial z} = i \sum_{n=1}^{\infty} n \underline{C}_n z^{n-1} = i \underline{C}_1 + 2i \underline{C}_2 z + 3i \underline{C}_3 z^2 + \dots$$

The first coefficient can be found taking the field at $z = 0$. Expanding gives:

$$E_x - iE_y = (ia_1 - b_1) + 2(ia_2 - b_2)(x + iy) + 3(ia_3 - b_3)(x + iy)^2 + \dots$$

From this expansion, the first two expansion coefficients for a_n and b_n can be found as:

$$a_2 = \frac{1}{2} \frac{\partial E_y}{\partial y} \Big|_{x=0=y}; \quad b_2 = -\frac{1}{2} \frac{\partial E_x}{\partial x} \Big|_{x=0=y}.$$

Lastly, it is common to normalize the fields so the strength can be expressed relative to the particle energy. The factor $[B\rho]$ (“Brho” or rigidity) is common and defined as:

$$[B\rho] \equiv \frac{\gamma_b \beta_b m c}{q} \tag{3.9}$$

For a pure quadrupole field, the dipole term coefficients a_1 and b_1 will be zero and the leading term will be the a_2 and b_2 coefficients. This will give the quadrupole term and the corresponding normalized transverse fields will be:

$$\frac{E_x}{\beta_b c [B\rho]} = -\frac{q}{\gamma_b \beta_b^2 m c^2} \frac{\partial E}{\partial x} \Big|_{x=0=y} x,$$

$$\frac{E_y}{\beta_b c [B\rho]} = \frac{q}{\gamma_b \beta_b^2 m c^2} \frac{\partial E}{\partial y} \Big|_{x=0=y} y.$$

For a perfect quadrupole, the gradients are equal and opposite. Denoting the on-axis gradient G as $G(z) = -\partial E_x/\partial x|_{x=0=y} = \partial E_y/\partial y|_{x=0=y}$, the linear applied fields (E_x^a and E_y^a) from a perfect pure transverse quadrupole are

$$\frac{E_x^a}{\beta_b c [B\rho]} = -\frac{qG}{\gamma_b \beta_b^2 mc^2} x = -\kappa(z)x, \quad (3.10a)$$

$$\frac{E_y^a}{\beta_b c [B\rho]} = \frac{qG}{\gamma_b \beta_b^2 mc^2} x = \kappa(z)y, \quad (3.10b)$$

where the focusing function $\kappa(z)$ has been introduced and is defined as:

$$\kappa(z) = \frac{qG}{\gamma_b \beta_b^2 mc^2} \quad (3.11)$$

for a purely transverse ESQ. Note that we are allowing the gradient function $G(z)$ to vary in z to account for the fringe field of short elements.

3.1.3 The Kapchinskij-Vladimirskij (KV) Envelope Equations

Eq. 3.1 is the general form for the equations of motion with the applied forces F^a and self-field forces F^s . Using Eqs. 3.8 and 3.10, the full single particle equation of motion can be written for the transverse directions as:

$$x'' + \frac{(\gamma_b \beta_b)'}{(\gamma_b \beta_b)} x' + \kappa(z)x - \frac{2Q}{(r_x + r_y)r_x} x = 0, \quad (3.12a)$$

$$y'' + \frac{(\gamma_b \beta_b)'}{(\gamma_b \beta_b)} y' - \kappa(z)y - \frac{2Q}{(r_x + r_y)r_y} y = 0. \quad (3.12b)$$

Recall that the particle is assumed to be moving within an elliptical beam of uniform particle density with major axes r_x and r_y . Additionally, applied ESQ focusing is provided by a

quadrupole with purely transverse fields. Note that the applied ESQ focusing is equal and opposite in the two planes so that while one plane is being focused by the ESQ the other plane is defocused. In general, the focusing can also be from other field elements (such as acceleration gaps) and the equations of motion can be rewritten to reflect this with separate x - and y -plane focusing functions as:

$$x'' + \frac{(\gamma_b \beta_b)'}{(\gamma_b \beta_b)} x' + \kappa_x(z)x - \frac{2Q}{(r_x + r_y)r_x} x = 0, \quad (3.13a)$$

$$y'' + \frac{(\gamma_b \beta_b)'}{(\gamma_b \beta_b)} y' + \kappa_y(z)y - \frac{2Q}{(r_x + r_y)r_y} y = 0. \quad (3.13b)$$

Additional insight can be gleaned by analyzing Eq. 3.13a for different cases.

Case A: For quadrupole focusing with $\kappa_x = \kappa(z)$ $\kappa_y = -\kappa_y$, assume no acceleration $\gamma_b \beta_b \approx \text{const.}$ and that space charge effects are negligible with $Q \approx 0$. Lastly, take $\kappa(z)$ to be periodic such that, for a lattice period L_p , $\kappa(z + L_p) = \kappa(z)$.

With these assumptions, the particle equations of motion take the form of Hill's equations [15]:

$$x(z)'' + \kappa(z)x(z) = 0,$$

$$y(z)'' - \kappa(z)y(z) = 0.$$

Hill's equation can be solved using a phase-amplitude formulation [7, 10, 12]. We assume a solution of the form $x(z) = A(z) \cos \psi(z)$ which we then plug into Hill's equation to give:

$$x'' + \kappa(z)x = [A'' + \kappa A - A\psi'^2] \cos \psi - [2A'\psi' + A\psi''] \sin \psi = 0.$$

Since we have used two functions, $A(z)$ and $\psi(z)$, to describe the single function $x(z)$, we are free to introduce a constraint. We will choose the coefficient for $\sin \psi$ to be 0 giving the first constraint:

$$2A'\psi' + A\psi'' = 0. \quad \text{Constraint 1}$$

This leads to a second constraint for the $\cos \psi$ coefficient to satisfy Hill's equation for all z :

$$A'' + \kappa A - A\psi^2 = 0. \quad \text{Constraint 2}$$

Using Constraint 1, one can show that $A^2\psi' = \text{const}$. We then rescale $A(z)$ such that $A(z) = A_i w(z)$ where A_i is the initial amplitude and $w(z)$ is a real-valued function. Constraint 1 then gives $w^2\psi' \equiv 1$. Rewriting $x(z)$ in terms of these functions gives:

$$x(z) = A_i w(z) \cos \psi(z), \quad (3.14a)$$

$$x'(z) = A_i w'(z) \cos \psi(z) - \frac{A_i}{w(z)} \sin \psi(z), \quad (3.14b)$$

with a similar solution for $y(z)$. Here A_i is the initial amplitude, $w(z)$ is a real-valued function that we can take $w \geq 0$, and $\psi(z)$ is a real-valued phase function. A useful dimensionless parameter called the undepressed phase advance σ_0 can be calculated by integrating Constraint 1 over a single lattice period L_p :

$$\sigma_0 \equiv \Delta\psi(z) = \int_{z_i}^{z_i+L_p} \frac{d\tilde{z}}{w^2(\tilde{z})} \quad (3.15)$$

where $\Delta\psi(z) = \psi(z) - \psi_i(z_i)$ and z_i is the starting z -location. Quantities ψ , w , and σ_0 can be subscripted x and y to distinguish the x and y -planes if needed. The phase advance

σ_0 provides a normalized measure of focusing strength and $\sigma_0 < 180^\circ/\text{period}$ is necessary for single particle stability. One can intuitively understand σ_0 by taking what is called a smooth approximation to $\kappa(z)$. In this approximation, $\kappa(z)$ is “smeared” out such that it is a constant value over L_p . In this case, the solution to Hill’s equations are the common harmonic solutions and $\kappa(z)$ is redefined as the undepressed betatron wavenumber $\kappa_{\beta 0}$ which is constant ($\kappa(z) = \kappa_{\beta 0}^2 = \text{const}$). Then $\sigma_0 = \kappa_{\beta 0} L_p$. It should be stressed that σ_0 only depends on the lattice properties and not the initial conditions of the particle. The usefulness of σ_0 is put succinctly in Dr. Lund’s lecture notes [10]:

The phase advance σ_0 is an extremely useful dimensionless measure to characterize the focusing strength of a periodic lattice. Much of conventional accelerator physics centers on focusing strength and the suppression of resonance effects. The phase advance is a natural parameter to employ in many situations to allow ready interpretation of results in a generalizable manner.

Case B: We keep the same assumptions as Case A except we include space charge so that $Q \neq 0$.

Incorporating the space charge shifts $\kappa_{\beta 0}$ resulting in the depressed betatron wavenumber κ_{β}^2 and depressed phase advance σ :

$$\kappa_{\beta}^2 = \left(\kappa_{\beta 0} - \frac{2Q}{(r_x + r_y)r_x} \right)^2.$$

Space charge will act to dampen the transverse oscillations. One can quantify the degree of space charge in system by calculating σ/σ_0 . A value close to zero represents a space charge dominated beam (cold beam) while a value close to 1 represents a thermally dominated beam. For continuous focusing, with both planes equal focusing we can take $r_x = r_y = r_b$

and $\kappa_\beta^2 = (\sigma/L_p)^2 = \kappa_{\beta 0}^2 - Q/r_b^2$ to estimate the depressed phase advance σ .

The phase-amplitude formulation can be used to form the Courant-Snyder (CS) invariant [14, 16]. Recall that A_i was the initial amplitude and some manipulation of Eq. 3.14 will yield the invariant:

$$\left(\frac{x}{w}\right)^2 + (wx' - w'x)^2 = A_i^2 = \text{const.} \quad (3.16)$$

From the quadratic in x, x' CS invariant we can see that the particle's trajectory in $x - x'$ phase space is described by a set of nested ellipses. The CS invariant is commonly expressed by defining the so called Twiss Parameters:

$$\alpha(z) \equiv -w(z)w'(z), \quad (3.17a)$$

$$\beta(z) \equiv w^2(z), \quad (3.17b)$$

$$\gamma(z) \equiv \frac{1}{w^2(z)} = \frac{1 + \alpha(z)^2}{\beta(z)^2}. \quad (3.17c)$$

Rewriting Eq. 3.16 in terms of the Twiss Parameters gives:

$$\gamma x^2 + 2\alpha x x' + \beta x'^2 = A_i^2, \quad (3.18)$$

which is the standard form of a rotated ellipse in $x-x'$ phase-space. The Twiss Parameters are functions $w(z)$ and $w'(z)$ which are functions of the lattice only and the size A_i set by the initial conditions. Therefore, the shape and orientation of the CS invariant are functions

of the lattice. One can integrate over the ellipse to get the phase-space area obtaining:

$$\text{Area} = \int_{\text{ellipse}} dx dx' = \pi A_i^2,$$

$$A_i^2 \equiv \epsilon,$$

where the last line defines the single particle emittance ϵ . The single particle phase-space area $\pi\epsilon$ is a conserved quantity and is widely used in accelerator physics. Fig. 3.2 is an drawing of the $x - x'$ ellipse with various points of interest labeled in terms of the Twiss Parameters. In terms of ϵ , the phase-space solutions to Hill's equation are

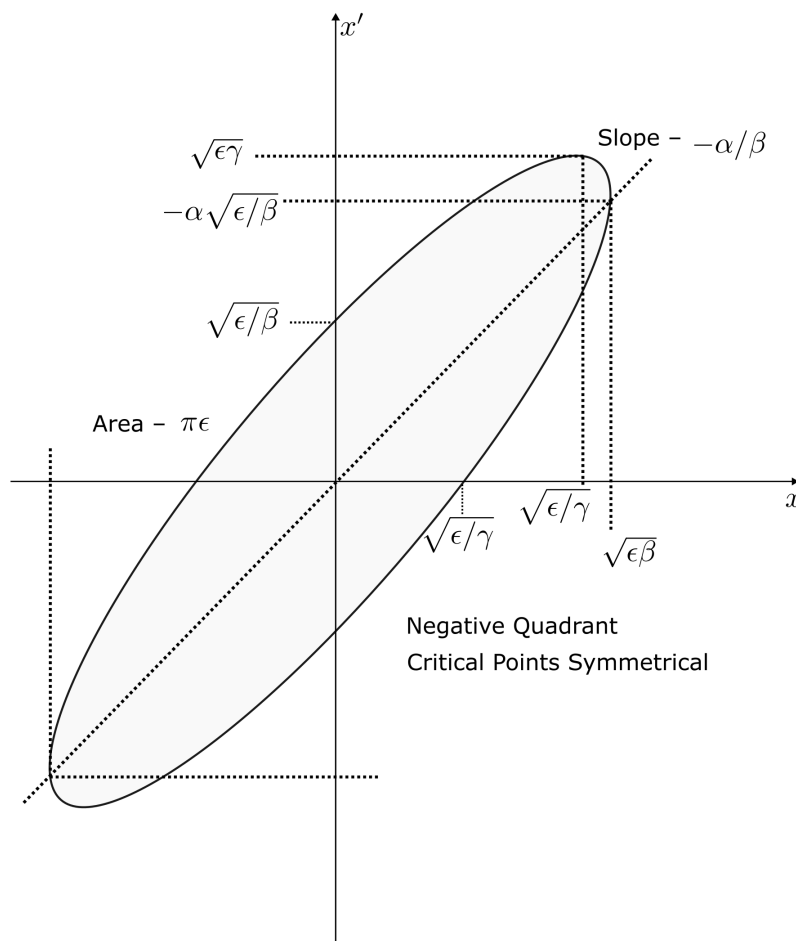


Figure 3.2: Phase space CS ellipse in $x - x'$ with points of interest labeled in terms of the Twiss Parameters α , β , and γ . Figure adopted from Ref. [10].

$$\begin{aligned}
x(z) &= \sqrt{\epsilon_x} w_x(z) \cos \psi(z), \\
y(z) &= \sqrt{\epsilon_y} w_y(z) \cos \psi(z),
\end{aligned}$$

where we have labeled the x and y plane quantities with the relevant subscripts.

The boundary in x and y can be found by setting $\cos \psi(z) = \pm 1$ in the solution, which for the maximum value of single particle emittance gives the beam edge as:

$$\begin{aligned}
r_x(z) &= \sqrt{\epsilon_x} w_x(z), \\
r_y(z) &= \sqrt{\epsilon_y} w_y(z).
\end{aligned}$$

Rewriting Hill's equations in terms of r_x and r_y results in a set of coupled, second-order, nonlinear differential equations that describe the evolution of the envelope edge known as the as the Kapchinskij-Vladimirskij (KV) envelope equations [17]. If the acceleration is included (see Ref. [10] for transformation needed to include acceleration in KV equations), then the full KV equations for r_x and r_y for linear applied fields are

$$\begin{aligned}
r_x''(z) + \frac{(\gamma_b \beta_b)'}{(\gamma_b \beta_b)} r_x + \kappa_x(z) r_x(z) - \frac{2Q}{r_x(z) + r_y(z)} - \frac{\epsilon_x^2}{r_x^3(z)} &= 0, \\
r_y''(z) + \frac{(\gamma_b \beta_b)'}{(\gamma_b \beta_b)} r_y + \kappa_y(z) r_y(z) - \frac{2Q}{r_x(z) + r_y(z)} - \frac{\epsilon_y^2}{r_y^3(z)} &= 0.
\end{aligned} \tag{3.19a}$$

A statistical analysis can be carried out to derive a self consistent distribution that satisfies the underlying assumptions: uniform density within an elliptical beam. Such a distribution exists and is called the KV distribution [17] after Kapchinskij and Vladimirskij. The KV distribution is the only solution to the time-dependent Vlasov equation (see [7, 10]) when the applied and self-fields are linear, and the transverse and longitudinal equation

of motions are decoupled. Since it is a solution to Vlasov's equation, the 4D transverse phase-area is conserved, i.e., the transverse emittance.

The details of the KV distribution (its derivations and physical interpretations) are fascinating but fairly complex. We neglect this discussion as it would take up a significant digression from the main topics of this dissertation. However, complete lecture notes and a video discussion can be found in the space charge course taught at USPAS [10] and in a review paper authored by one of the course lecturers [18]. Using the KV distribution, the statistical root-mean-square (rms) envelope edges and angles can be calculated from the statistical moments giving:

$$\begin{aligned} r_x &= 2\sqrt{\langle x^2 \rangle_{\perp}}, & r_y &= 2\sqrt{\langle y^2 \rangle_{\perp}}, \\ r'_x &= 2\frac{\sqrt{\langle xx' \rangle_{\perp}}}{\sqrt{\langle x^2 \rangle_{\perp}}}, & r'_y &= 2\frac{\sqrt{\langle yy' \rangle_{\perp}}}{\sqrt{\langle y^2 \rangle_{\perp}}}, \end{aligned} \quad (3.20)$$

where $\langle \dots \rangle_{\perp}$ denotes a transverse statistical average over the 4D phase-space. Lastly, the statistical rms-edge emittance is given by:

$$\epsilon_x = 4\epsilon_{x,\text{rms}} = 4[\langle x^2 \rangle \langle x'^2 \rangle - \langle xx' \rangle^2]^{1/2}, \quad (3.21a)$$

$$\epsilon_y = 4\epsilon_{y,\text{rms}} = 4[\langle y^2 \rangle \langle y'^2 \rangle - \langle yy' \rangle^2]^{1/2}, \quad (3.21b)$$

These edge measures of beam statistical size r_x , r_y , r'_x , and r'_y and phase space area ϵ_x and ϵ_y in Eqs. 3.20 and 3.21 are also used when the beam is not KV.

For the case of a KV beam:

$$\gamma_b \beta_b \epsilon_x = \text{const}, \quad (3.22a)$$

$$\gamma_b \beta_b \epsilon_y = \text{const}. \quad (3.22b)$$

In the presence of acceleration, Eq. 3.22 is applied in integrating the KV envelope equations. Generally the emittance will not be conserved in an acceleration lattice and the acceleration gaps will dampen the emittance.

We cannot hope to achieve a KV distribution as it is a 4D hyper-ellipsoid. Nonetheless, the KV model representing a real distribution in an rms equivalent sense can function well. The issue is stated succinctly by Reiser [7]:

The K-V theory is a good and very useful approximation for beams where the current remains well below the space charge limit. This is true for practically all accelerators and other devices . . . It must be kept in mind, however, that the KV model does not include nonlinear effects that increase the emittance.

This prompts the question: how can we possibly model a real accelerator system with equations that assume such a distribution and that will change the emittance due to intrinsic non-linear effects? Lapostolle [19] and Sacherer [20] developed the concept of equivalent beams (discussed in [7]) which allows one to study the KV equivalent of a laboratory beam of the same species, current, and kinetic energy if second moments of the distribution are the same. Furthermore, this approximate equivalence can be further justified if the emittance stays approximately the same over the lattice.

3.2 Summary

This chapter presented a brief overview of the key concepts and equations that are of importance to the transverse dynamics analysis that will be presented in Part II of this dissertation—especially Ch. 7. Derivations for the transverse equations of motion in the presence of linear applied and self field were sketched. Key parameters such as the generalized perveance Q and rms-edge emittance ϵ_{rms} were defined. Of particular importance are the KV envelope equations, which will be the primary description used in Ch. 7 to model the RF MEMS accelerator system in the presence of accelerating gaps and ESQ focusing. There we will find that although the beam is not a KV distribution, the KV model in an rms equivalent beam sense will accurately model our situation.

Chapter 4. Self Consistent Simulations

Warp [21, 22] is a multi-dimensional particle-in-cell (PIC) code with an extensive hierarchy of models that was developed to model high current, high brightness beams for heavy-ion driven inertial confinement fusion (HIF). However, the software is not limited to HIF modeling. Warp has extensive capabilities and can self-consistently model a wide variety of accelerator systems. Detailed conducting elements can be placed on the mesh with time-varying or static voltages applied and with various options for transverse and/or longitudinal boundary conditions. In a nutshell, the software will first perform a field-solve on the user-defined mesh for all the conductor objects placed at subgrid resolution. Particles, either single species or mixed, will then be injected and deposited on the mesh based on user settings. Once deposited, the self-field is calculated by finding a global solution to Poisson's equation (including conductors with present biases) and the particles are advanced using a 2nd order leap-frog algorithm. A variety of symmetry options are available to reduce field solve times and options exist to implement adaptive mesh refinement (AMR) for more efficient simulations by increasing grid resolution in regions where finer detail is needed while keeping less interesting areas at a grainer resolution to save computational time. Warp also has a variety of pre-defined conductor objects as well as the capability to create conductors through its conductor building tool-kit. The numerically intensive parts of the code are written in Fortran, but the code tools are wrapped within a Python interface for easier use. This allows very flexible code setup and running with dynamic steering. Diagnostics are written mostly in Python to allow flexible adaptation to problems.


```

1  import warp as wp
2  # Mesh Creation and time step
3  wp.w3d.zmin = minimum z
4  wp.w3d.zmax = maximum z
5  wp.w3d.nz = number of grid points to use
6  wp.top.dt = time step
7  # ... same setup for x and y
8
9  # Set boundary conditions
10 wp.bound0 = boundary condition at minimum z (dirchlet, neumann,
reflect)
11 wp.boundnz = boundary condition at maximum z (dirchlet, neumann,
reflect)
12 wp.boundxy = transverse boundary conditions (dirichlet, neumann,
periodic,)
13
14 # Can also set particle boundary conditions
15 wp.top.pbound0 = ...
16 wp.top.pboundnz = ...
17
18 # Beam Specifications
19 beam = wp.Species(type=wp.Argon, charge_state=1, name="Ar+")
20 beam.ekin = beam kinetic energy
21 beam.ibeam = current of beam
22 ... # More specifications available like beam radius and
23 ... # angle, emittance, etc.
24
25 # Specify injection. The number of particles injected, particle weight

```

```

,
26 # injection type, beam distribution of injection and
27 # much more can be specified.
28
29
30 # Create and install conductors. Once
31 conductor = wp.Annulus(...)
32 wp.installconductors(conductor)
33
34 # Specify solver type. RZ, four-fold, two-fold, or
35 # full 3D options available.
36 solver = wp.MRBlock3D() # full 3D symmetry
37 solver.ldosolve = True # Solve for self-fields, i.e. include space
charge
38 wp.register(solver)
39
40 # Generate the PIC code, allocate storage for arrays,
41 # and do initial field solve
42 wp.package("w3d")
43 wp.generate()
44
45 # Advance particles. Depending on the settings, diagnostics will be
collected
46 # each time step and plots can be generated in a cgm file.
47 for i in range(desired_time_steps):
48     wp.step() # One time step advance
49     ...
50

```

```
51 # Do analysis. Various outputs are specified in the top.v file for a
    multitude of different calculations Warp performs based on the user-
    setting for the information to be gathered.
52 do analysis of simulation
```

Listing 4.1: Outline of a Python script setting up a simulation in Warp.

A typical script outline is given in pseudo-code in Lst. 4.1. Scripts can become highly complex and lengthy. But the flexible interface allows code tools to be adapted to the problem and rapid tool checks added without modifying and re-compiling source code.

Warp has multiple diagnostic tools and will store statistical moment histories if the options are set to do so. In particular, the `ZCrossingParticles` tool will be used in the data collection in later sections (see Ch. 8). This diagnostic acts as a beam position monitor (BPM) and will record particle data (transverse position, time of crossing, velocities, momentum, etc.) at the given z -location for particles crossing it. As invaluable as Warp is, it can be both difficult to setup and computationally expensive. As will be discussed later, using Warp exclusively in 3D simulation mode to optimize the system that this dissertation is concerned with would quickly become intractable. In this dissertation, we use Warp to help formulate and verify rapidly running reduced physics models to better optimize the system and then use full 3D Warp source to target simulation to verify final system performance. Thus, full detail Warp simulations are applied in the last step in the design and optimization before testing in real-experiment. Fig. 4.1 gives an ideal flow of simulation to experiment. Reduced physics models focusing on distinct aspects of the system are first developed using Warp in a focused setting and are used to probe system parameters and design. Optimization cases can be developed quickly with rapid runs clarifying physics to optimize choices. This then acts as a guide to use Warp for more sophisticated evaluation of combined effects

in the full system. The initial stages are done repeatedly and once there is established confidence in design, better optimized experiments can be carried out. Since experiments are typically more difficult and less flexible, this cycle is repeated less frequently. Ideally, this whole process is repeated until convergence is reached with a design expected to perform strongly. This dissertation is primarily focused on the first stage in the cycle, using reduced physics models to probe the parameter/design space which then guides more optimal full Warp simulations. In this thesis, we apply this cycle to generate better designs for future experiments. Unfortunately, project funding ran out before laboratory experiments could be carried out to experimentally verify improvements found. It is hoped that results will point to better approaches for future experiments should future funding become available.

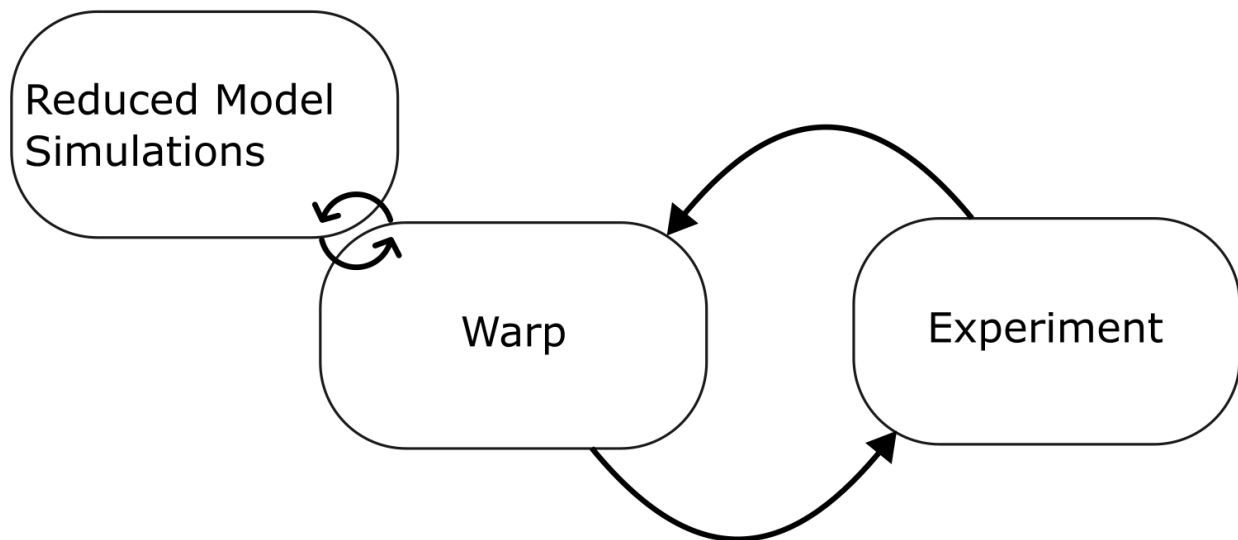


Figure 4.1: Flow of Warp simulations and experiment. Reduced models are constructed using Warp and applied to rapidly probe the parameter and design space. This narrows the scope so that the more costly 3D simulations can be ran with Warp for final integrated system performance. This first stage is repeated numerous times as signified by the tight double arrow. Once optimizations are carried out, the improved design can be verified with Warp before experimental evaluation. Because the experiment is more constrained, this second stage is iterated less frequently.

Part II

Compact Multi-Beam RF Linear Ion Accelerators

Chapter 5. RF-MEMS System

5.1 Introduction

Out of the accelerator zoo, we are concerned with the RF linac accelerator type. RF-linacs are pervasive in industrial and medical applications. Within the realm of possible applications there are multitudes of required performance needs. For example, medical applications for cancer therapy using linac accelerators require ions to be of the order of 10 MeV and focused on a spot size that is on the order of 1 cm^2 [23]. Ion implantation is another common use of RF-linac for which the range of energy and currents widely vary. Technologies require highly stable and collimated beams with currents ranging from a few μA to 100 mA and energies from 100 eV to $\sim 10 \text{ MeV}$ [24]. Additionally, accelerator technology is used in food sterilization, nondestructive testing and imaging, and propulsion systems for low mass $\sim 1 \text{ kg}$ satellite systems. With this in mind it is immediately obvious the benefit a compact and cost-effective RF-linac architecture with variability in the delivered current and energy would be. Reducing cost and size of a linac machine (or any complex system) has gained momentum due to advances in the field of micro electromechanical systems (MEMS). MEMS technology has revolutionized many fields and accelerator technologies are an excellent arena to enable applications. The development of compact and cost effective multi-beam array RF-linacs using MEMS technology (RF-MEMS) to accelerate an array of ion beams is the focus of Part II of this dissertation. An array of beams with compact cross section using MEMS technologies allows economical scaling to high net currents by simply using more beams to fit the application needs.

Part II is organized as follows. This chapter will introduce the RF-MEMS system. An overview of the system will be given with various references that do not fit the scope here. The acceleration array gaps and the ESQ arrays will be overviewed along with their associated fields. We will discuss the work done to optimize the ESQ array geometries such that the leading order error in the quadrupole field expansion is minimized. Once we have an optimum array geometry we investigate the ESQs further to quantify other useful properties for design such as the voltage breakdown limit and spacing properties. With a firm grasp of the overall system architecture and individual conductors, we proceed by developing reduced physics models used to efficiently simulate the system. These array optimization and reduced physics models employ fields extracted from realistic array elements modeled using Warp. This discussion begins with the longitudinal motion in Ch. 6. In this chapter, the computational tools developed and used to model the longitudinal phase-space is the primary focus along with models benchmarked using Warp. We then move on to analyzing the transverse motion in Ch. 7 and develop reduced physics models used to simulate the physics therein. We will show how the developed model is in excellent agreement with Warp and apply the model to design a lattice with improved transverse focusing to radially confine the beam. Additionally, the model is used to quantify useful performance metrics such as the phase advance σ and various measures of the beam envelope. Lastly, we combine optimization results of the previous chapters to perform 3D self consistent simulations with Warp in Ch. 8. Due to the complex nature of the accelerator system, the resulting self consistent simulations are complex, so we begin this chapter by overviewing the simulation. We then proceed gently by first carrying out simulations in a short lattice structure comprised of four acceleration gaps and a pulsed beam injection. This allows us to capture and observe much of the physics involved that is otherwise obscured by longer lattice structures and continuous

injection while characterizing important gap focusing properties. Finally, we perform full source to target simulations of a 16 gap design with continuous injection and discuss results. Each chapter ends with section that summarizes findings and also briefly describes areas of improvement and/or further research avenues. Unfortunately, lack of funding precluded experimental verification of advances in RF MEMS systems developed. However, results obtained promise much improved system performance in future experiments in the hopeful event that funding will be secured. Indeed, advances documented promise improved system performance to render many applications attractive.

5.2 Overview of Compact Multi-Beam MEMS Linac Structures

Although a linac is a technologically complex machine, its basic components are relatively simple. A schematic of the multi-beam RF-MEMS device is shown in Fig. 5.1 [25, 26, 27]. At the left of the schematic is the ion source. The source is housed in an insulated “cage”

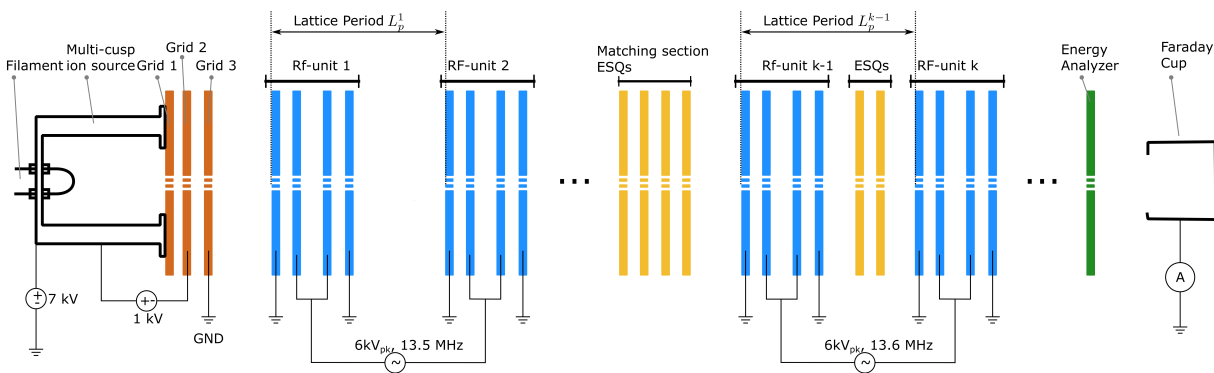


Figure 5.1: RF-MEMS Linac schematic. Note that a single RF-unit is comprised of two RF acceleration gaps. Two single RF-units are connected to the same voltage source but multiple voltage sources will allow for frequency changes down the accelerator line. After the matching section, the Linac can be decomposed into a periodic lattice consisting of two RF-units.

and floated to 7–10 kV. Within the source, an ion gas is then heated by a filament to create a predominantly singly ionized plasma which is extracted to produce an array of beams (beamlets) of ions at about 7 keV and a current of $\sim 10 \mu\text{A}$ per beamlet. A distinguishing feature of the RF-MEMS system is that an array of beamlets are accelerated rather than a single beam in conventional linac devices. Also, the conductor elements are wafers of etched and machined printed circuit boards (PCBs) which reduces the cost of manufacture substantially and allows for batch manufacturing. The injection is continuous (CW) unless stated otherwise. Fig. 5.2 shows a close up image of an accelerating wafer with a 120 beamlet array. Note that neighboring beam apertures are closely spaced so the transverse footprint does not rapidly scale with beam number. Following the source are RF-units which are comprised of two RF-acceleration gaps operating 180° out of phase relative to one another

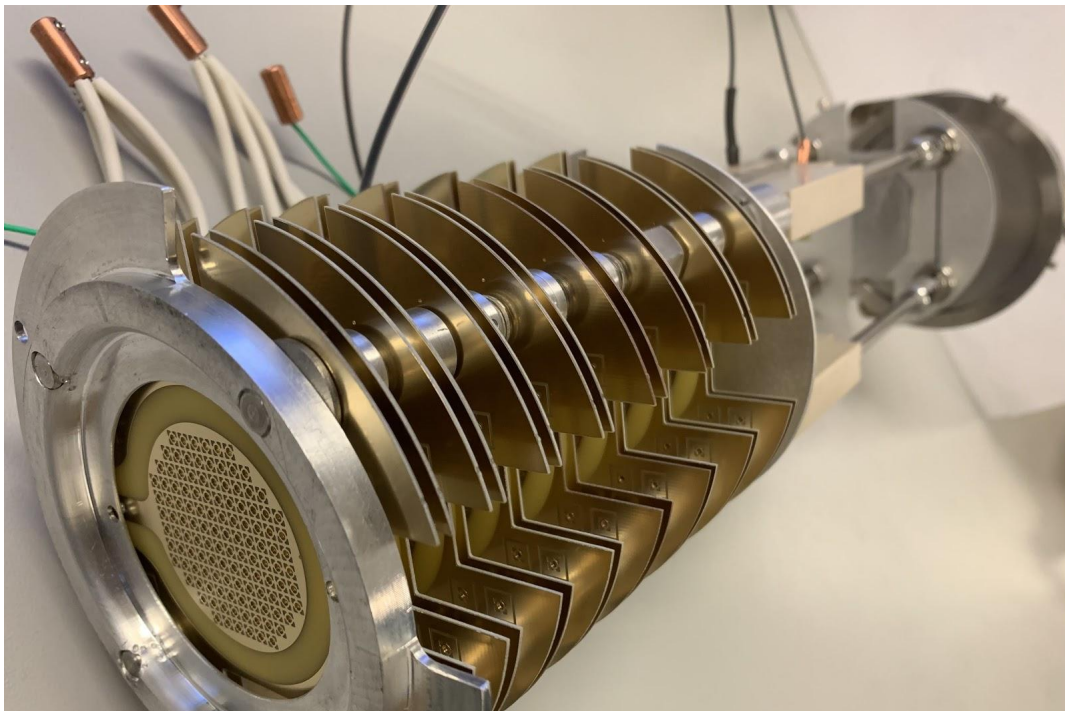


Figure 5.2: Acceleration gap wafers for a 120 beamlet array design. The 8 acceleration gaps are mounted on a four rods that hold the wafers in place. This design offers an easy way of adjusting the spacing between wafers to manipulate the RF phasing by changing the distance from one gap center to another.

to implement π -mode acceleration. This π -mode design choice allows for better longitudinal compactness of the system. Note that this first section of RF-gaps does not have periodic doublet focusing but acceleration gaps after the matching section do. This unorthodox design is chosen because we find that the acceleration gaps provide strong transverse focusing and can be used immediately with extra quadrupole focusing following extraction till the energy gains where the effect becomes weaker. This design will be further explained in the Ch. 7. The acceleration gap width is about 2 mm in axial length and the individual gap wafers are about 700 μm thick. A typical RF gap accelerating voltage is 7 kV but is operation dependent and can be within the range of 5–7 kV. Accelerating voltages are applied with a sinusoidal variation at a frequency of 13.6 MHz. Note that the gap voltage is comparable to the injection voltage so that initial rate of acceleration is very high. This contrasts the usual acceleration situation where gap energy changes are perturbative.

After the matching section are RF-units with periodic doublet electrostatic quadrupole (ESQ) focusing that lies in a field-free region where ESQs can be inserted. The schematic shows a doublet configuration, but if there is space and more focusing is needed, then ESQ wafers can be “stacked” end to end. The Focusing strength depends on the beam energy and the ESQ bias as outlined in Ch. 7. The ESQ bias V_q will need to increase as the beam energy increases in successive lattice periods to maintain focusing strength, but must remain below breakdown. For the ESQ parameters employed, the breakdown voltage $V_{q,\text{br}}$ is estimated to be at approximately $V_{q,\text{br}} \sim 1 \text{ kV}$, and typically we seek to operate at the less than half this limit. With an ESQ wafer thickness of about 0.695 μm the ESQs are axially thin compared to conventional ESQs. In between the two acceleration portions of the system is an ESQ matching section that will match the beam from the initial gap-only phase where there is acceleration and no ESQ focusing, to the second phase where ESQ doublet focusing is applied

Initial/Source Values per Beamlet	
Ion Species	Ar ⁺
Beam Energy $\mathcal{E}_{b,i}$	7.0 keV
Beam Current I_b	10 μ A
Temperature T_b	0.1 eV
Rms Edge Emittance ϵ_{rms}	1.34 mm-mrad
Geometric Design Values	
Source Aperture Radius r_s	0.25 mm
Gap Aperture Radius r_p	0.55 mm
Gap Width g	2.0 mm
ESQ Rod Radius R	0.717 mm
ESQ Length ℓ_q	0.695 mm
Lattice Period L_p	~ 20 mm
Operational	
Gap Voltage V_g	~ 6 kV
ESQ Voltage V_q	~ 100 V
RF Frequency f_{rf}	~ 13.6 MHz
RF Period τ_{rf}	~ 73.5 ns
RF Wavelength λ_{rf}	~ 22.04 m

Table 5.1: Single beamlet parameters for the RF-MEMS system. Values shown as $\sim \#$ have some degree of variability due to operation conditions or changes in beamlet parameters, for example, L_p will change with increasing beam kinetic energy \mathcal{E}_b to maintain RF synchronism. The R listed is the optimal value found in Sec. 5.4.

every period. The matching section takes the round beam emerging from the injector and matches it to the needed alternating plane flutter to efficiently focus in the downstream ESQ and acceleration periods. The schematic shows four ESQs being used, but more can be inserted to better optimize. The whole system is in vacuum with a pressure of ~ 3 mTorr. At the end of the acceleration lattice is a collimating slit. After passing through the slit, the beam enters a deflector plate that acts as an energy analyzer deflecting ions of a certain energy into a Faraday cup (Fcup).

Table 7.1 collects key parameters per beamlet for the system which are organized into three groups: Source, Geometric, and Operational. Source parameters are initial values of

the individual beamlets after extraction. The geometric parameters are values of various components of the system such as the length of the acceleration gap and the radius of the clear bore aperture of each beamlet. Note that the lattice period is for the first period of the acceleration lattice and is generally not a fixed quantity. The listed value is for an initial kinetic energy of 7 keV and maximum acceleration in the first two gaps. Operational quantities relate to the system settings for acceleration and focusing. Also note that the gap voltage V_g , ESQ focusing voltage V_q , and RF-frequency f_{rf} are not fixed quantities and can be tuned for specific applications.

The RF-MEMS system can be separated into two project areas. One area focuses on the application of the RF power amplifier developed by Airity technologies [28]. Employing these RF power amplifiers to deliver consistent and stable accelerating voltages without damaging the circuitry proved challenging. Parameters employed here are consistent with viable operating conditions found. Another area focuses on the analysis and optimization of the accelerated beam along with the accelerating and focusing elements (RF acceleration gaps and focusing ESQs). It is this latter area that this dissertation is directed toward—to guide more optimal choices in future experiments.

5.3 Fields from RF Acceleration Arrays

The acceleration gaps are created by first starting with an FR-4 PCB with copper cladding on both sides [29]. Laser micro-machining is then used to pattern the PCB and drill an array of aperture holes (see Ref. [29] for more details). Fig. 5.3 is a schematic showing the array of acceleration apertures and a close up of an individual gap. The whole PCB is excluded for simplicity. Conducting material (in this case copper) is grey, the FR-4 dielectric

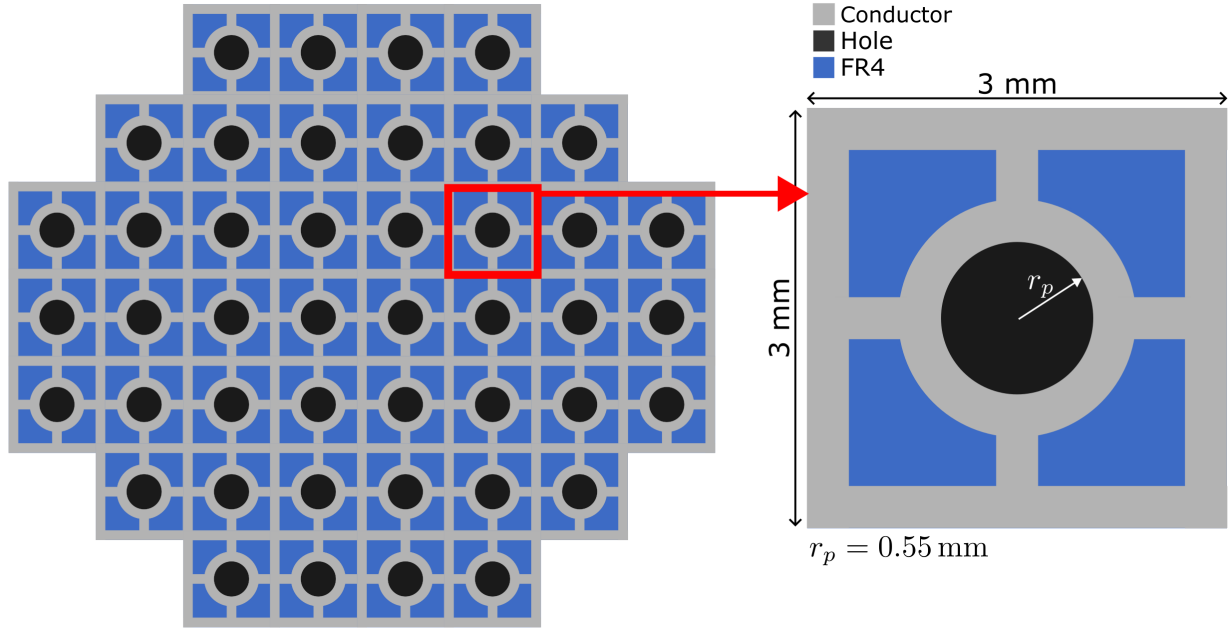


Figure 5.3: Front view of the array of acceleration gaps (left) and a close-up of a single gap (right). The conducting copper material is shown in grey. In blue is the FR-4 dielectric material that the PCB is comprised of. The PCB is copper cladded on both sides and the pattern and aperture hole (shown in black) created using laser cutting. Image adapted from Ref. [29].

material shown in blue, and black corresponds to vacuum apertures where the beamlets will be. The FR-4 material is not simulated and instead the blue area is treated as vacuum. The conducting rings surrounding the clear apertures has thickness of 0.3 mm. The connecting prongs have a length of 0.5 mm and width of 0.3 mm. Lastly, the surrounding square has a thickness of 0.5 mm. With the dimensions quantified, the conductor is created in the Warp code and the field is calculated. The gap geometry is relatively simple and there was not a priority to optimize the gap properties. Thus, these values are considered fixed quantities unless stated otherwise.

5.3.1 Electric Field Profile of the RF-Gap

In later sections the extracted on-axis electric field will be used to advance particles in a reduced physics model simulations that can be quickly ran and used to gain insight on the longitudinal and transverse dynamics. The Warp electrostatic field solver is applied to calculate the gap array RF field using a static bias V_g . The time dependent field is then obtaining by scaling a time modulation of the gap voltage. It is crucial that periodic boundary conditions are used to simulate the array geometry of the gaps. Fig. 5.4 shows the Warp-extracted z -component of the on-axis electric field normalized by the maximum amplitude ($E_z(r = 0, z)/\text{Max}[E_z(r = 0, z)]$) for two gaps that are 180° out of phase. The

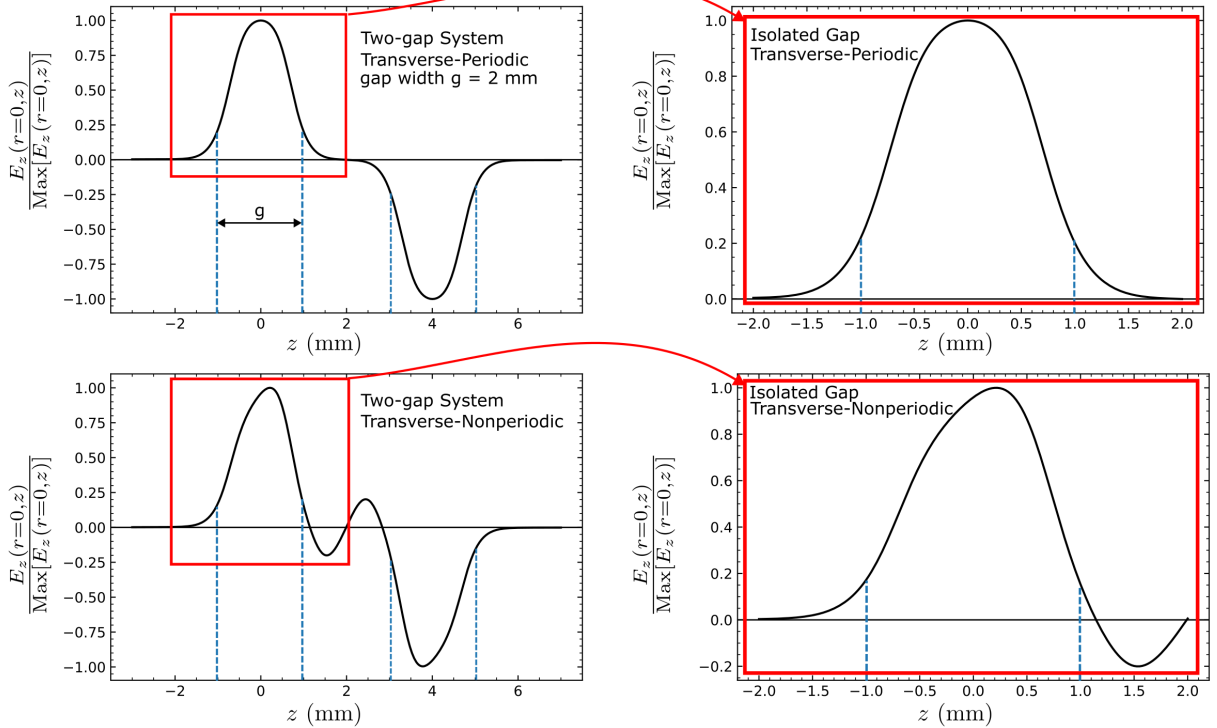


Figure 5.4: Extracted on-axis electric field $E_z(r = 0, z)$ using Warp for transverse-periodic boundary conditions (top row) and Dirichlet-boundary conditions (bottom row). Two RF-gaps are needed to correctly simulate the transverse periodic-array geometry. If a single gap is used, then the fields will be strongly altered by the boundary conditions as shown in the bottom row. The gap width g is labeled as blue vertical dash lines. The field data can be isolated by extracting the data within the red box.

top row shows the profile when the transverse conditions of the Warp field solver are set to periodic. The bottom row shows when the transverse conditions are set to Dirichlet or Neumann instead. It is apparent that these two profiles are drastically different and the choice will significantly impact the longitudinal dynamics, as well as the acceleration, of the particle. Periodic is the correct choice when modeling the beam *within* the array, such as that shown in Fig. 5.3. However, the non-periodic selection shows the field profile of the cells that surround the array. This suggests that it may be necessary to restrict beam acceleration to cells within the array and use the outer edge elements as dummy cells that still have an applied voltage but no beam current. Doing so will help ensure a desirable acceleration profile. An analysis was done to see how many layers are needed to surround a particular cell in the array so that the field is reasonable consistent. Fortunately, only one surrounding layer is needed. The top right image of Fig. 5.4 shows a sample extraction of $E_z(r = 0, z)$ in the periodic array. In later sections (see Ch. 6) the extracted field is used to advance particles in a 1D reduced physics model—that is, all particles are assumed on-axis and particles can only move in $\pm z$ along the accelerator axis. The normalized field can be scaled linearly to represent any applied gap RF voltage. Particles closer to the conductors (further off-axis) will generally experience a stronger magnitude field. This difference was less than a percent at approximately 80% of the aperture radial extent from the center axis of the aperture. Further from the center than that, the difference was only approximately 5% out to the radial edge. Therefore, we can safely ignore the added complexity of modeling off-axis change in acceleration that particles experience in an extended beam radial profile and assume the field-strength is uniform in radius from the center.

In addition to choosing the correct boundary conditions, to properly model the periodicity and extract the correct electric profile, a two-gap system must be simulated. Without

the second gap, the periodicity is not captured does and the field profile will be incorrect. Therefore, in order to extract a single gap profile, two gaps, are simulated. The extraction is done by properly isolating the field in either gap. This is done by extracting the upstream field from $[z_1, z_2]$ where z_1 is the start of the mesh boundary and z_2 is the geometric center of the two gaps. If the first gap is centered on $z = 0$ then $z_1 = -z_2$. With the isolated field of a single gap, a relationship can be established between the applied wafer voltage and the peak gap voltage and electric field that results. Assuming a linear relationship, the applied voltage V_a will be related to the on-axis peak voltage $V_p(r = 0)$ by $V_a = \xi V_p(r = 0)$ where ξ is some scaling parameter. Choosing any voltage, executing Warp, and then extracting the peak values yields the relationship for $r_p = 0.55$ mm:

$$V_a = 0.9962V_p(r = 0). \quad (5.1)$$

Eq. 5.1 can be used on the extracted normalized field to determine the scale factor for a desired voltage. Note that the on-axis voltage is slightly lower than the applied voltage and therefore, to get the desired energy gain, a slightly higher voltage will have to be used. This does not account for non-linear effects that may be present in the system, such as conductor defects, sparking, etc.

5.4 Optimized Electrostatic Quadrupole (ESQ) Focusing Array Fields

As the beam is transported through the system thermal spread, space charge forces, and transverse forces from the acceleration gap will cause the transverse extent of the beam to

expand. This beam expansion will reduce the beam intensity and lead to radial particle losses as the beam begins to scrape the system walls. Such beam losses can damage structures and charge dielectrics. To prevent these effects, ESQ doublet focusing can be added in the field free region between successive pairs of gaps. The ESQs here are comprised of cylindrical rods with rod radius R and axial length ℓ_q . Non optimal choice of R may result in significant contributions from the leading order error term in the multipole expansion giving undesirable nonlinear applied focusing. In this section, we discuss the development and application of a computational tool for modeling and optimizing the array of ESQs to minimize the leading order error term in the multipole field expansion while making a compact array. This section is organized as follows. In Sec. 5.4.1, the theoretical background and numerical procedure for calculating the multipole coefficients is discussed. Sec. 5.4.2 uses the developed computational tools to reproduce a known result for an isolated, long ESQ to benchmark the algorithm. We then analyze various geometrical parameters for short and radially packed arrays to optimize the field quality. We finally analyze the voltage breakdown limits for safe operation and ESQ spacing considerations to prevent field overlap between the doublet ESQs.

5.4.1 Numerical Methods

We seek to minimize the leading order error term in the multipole expansion of the electric field by choosing the optimized geometrical configurations for the ESQ arrays. Fig. 5.5 shows a face-on view of a prototypical ESQ design comprised of cylindrical rods. These rods have radius R , axial length ℓ_q , and applied voltage V_q . The beam passes through a circular clear-bore aperture of radius r_p (the rods touch the clear aperture along the coordinate axes) shown as a dotted circle in the figure.

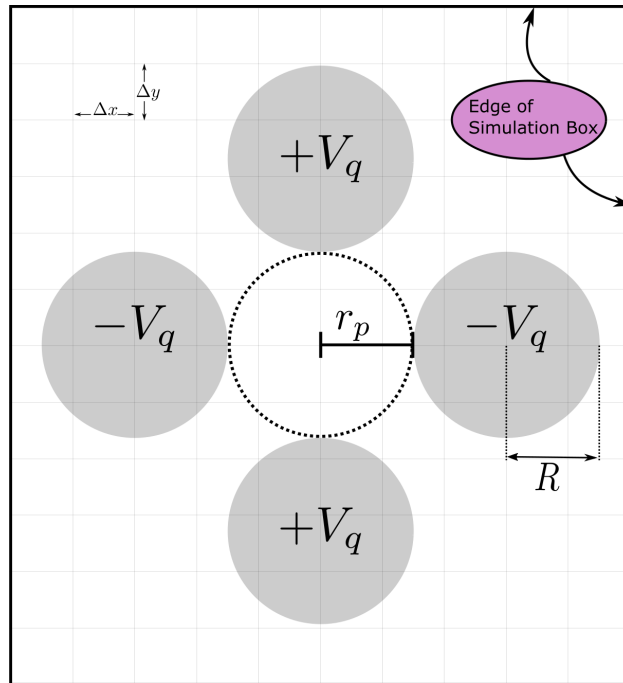


Figure 5.5: Transverse view of a single electrostatic quadrupole in a simulation box with grid sizes $\Delta x = \Delta y$. The conducting rods are shaded in grey with the applied voltage V_q . The rods have a radius of R and the dashed line shows the clear bore aperture of radius r_p . White space represents vacuum.

The short axial length ℓ_q of the ESQ arrays (formed by thin PCB wafers), relative to r_p , gives rise to a significant axial fringe field. We used Warp to model an ESQ made of solid cylinders with $r_p = 0.55$ mm, $\ell_q = 1.264r_p$, and rod radius $R = 1.14r_p$ to extract the 3D electric field $\vec{E}(x, y, z)$ on the simulation mesh. The electric field gradient $G(z) = \partial E_x(z)/\partial x|_{x=y=0}$ is calculated one grid cell off-axis to minimize nonlinear field content. Fig. 5.6 shows the electric field gradient obtained. Denoting the max gradient at the geometric center of the quadrupole as $G^* = G(z = 0)$, we define the effective length ℓ_{eff} as [13]:

$$\int_{z_1}^{z_2} G dz = G^* \ell_{\text{eff}}. \quad (5.2)$$

Here, z_1 and z_2 are sufficiently large to capture the axial extent of the fringe field upstream and downstream of the ESQ. As illustrated in Fig. 5.6, this system has $\ell_{\text{eff}} = 1.306$ mm, an 88% increase from ℓ_q . The result $\ell_{\text{eff}} > \ell_q$ shows that this design will provide more focusing than might be expected based on physical axial length ℓ_q . This helps increase the strength of the short focusing elements. We parametrically evaluated how the physical length of the quadrupole relative to the aperture radius impacts the effective length (see Fig. 5.7). For $\ell_q/r_p < 3$, ℓ_{eff} can be $\sim 40\%$ or more greater than ℓ_q . For $\ell_q/r_p > 4$, ℓ_{eff} approaches ℓ_q . This result provides guidance on the trade offs between creating a more compact system by shortening ℓ_q and the amount of focusing that can be achieved.

The ESQ arrays are optimized to minimize multipole field errors. We first integrate the 3D electric field $\vec{E}(\vec{x})$ over the axial coordinate z and normalized by ℓ_{eff} [30] to obtain the

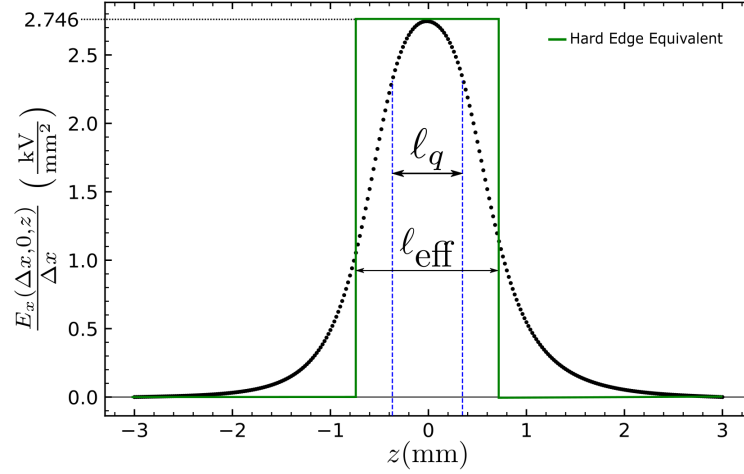


Figure 5.6: Electric field gradient calculated one grid-cell off-axis for a single isolated ESQ made of solid cylindrical rods. The blue vertical dashed line demarcates the physical length of the ESQ ℓ_q . The green line shows the hard-edge equivalent with a total field extent equal to the calculated effective length ℓ_{eff} at the mid-ESQ gradient.

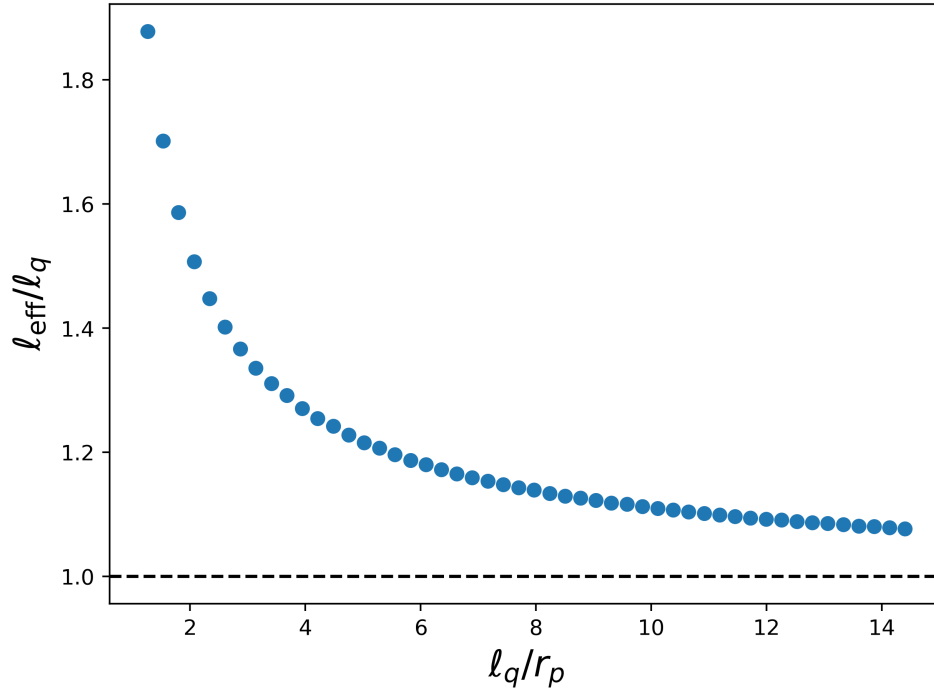


Figure 5.7: Ratio of physical ESQ length ℓ_q to calculated effective length ℓ_{eff} as a function of axial length normalized by the clear aperture radius ℓ_q/r_p .

averaged 2D transverse field components \bar{E}_x and \bar{E}_y

$$\begin{aligned}\bar{E}_x &= \frac{1}{\ell_{\text{eff}}} \int_{z_1}^{z_2} E_x(x, y, z) dz, \\ \bar{E}_y &= \frac{1}{\ell_{\text{eff}}} \int_{z_1}^{z_2} E_y(x, y, z) dz.\end{aligned}\quad (5.3)$$

Again, z_1 and z_2 are again sufficiently large to capture axial extent of the field upstream and downstream of the ESQ field. For cases of ESQ pairs in doublets, one of the the coordinates can be the zero field point between the opposite biased ESQ pairs. The Maxwell Equations of the this 2D-field are $\nabla \cdot \bar{\vec{E}} = 0$ and $\nabla \times \bar{\vec{E}} = 0$ where $\bar{\vec{E}} = \bar{E}_x \hat{x} + \bar{E}_y \hat{y}$. This fields satisfies the Cauchy-Riemann conditions for a complex field $\underline{\bar{E}}^* = \bar{E}_x(x, y) - i\bar{E}_y(x, y)$ with $i = \sqrt{-1}$ and complex coordinate $\underline{z} = x + iy$. Thus, expanding the field in a Laurent series we have:

$$\underline{\bar{E}}^*(x, y) = \bar{E}_x(x, y) - i\bar{E}_y(x, y) = \sum_{n=1}^{\infty} \underline{b}_n \left(\frac{\underline{z}}{r_p} \right)^{n-1}. \quad (5.4)$$

Here, \underline{b}_n is a complex number representing the multipole moments scaled such that the \underline{b}_n coefficients are measured in field units. In a polar representation with $\underline{z} = r e^{i\theta}$, Eq. (5.4) is expressed as

$$\underline{\bar{E}}^*(x, y) = \bar{E}_x(x, y) - i\bar{E}_y(x, y) = \sum_{n=1}^{\infty} \underline{b}_n \left(\frac{r}{r_p} \right)^{n-1} e^{[i(n-1)\theta]}. \quad (5.5)$$

Expressing $\underline{b}_n = A_n - iB_n$, orthogonality can be exploited to calculate:

$$\begin{aligned}A_n &= \frac{1}{2\pi} \left(\frac{r_p}{r} \right)^{n-1} \int_0^{2\pi} \left[\bar{E}_x \cos[(n-1)\theta] - \bar{E}_y \sin[(n-1)\theta] \right] d\theta \\ B_n &= \frac{1}{2\pi} \left(\frac{r_p}{r} \right)^{n-1} \int_0^{2\pi} \left[\bar{E}_x \sin[(n-1)\theta] + \bar{E}_y \cos[(n-1)\theta] \right] d\theta.\end{aligned}\quad (5.6)$$

To evaluate the coefficients in Eq. (5.6), the Warp fields on a 3D Cartesian rectangular parallel-piped mesh were axially integrated using Eq. 5.3 to obtain 2D fields on a transverse square mesh. Field values at mesh points are linearly interpolated to desired coordinates using volume weighting [31]. The multipole coefficients are then calculated using Eq. 5.6. Enough azimuthal mesh points (N_θ) are taken along a circular interpolation path to resolve the maximum order n_{\max} of the extracted multipole by taking

$$N_\theta = \frac{\sqrt{2}\pi n_{\max} r}{\Delta x}. \quad (5.7)$$

Here Δx is the grid spacing in x , $\Delta y = \Delta x$, and r is the integration radius. We chose r to be approximately one grid cell within r_p since the multipole error terms will be strongest in regions close to the ESQ surfaces (see Fig. 5.8). Symmetry considerations show that only $n = 2$ (Quadrupole), and $n = 2 + 4p$ for $p = 1, 2, 3, \dots$ are allowed [32]. Therefore, the leading

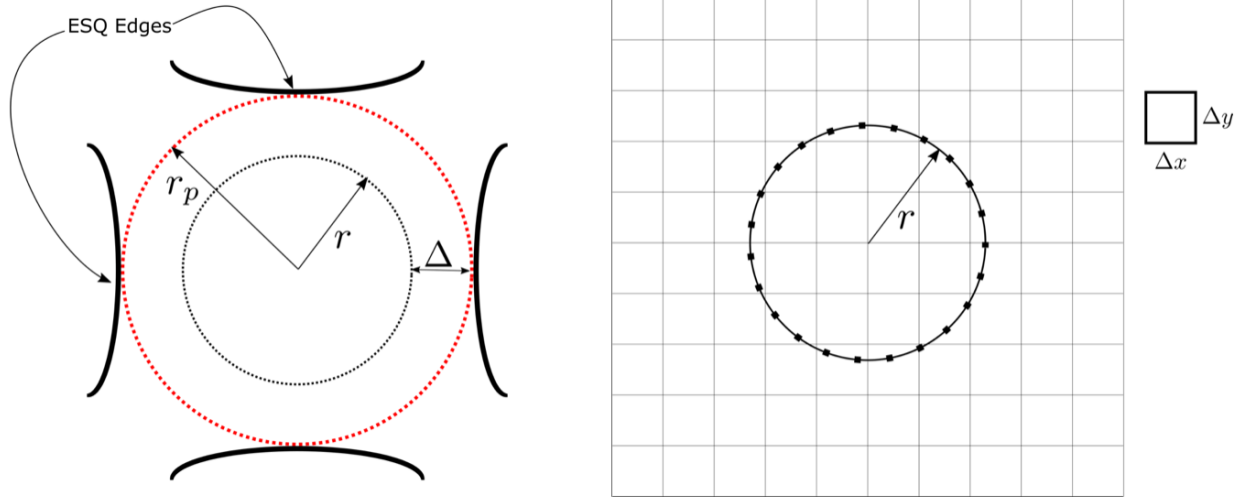


Figure 5.8: (Left) The clear bore aperture is displayed as the dotted red circle with radius r_p . The integration path is shown as a dotted black line at radius r . The radial distance between r_p and r is Δ which is chosen to be grid cell away from the ESQ edges shown in solid black lines. (Right) Schematic of the simulation mesh with field interpolation points passing through the discretized space with simulation cell box sizes of Δx and Δy .

order error term is $n = 6$ or 12-pole terms. By symmetry, $B_6 = 0$ and A_6 will generally be nonzero. Simulations are consistent with these symmetry considerations to within expected numerical errors. We used a cubic mesh with grid spacings $dx = dy = dz = 10\mu\text{m}$. The transverse mesh size was 3 mm (size of the unit cell). The axial size of the mesh is ≈ 6 mm which gives ≈ 3 mm of spacing from the ESQ axial edge and the end mesh boundary on either side of the ESQ. Using Eq. 5.7, this gives $N_\theta \approx 1460$.

5.4.2 Optimizing ESQ Arrays to Minimize Leading Order Error in Field Expansion

5.4.2.1 Benchmarking for a Single Axially Long ESQ

To simulate a long quadrupole ($\ell_q \gg r_p$), the longitudinal mesh boundaries were set to periodic and placed on the axial rod extents to model an infinitely extended quadrupole with pure 2D transverse fields. Meanwhile, the transverse boundary conditions were set to Dirichlet with the simulation box large enough to contain the rods. To resolve the mesh elements and provide enough sample points to accurately interpolate points along the integration path at radius r , the resolution on the mesh was fixed at $5\mu\text{m}$. To analyze the correlation between the rod radius to aperture radius ratio (R/r_p) and the leading order multipole error term (A_6), we calculated the normalized error, $A_6/|A_2|$, for several rod radii while keeping the aperture radius fixed at $r_p = 0.55$ mm. Referencing Fig. 5.9, it was determined that A_6 is zeroed when $R/r_p = 1.145$. The result for a single isolated long quadrupole is within 0.1% of the axially long ESQ result previously found by Faltens and verifies the algorithm [1].

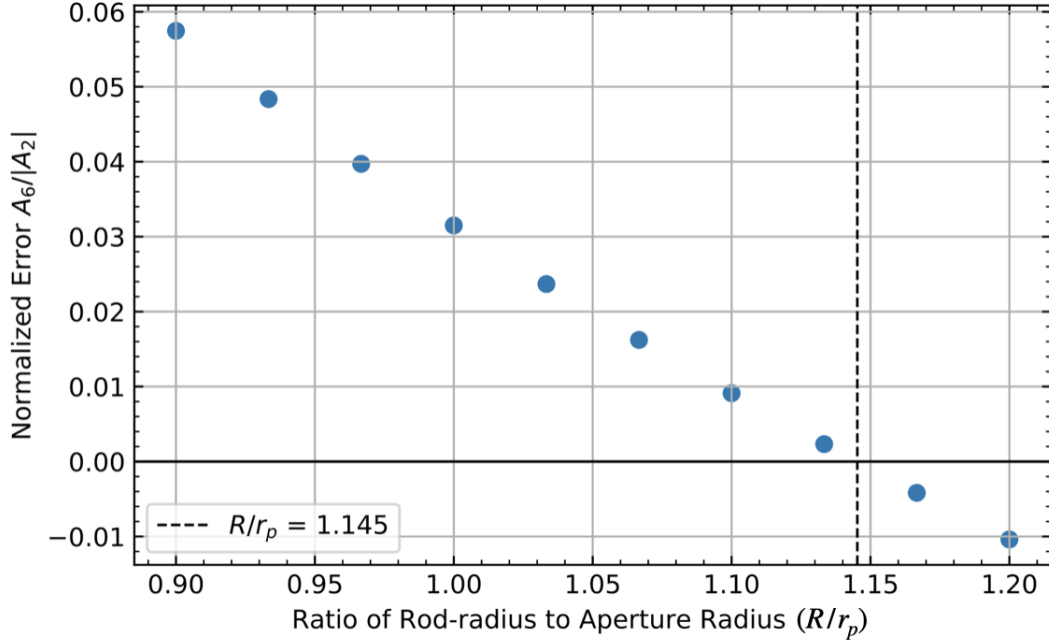


Figure 5.9: Scaled leading order multipole error field term A_6 as a function of rod radius R/r_p for an infinitely long quadrupole. The vertical dashed line is at $R/r_p = 1.145$ where the error is nulled.

5.4.2.2 Finite Axial Length ESQ Optimization for Full and Chopped Rod Geometries

Fig. 5.10 shows three sets of curves for a finite length quadrupole with $\ell_q = 0.695$ mm, $r_p = 0.55$ mm, and three different rod geometries. The blue curve with circle markers corresponds to a geometry using full rods, the green curve with triangle markers is for half-chopped rods, and the third magenta curve with “x” markers is a quarter-chopped rod. All three curves were created with Dirichlet boundary conditions for both transverse and longitudinal directions with boundaries placed sufficiently far away not to significantly impact the results shown. The curves for the full and half-chopped rods approximately share the same optimum at $R/r_p = 1.318$. However, the quarter-chopped rod has a new, modestly shifted, optimum at $R/r_p = 1.330$, a one percent difference. From this, we deduce that the curvature of the

rod along with the finite axial length are primary factors that shift from the infinitely long ESQ optimum $R/r_p = 1.145$ discussed in Sec. 5.4.2.1. This suggests one can create more transversely compact focusing arrays by using chopped rods while maintaining the same curvature and varying the rod radius to minimize field errors. However, caution must be applied because the transverse chop locations may give strong field enhancement at edges and contribute to voltage breakdown at a lower voltage (see Sec. 5.4.2.3).

Lastly, we use periodic transverse boundary conditions instead of Dirichlet boundary conditions and used the simulation box to chop the rods in half to represent an infinite transverse array. The optimum rod radius to null the leading order error field term for this setup was found to be $R/r_p = 1.336$, a less than 0.5% difference from the value found with Dirichlet transverse boundary conditions. This suggests that the periodic array geometry does not significantly shift the optimum point to minimize leading order field errors. Note also that half-chopped rods in a periodic array also eliminate transverse edges that may impact voltage holding.

5.4.2.3 Voltage Breakdown Limits for ESQ Array Geometry

Required focusing strength provided by the ESQs will necessitate higher bias voltages and/or longer axial lengths to compensate for increased kinetic energy of the ions as they advance through the acceleration lattice. As kinetic energy increases, the lattice period also increases allowing for axially longer ESQs (lengths can be increased by stacking PCB wafers) to help with voltage holding. For distances less than 1 cm, breakdown in typical low energy transport conditions is expected at electric field strengths of 10^7 V/m [1]. The max 2D electric field

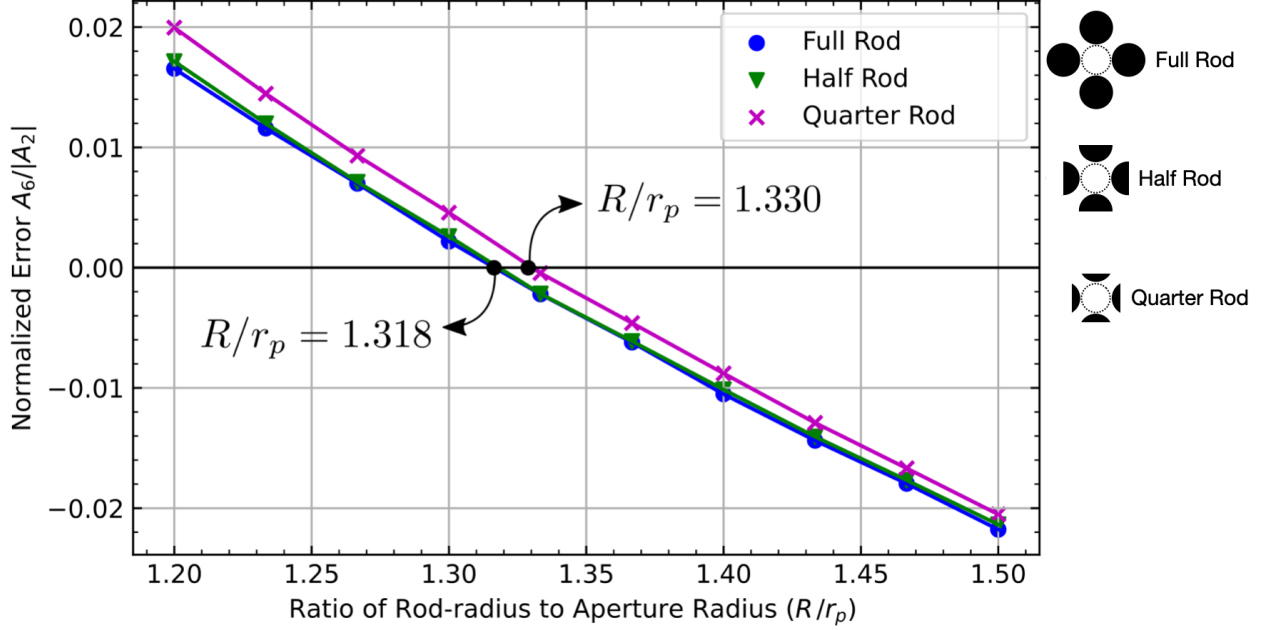


Figure 5.10: Leading error term normalized by the quadrupole term A_6/A_2 in the multipole field expansion for a finite length ESQ of $\ell_q = 0.695$ mm using full (blue circles), half-chopped (green triangles), and quarter-chopped (magenta “x”s) rod geometries. To the right are illustration of the chopped rod geometries.

for an ideal quadrupole occurs at $R = r_p$ and is given by

$$E_{2D} = \frac{2V_q}{r_p}. \quad (5.8)$$

Substituting the breakdown field limit of 10^7 V/m and solving for the quadrupole voltage V_q gives a maximum limit of $V_{q,br} = 2.75$ kV. Typically, fields are limited to less than half the breakdown value to provide safety margins giving a practical limit of $V_{q,\frac{1}{2}br} = 1.37$ kV. To evaluate the impact of transverse and longitudinal rod chopping, we used Warp to perform a voltage scan of an isolated ESQ with $\ell_q = 0.695$ mm and $r_p = 0.55$ mm using periodic transverse boundary conditions on a 3D grid. The simulations calculated the electric field for a mesh with $10 \mu\text{m}$ grid spacing in the transverse direction and $20 \mu\text{m}$ grid spacing in the longitudinal direction. The voltage scan was performed over a range of voltages from

0.1 kV to 1 kV and the magnitude of the 3D electric field $|E(x, y, z)|$ was computed at each point on the mesh. The half breakdown limit occurred at an applied voltage of 531 kV, a 61% reduction from the 2D field estimates above. This breakdown value occurred at the longitudinal grid point at the axial edge of the ESQ and transversely at the grid point closest to the oppositely biased rod.

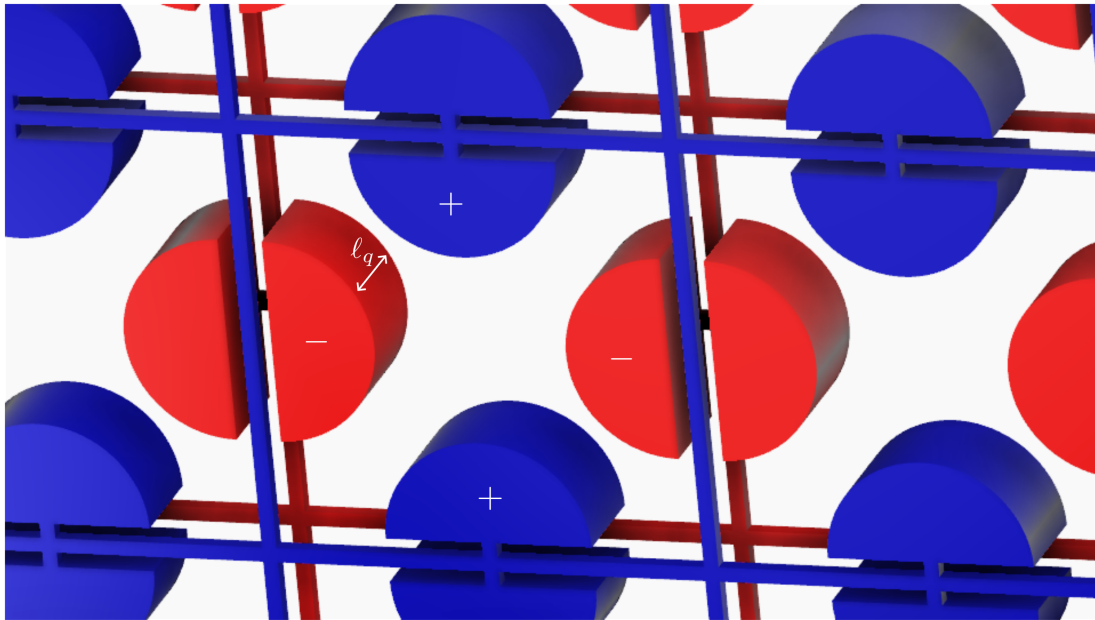
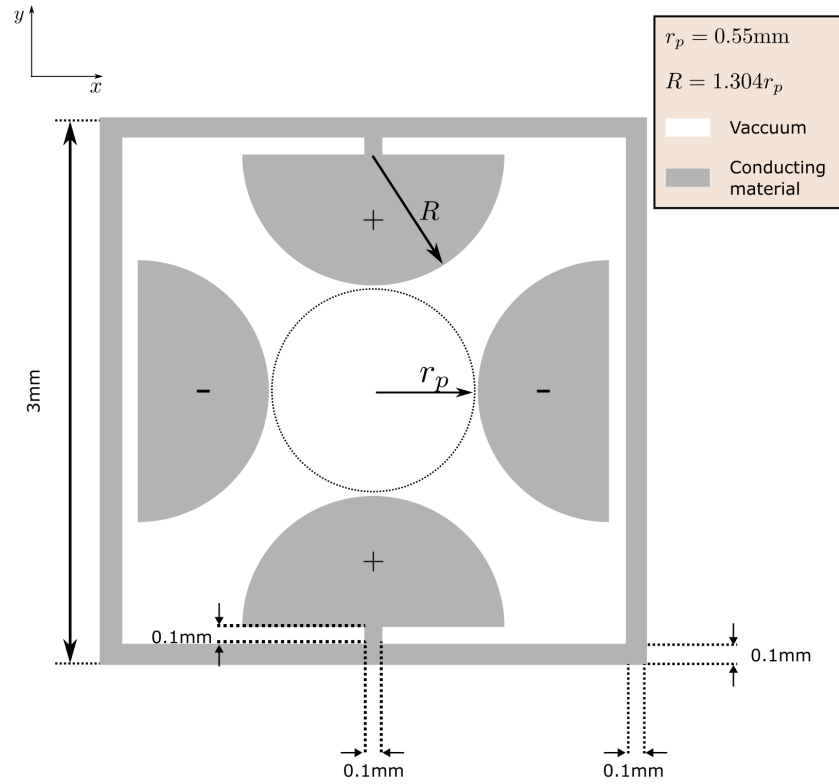


Figure 5.11: (Top) Schematic of a single cell from the realistic transverse-periodic array ESQ design used in simulations. The clear bore aperture is shown as the dotted black circle with radius r_p . The half-chopped quadrupole rods have radius R . Conducting objects are shaded grey and white-space is vacuum. (Bottom) Transversely periodic ESQ array modeled in simulation. Positively biased structures are colored blue and negatively biased structures are red.

5.4.2.4 Optimization Considerations From Realistic Geometry

The manufactured geometry varies from the idealized geometry in both the transverse and longitudinal directions as can be seen in Fig. 5.11. Transversely, there is a surrounding conducting square with prong attachments to the rods. This structure is printed on each longitudinal side of a single PCB wafer. On one side, the conducting structures are positively biased while on the other end, the prong attachments are switched to the other rods with all conductors being negatively biased. Note that the surrounding material is plastic. Because Warp does not have the capability to simulate dielectric material in the full 3D solver, we take the surrounding material to be vacuum. This will tend to overestimate fields in the regions where there is dielectric. The conducting square that surrounds the rods and the attaching prongs place a physical limit on the size of the poles. This limit ensures that the biased rods do not touch the surrounding conducting square and that the rods do not cross into the aperture space. We take the half-chopped rod geometry as our design to create a more compact structure. With these physical constraints, the physical rod radius must satisfy $R_{\max} \leq 0.75 \text{ mm} = 1.36r_p$. Using the same analysis as for the ideal geometry and a half-chopped structure with $r_p = 0.55 \text{ mm}$ and $\ell_q = 0.695 \text{ mm}$ an optimum ratio of $R/r_p = 1.304$ was found (see Fig. 5.12).

Following the same procedure as before for the idealized geometry, the half breakdown-limit was found to be at $V_{q,\text{br}} = 0.531 \text{ kV}$. This places the safe operating limit at a much lower voltage of $V_{q,\frac{1}{2}\text{br}} = 0.267 \text{ kV}$. Before, the breakdown point occurred at the longitudinal chopped ends of the ESQs. In the realistic geometry, the point shifts to the space between the oppositely charge rod and surrounding square at the prong tips. Due to the aforementioned inability to simulate the dielectric material in Warp's full 3D solver, the

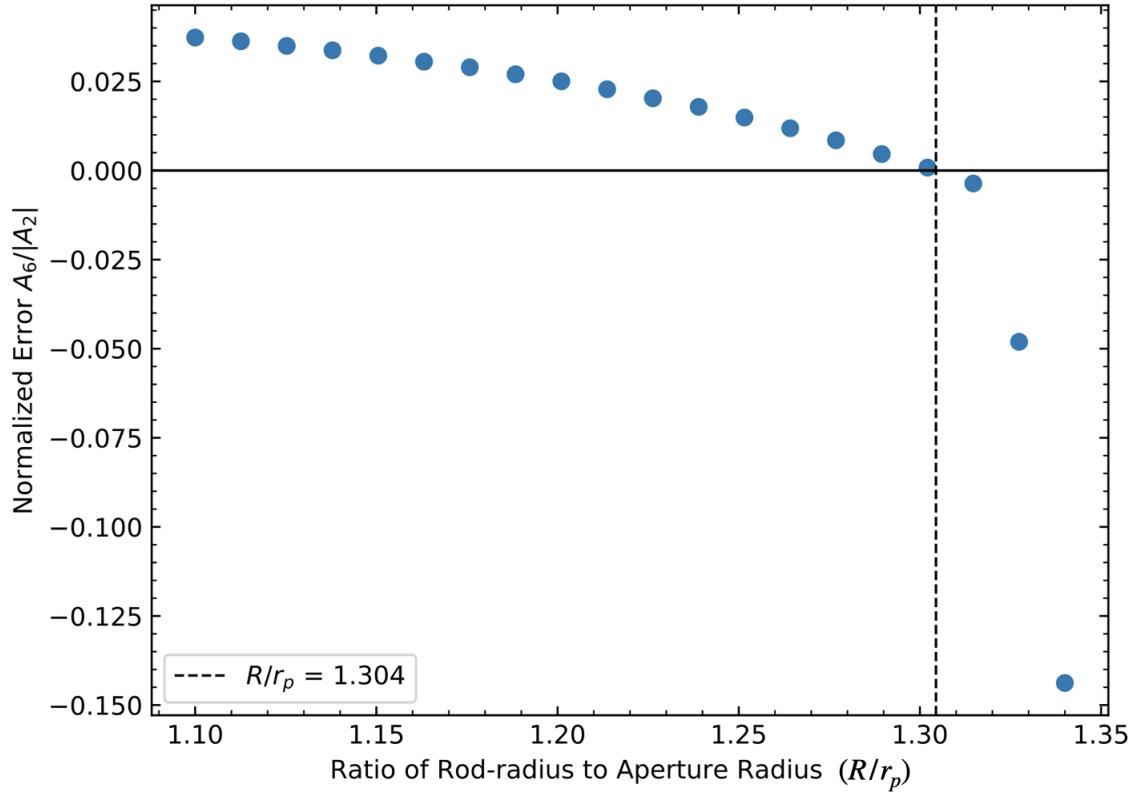


Figure 5.12: Leading multipole error field term as a function of the rod radius R/r_p for an RF MEMS ESQ design.

operating limit is likely higher since this area is surrounded by dielectric and breakdown will likely occur at the axial rod ends with similar limits as obtained in the ideal geometry at this location. For example, the dielectric material used in the PCBs is FR-4. This material has a dielectric constant ranging from 3.8 to 4.8. Thus, assuming the electric field scales linearly with the dielectric constant, we can assume the safety limit to occur within the range of $0.7 \text{ keV} \leq V_q \leq 0.9 \text{ keV}$. This safety limit defines the operating point at 50% the breakdown; therefore, this limit can be pushed higher with caution. Since it is unclear which breakdown limit we should use but we know the dielectric material will increase the limit, we will take the breakdown limit found in vacuum to be the half-breakdown limit from here on: that is, $V_{q, \frac{1}{2}\text{br}} = 0.531 \text{ kV}$.

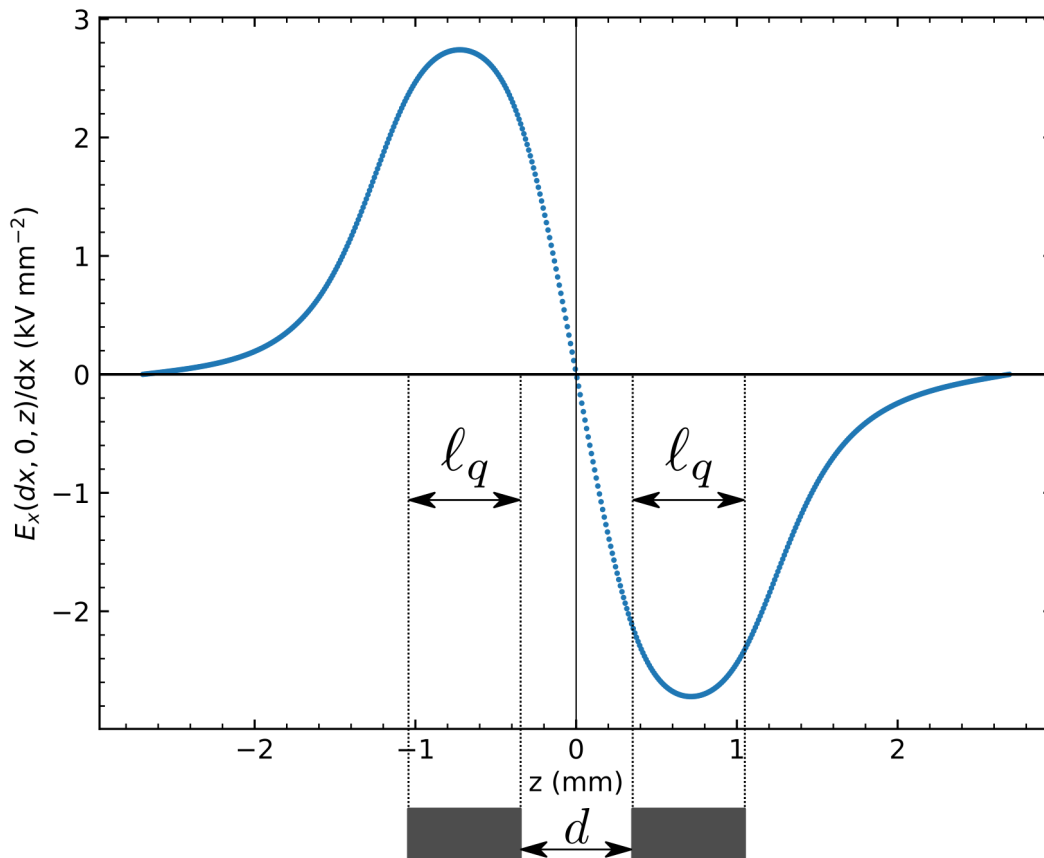


Figure 5.13: Gradient of ESQ Doublet structure. The axial extent of the ESQ elements are drawn below the plot (length ℓ_q), and axial ESQ separation distance is d .

In addition to the geometrical factors for ESQ optimization, the acceleration lattice will be comprised of a doublet with oppositely biased neighboring ESQs. The gradient of the two ESQs that are oppositely biased will be zero at the symmetry point located axially between the geometric center of the two ESQs. Due to the close separation and opposite bias, the axial separation distance d will have an impact on the effective length and therefore the overall focusing strength obtained. This is illustrated in Fig. 5.13 with $r_p = 0.55$ mm and $\ell_q = 0.695$ mm for two ESQs separated by $d = 0.695$ mm. To analyze the effects of d in the doublet structure, we performed simulations by placing an isolated doublet in a simulation box with periodic boundary conditions in the transverse direction and Dirichlet

boundary conditions in the longitudinal direction. The upstream quadrupole was held fixed at a sufficient distance from the upstream grounded plane to ensure that the grounded plane had no significant impact on the field. The downstream quadrupole was then placed at varying d and the effective length for the upstream quadrupole was calculated by integrating the gradient from the upstream grid extent to the downstream zero point and applying Eq. 5.2 with G^* corresponding to the max field in the upstream ESQ. Results obtained are shown in Fig. 5.14. In the figure, we plot $\Delta\ell_{\text{eff}} = \ell_{\text{eff}}|_{\text{doublet}} - \ell_{\text{eff}}|_{\text{isolate}}$ normalized by the $\ell_{\text{eff}}|_{\text{isolate}}$ using the effective length of the isolated ESQ and varying d between the two ESQ extents. Results show that close ESQ spacing with $d \lesssim 2.5$ mm can significantly degrade effective focusing strength. This result provides guidance on spacing requirements for the ESQ doublet system both between the ESQ pairs and the ESQ and neighboring grounded acceleration gap wafer. Generally, one wants the neighboring ESQ doublets to have maximum d separation to focus more efficiently. This will also help increase effective length. Both effects will help control ESQ biases needed. Grounded planes of the acceleration gaps closest to the upstream and downstream ESQs should also have sufficient spacing to avoid enhancing peak fields.

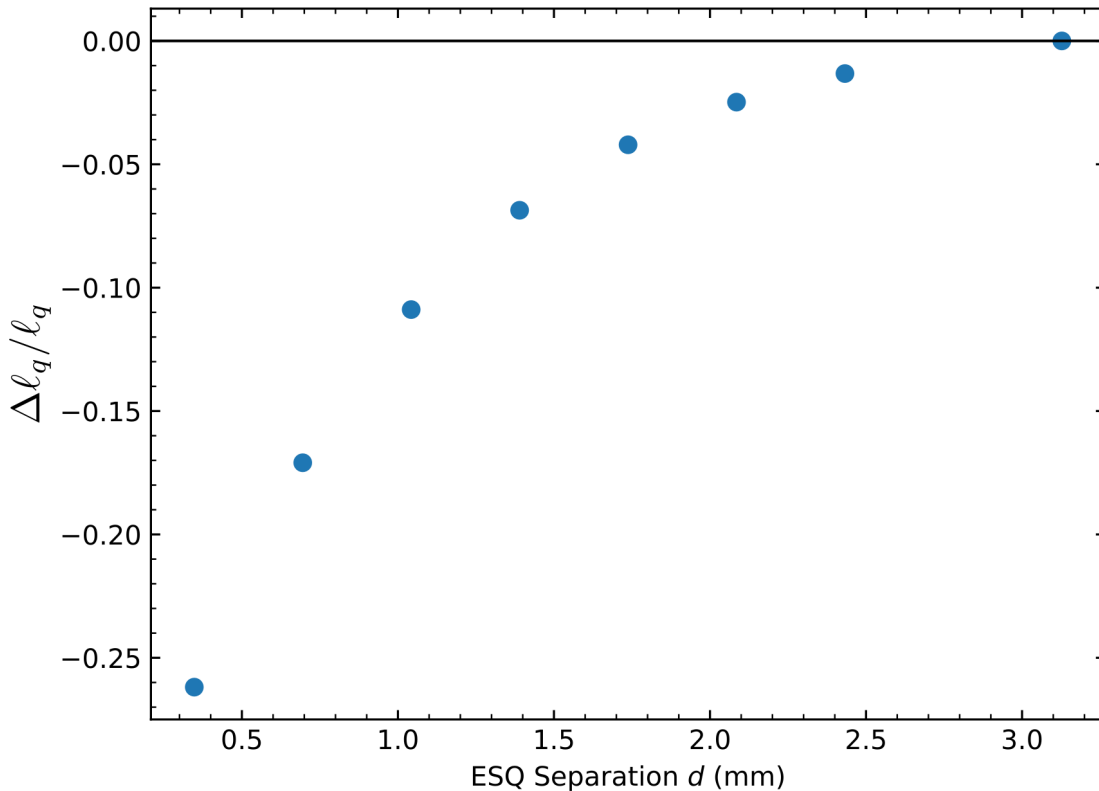


Figure 5.14: Difference in the calculated effective length ℓ_q as the ESQ doublet separation distance d is varied.

5.5 Summary and Concluding Remarks

The ESQs in our system will provide transverse focusing to limit particle losses and increase the number of particles delivered on target. Creating a pure linear focusing quadrupole field is not possible and there will be contributions to the field from high order error field terms. We developed a computational tool to simulate ESQ geometries and optimize geometric choices to minimize the leading order error field term A_6 in the quadrupole field expansion. We found that for $r_p = 0.55$ mm and $\ell_q = 0.695$ mm that using a half-chopped rod of radius $R = 1.304r_p$ is the optimal choice. We also found that the effective length of these short quadrupoles are considerably longer than their physical lengths. Our tool also enabled

us to predict where voltage breakdown is likely to occur and at what voltage. Due to the inability of Warp to simulate dielectrics in 3D, the breakdown limit estimated is conservative and significantly higher ESQ biases are possible when the FR-4 dielectric is present. We therefore will take the breakdown limit found in vacuum to be the half-breakdown limit so $V_{q, \frac{1}{2}\text{br}} = 0.531$ kV. Lastly, we were able to characterize the effects of axial separation in the doublet structure for our ESQ design has on the field. We showed that for an axial separation of $d \geq 2.5$ mm the fields from the opposing doublets do not interfere significantly with each other to reduce focusing strength. The developed tool is robust and can be applied to new geometries with the limitation being the ability to accurately create the conductor in Warp. We use the optimized ESQ array fields in subsequent reduced physics transverse models described in Ch. 7. Results found in this chapter will allow refined ESQ focusing arrays to be constructed in future RF-MEMS experiments.

Chapter 6. Longitudinal Dynamics in RF-MEMS

In Ch. 2 the longitudinal dynamics in systems were reviewed. In the RF-MEMS system, the difficulty arises due to the strong acceleration because $qV_g \approx \mathcal{E}_b$ out the source which prevents a perturbative treatment. The low beam energy emerging from the source helps accelerate as rapidly as possible to keep the system compact. With non perturbative acceleration, defining an RF bucket is challenging, and the \mathcal{E}_b-t phase space evolution is better analyzed numerically. Using a self-consistent 3D simulation like Warp is possible but impractical. Varying the synchronous phase condition to balance longitudinal focusing and acceleration can be done by varying the gap-to-gap distances in the lattice to change the synchronous phase ϕ_s . Using a fixed set of injection conditions, this limits the parameter tuning to one variable. However, the synchronous phase is a continuous variable and for simultaneous acceleration and longitudinal focusing needs to be in the range $-\pi/2 < \phi_s < 0$ and can be tuned for each gap. This brute force approach to explore optimal tuning approaches infeasibility with higher energies. Furthermore, the complexities captured in the Warp simulation (space charge dynamics, transverse dynamics, particle scraping) make analysis of the longitudinal dynamics difficult to disentangle from other effects. For these reasons, a rapid running reduced physics model is needed. The model should be able to rapidly run for any number of gaps and designs. It should also be comparable to Warp when making a comparison for a test case. In this chapter, the development and application of such a model is presented.

6.1 Numerical Methods

In Sec. 5.3, the extraction of the gap array fields from Warp field solves was presented. To simulate particles moving RF-varying fields, the normalized RF-fringe field is extracted. Since the gap geometry is constant throughout the lattice, this extraction only needs to be done once and then the time dependent RF field in each gap can be found by scaling. However, as noted in the earlier section, care must be taken to include two gaps in the Warp simulation or else the periodic array geometry will not be properly incorporated into the field-solve if done incorrectly. This will give a strongly altered field and invalidate the results in the simulation. With the extracted fringe data, the position and field array can be “stitched” into a complete lattice by translations and scaling. Fig. 6.1 is an illustration to help explain the process for stitching in the extracted z -data. Initially, the complete z -array is created with some specified resolution dz and n -points. The gap center is located in this initial array between z^k and z^{k+1} . Here, the extracted p -points of z -data (\mathbf{z}_{ext}) is incorporated into the lattice. In general, $p \neq n$ and $dz \neq dz_{\text{ext}}$. A completed z -mesh will have $n + m$ grid points. An identical algorithm is performed in tandem to incorporate the field data $E_{z,\text{ext}}$ between E_z^k and E_z^{k+1} . This process is repeated for the specified number of gaps and the normalized $E_{z,\text{ext}}$ is scaled to represent the acceleration voltage with the scaling factors discussed in Sec. 5.3 so that the on-axis field is properly connected to the gap biases.

With the completed position and field data, a reduced physics simulation for transporting particles down the lattice with RF-varying fields can be formulated. To explain the algorithm, consider a single particle at the $(j - 1)$ -th point in the advancement. The particle is of mass m at current a position z^{j-1} , time t^{j-1} , and kinetic energy \mathcal{E}^{j-1} . To advance the particle

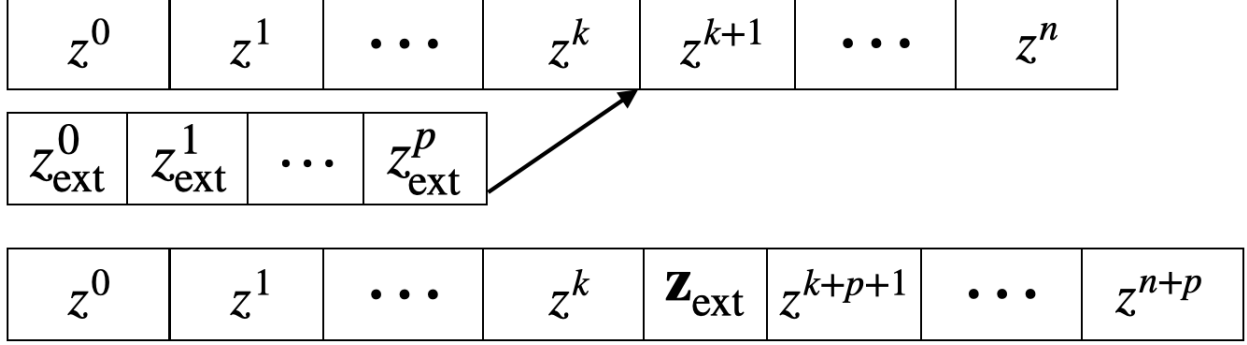


Figure 6.1: Visual representation of the array-stitching process for the position data in the reduced physics longitudinal model. An initial z -array is created with n -grid points. The extracted gap data has both field data and position data. Here, the position data \mathbf{z}_{ext} is stitched into the initial array between z^k and z^{k+1} . The resultant mesh will have $n + p$ grid points. An identical algorithm is performed in tandem for the field array and extracted field data.

to z^j , the particle velocity v^{j-1} must be calculated. Then, the velocity is used to find the time it takes to travel $\Delta z^j = z^j - z^{j-1}$. Note that this distance should be calculated each step since it is not guaranteed that the mesh is uniformly discretized. This time is then used to compute the future electric field $E_z(z^j, t^j)$ and calculate the work done on the particle by the field. Symbolically, the non relativistic algorithm is:

$$\begin{aligned}
 v_z^{j-1} &= c \sqrt{\frac{2\mathcal{E}^{j-1}}{mc^2}}, \\
 t^j &= t^{j-1} + \frac{\Delta z^j}{v_z^{j-1}}, \\
 \mathcal{E}^j &= \mathcal{E}^{j-1} + q\Delta z^j E_z(z^j, t^j)
 \end{aligned} \tag{6.1}$$

where c is the speed of light. The electric field at the j -th grid point at t^j is given by $E_z(z^j, t^j) = E_z^j \cos(\omega t^j)$ with $\omega = 2\pi f_{\text{rf}}$ and E_z^j is the j -th element of the field data. It is possible for particles to get decelerated and acquire negative velocity in the simulation. When this occurs, these particles are “scraped,” meaning that a filter is applied so that

particles with $v \leq 0$ (or negative kinetic energy) are ignored. Particles are loaded by initially distributing about the design particle which is always placed at $z = 0$. Thus, particles will be between $|z_{\text{load}}/2|$ where z_{load} is some distance set by the user. Then, the time for each particle to reach $z = 0$ is calculated and saved. This initializes all particles to $z = 0$ with an initial time. Particles with an initial negative time are “early” relative to the design particle while particles with initial positive time are “late.”

6.2 Benchmark of the 1D Model with Warp in a 4-Gap System

To verify that the 1D code properly describes the longitudinal dynamics of the RF-MEMS system, a simulation of the 1D code was benchmarked against Warp simulations. A four gap system with $V_g = 7 \text{ kV}$ was modeled for maximum acceleration with $\phi_s = 0$ at all gaps. A continuous and uniformly distributed beam was injected with initial kinetic energy $\mathcal{E}_{b,i} = 7 \text{ keV}$. Fig. 6.2 shows a comparison in the final energy distribution between the Warp simulation in orange and the 1D code in blue for a selection of particles satisfying $|\Delta t| = |t - t_s| \leq 0.5\tau_{\text{rf}}$, where τ_{rf} is the RF period. If this filtering is not done, then the comparisons can be misleading. This is because when Warp advances particle in time, particles will be lost for various reasons: absorbed at transverse boundary, decelerated to negative velocity, and too slow to reach the target before the simulation is terminated. In the 1D code, this is not an issue since particles are advanced in space and thus, all particles that maintained a forward velocity will be collected. Despite Warp simulating space charge effects and advancing particles in a 3D space, the 1D model reliably captures the energy distribution since longitudinal space charge effects appear minimal. This 1D longitudinal code

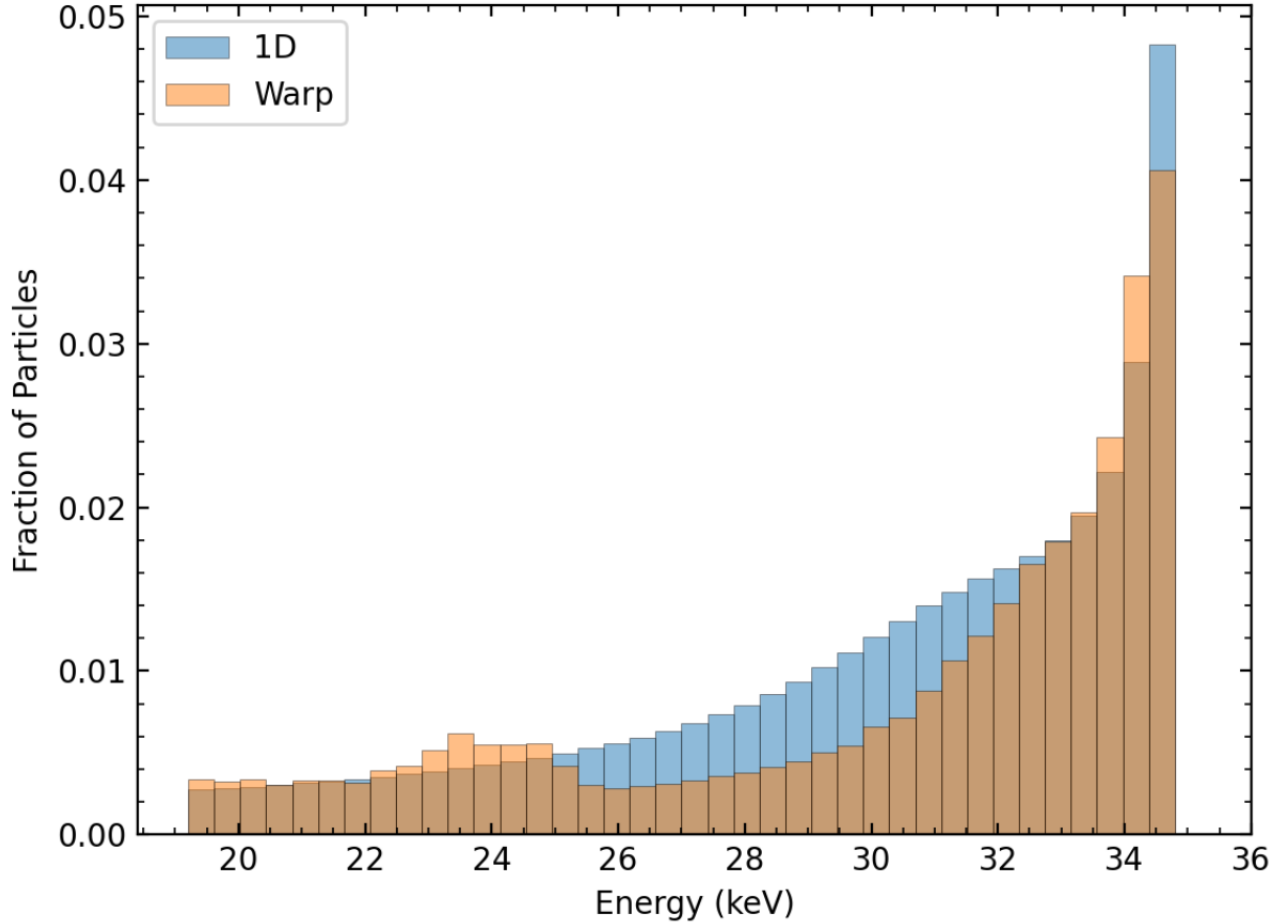


Figure 6.2: Comparison of a four-gap simulation with $V_g = 7$ kV and $\mathcal{E}_{b,i} = 7$ keV between the reduced physics 1D longitudinal code (blue) and 3D Warp simulations (orange). The binning is done over the final kinetic energies of the particles after arriving at a target after the fourth gap. Particle energies are filtered based on arrival time relative to the synchronous particle with $|\Delta t| = |t - t_s| \leq 0.5\tau_{rf}$.

runs rapidly in comparison with Warp: the 1D code takes less than a second, whereas the Warp simulation can take up to ten minutes for a low-resolution (coarse spatial discretization and large time step) runs.

6.3 Lattice Design for Optimal Acceleration, Bunching, or Both

With the reduced physics 1D longitudinal model validated, the parameter space can be explored to find optimal lattice designs. The definition of optimal may change based on the application. In most cases users want the most particles delivered at the highest kinetic energy. Especially for linacs, users want the shortest lattice possible necessitating high acceleration rates. The two criteria for high kinetic energy gain and longitudinal focusing for more particles conflict with one another—maximum acceleration gives least particle bucket capture and longitudinal focusing, and vice versa. As a demonstration of the algorithm, we will compare three different lattice designs: one for maximum acceleration, one for maximum bunching, and one that tapers the acceleration so that there is simultaneous particle capture and bunching while maintaining acceleration. This will be done for a four gap system to keep the analysis easier.

6.3.1 Initial Conditions and Parameter Settings

A cold (iso energetic) continuous beam of uniformly distributed Ar^+ ions emerging from the source will be simulated with $\mathcal{E}_b = 7 \text{ keV}$. The injection will be for one RF-period giving a beam pulse duration of τ_{rf} and axial length $L_b = v_b \tau_{\text{rf}}$. At the target, only particles with $|\Delta t| \leq 0.4 \tau_{\text{rf}}$ will be analyzed. Note, by tightening the range on Δt , we have made it more difficult to capture particles in the end diagnostic. This drives design optimization to increased particle capture when tuning parameters. We use an applied gap voltage $V_g = 6 \text{ kV}$ with the first gap starting at a maximally negative field. The voltage drop is a design choice that will prevent possible damage to the source. At extraction, such a strong acceleration

will stagnate and decelerate particles which would impinge on the source. Fig. 6.3 shows the extracted $E_z(r = 0, z)$ at $t = 0$ for a $N_g = 4$ where the fields are normalized by the calculated DC-field $E_{DC} = V_g/g$. Note that the field amplitude is greater than unity. When applying V_g to the conductors, the on-axis potential for the real conductor is less than this value. In order to ensure the desired on-axis voltage, the applied gap voltage must be slightly greater to compensate for finite 0.55 mm aperture radius. Because the fields shown are extracted from the real conductors, these fields are scaled to achieve the correct $E_z(r = 0, z)$. If this scaling is not done, then the RF-matching will be slightly mismatched for the design particle and desynchronization will grow gap-to-gap. Gap centers are marked by vertical gray dashed lines. Note the increase in spacing to maintain phase synchronism as the beam energy increases consistent with π -mode acceleration. Particles are advanced from $z = 0$ to a Faraday cup (Fcup) marked by a vertical red dashed line that is placed 10 mm from the grounded edge plate of the final gap.

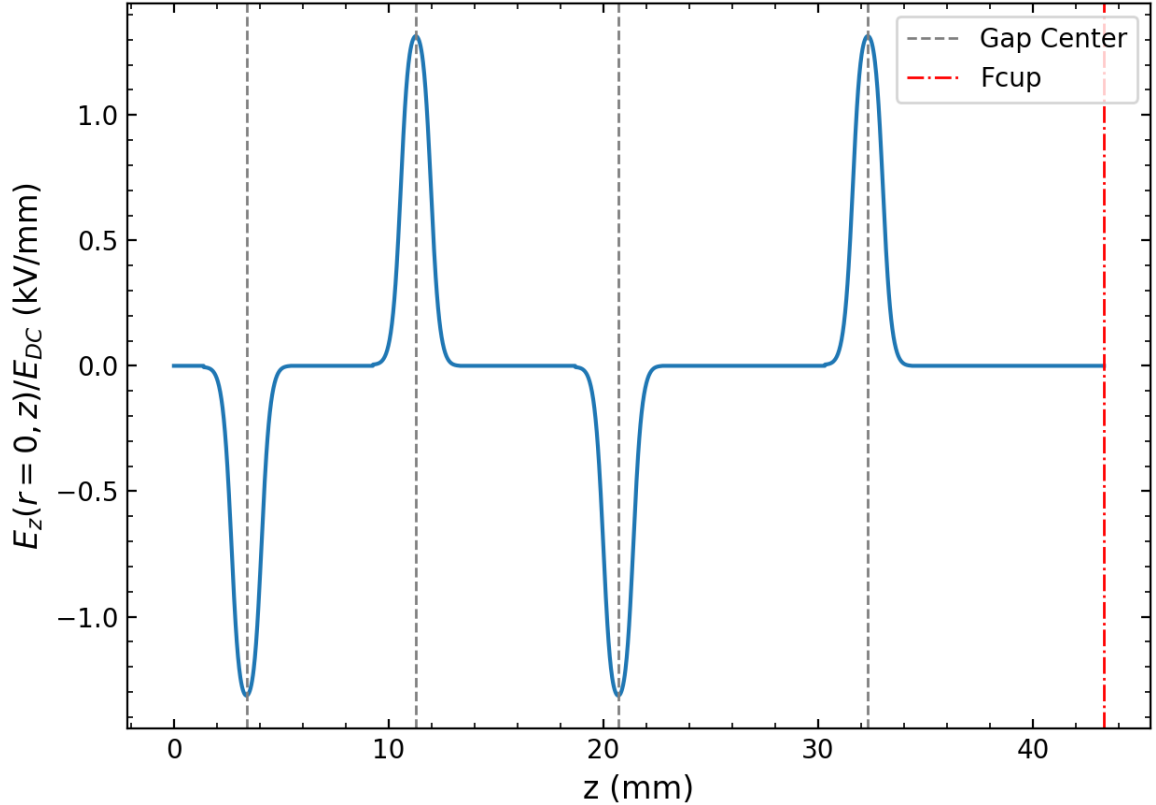


Figure 6.3: Normalized on-axis electric field for a 4-gap system at $t = 0$. The electric fields are normalized by the calculated DC-electric field $E_{\text{DC}} = V_g/g$ (V/m). The on-axis voltage will be less than the applied voltage in the real conductor. Since the field is extracted from the conductor, the field needs to be scaled to be greater than E_{DC} to ensure the on-axis voltage matches the desired voltage hence why the amplitudes in the plot are greater than unity. Each gap center is marked by a grey vertical dashed line and the distances between gap centers increases with increasing energy by $\beta\lambda_{\text{rf}}/2$ —the RF-resonance condition for gaps operating in the π -mode. The particle advance is terminated at the Faraday cup (Fcup) marked by a red vertical dashed line and placed 10 mm downstream from the final gap edge.

6.3.2 Comparing Lattice Designs

Fig. 6.4 shows a side-by-side comparison of the \mathcal{E} - t phase space for three design choices at the Fcup. Each plot is created using a kernel density estimator (KDE) to plot the 2D data of \mathcal{E} versus scaled relative time difference $\Delta t/\tau_{\text{rf}}$, where $\Delta t = t - t_s$ is the time relative to the synchronous design particle. The KDE plot is chopped at the extreme points where the data does not exist—a KDE plot will estimate data in these regions and can be misleading where

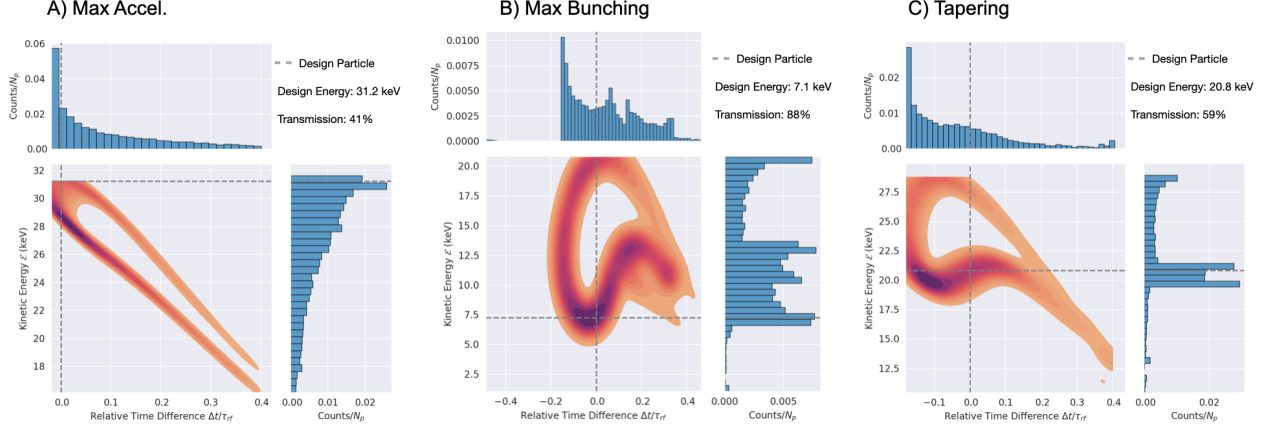


Figure 6.4: Reduced physics longitudinal model energy versus time Phase space plots and histograms for different design choices. Cases A), B), and C) correspond to: A) Maximum acceleration, B) Maximum bunching, and C) simultaneous acceleration and bunching.

there should not be particles with higher energy than possible. Corresponding \mathcal{E} and $\Delta t/\tau_{rf}$ histograms are given at the top and bottom respectively to help visualize the data. Vertical dashed lines correspond to the design particles energy and relative time. By definition, the vertical line will always be at $\Delta t = 0$. The intersection of the two lines represents where the design particle location is in phase-space. The final design particle kinetic energy and the total transmission percentage are noted in the top right corner of each plot.

In Fig. 6.4, designs for maximum acceleration and maximum bunching are shown in Panels A and B respectively with, unsurprisingly, drastically different phase spaces. In Panel A, gaps are spaced so $\phi_s = 0$ and particles are mostly clumped near the design particle with a monotonic decrease in the particle density towards lower energy and higher Δt . This is expected. Since the design particle is maximally accelerated it will be at the very tip of phase space at the highest energy possible. Particles close to the design particle will follow the design particle but, at each gap, receive less acceleration. With each pass through a gap, the density of particles near the design particle is reduced and the design particle continues to get further away in energy from the rest of the distribution. If the simulation were to be

run for more and more gaps, one would see counts in the bin near the design particle shrink more and more while bins at higher Δt and lower energy would increase from these particles drifting away from the design particle. Eventually, only the design particle will remain near $\Delta t = 0$ at maximum energy. The corresponding transmission is also the lowest as we should expect since there is no longitudinal focusing about the design particle. Maximum bunching with gaps spaced for $\phi_s = -\pi/2$ is shown in Panel B with 88% transmission compared to Panel A's 41% transmission. Although Panel B has the lowest energy gain with most particles being at 15 keV or lower. Although the phase space in Panel B is malformed from strong nonlinear longitudinal focusing, there are some telling features. The phase space wrapping is indicative of strong nonlinear longitudinal focusing. Recall that in a perturbative acceleration case, this phase space would look like an eye. Such strong acceleration distorts the shape, but we still see characteristic structure about the design particle. We verified this by running cases where the acceleration was perturbative for many gaps and were able to see the classic iris shape. By iteratively simulating the lattice with increasing V_g , we observed the iris distort into the shape shown. Also note the large region of particles missing between $-0.4\tau_{\text{ff}} \leq \Delta t \leq 0.2$. This is due to the strong acceleration since these earlier particles are going to see a strong decelerating field and will get pushed to negative velocities thus being filtered from the diagnostics. The small number of particles that made it through the lattice and were not filtered by time can only occur due to a fortuitous combination of phasing. These particles were able to keep up with the design particle in time until the last gap where they were mostly out of phase and experienced a strong deceleration once more. Although maximum bunching occurs when choosing $\phi_s = -\pi/2$, one must be cautious when in the non-perturbative regime. We will show later that one cannot rely on this design to give the best particle capture. To the far right is Panel C where we attempted to strike

balance between acceleration and bunching by tapering the acceleration from max bunching to max acceleration in the four gaps. In this case, the gap spacing was set to correspond to $\phi_s = (-\pi/2, -\pi/3, -\pi/6, 0)$. The familiar golf club structure is observed analogous to as discussed in Ch. 2. With this design, we can accomplish bunching and acceleration as noted by the 50.9% transmission with $\mathcal{E}_s = 20.1$ keV design energy. This test case only illustrates the methodology and has not been rigorously optimized to find the global optimum between transmission and acceleration.

All three cases have a void in the phase space where there is no data in seeming contradiction to the histograms. This is due to the cold beam injection and was verified by simulating a relatively hot beam with a thermal spread over the individual particle velocity that was a few percent of the beam velocity. When simulating the hot beam, there were no voids in the phase space distribution. This void arises because certain parts in phase space are not accessible due to the cold beam with strong acceleration and high transit time factor. These conditions create a thin band in phase space rather than a contiguous distribution.

As a final example we applied the 1D longitudinal model to a 16-gap lattice and obtained results are shown in Fig. 6.5. In this longer lattice case a new design case, Panel D), we call Fast Tapering is shown.

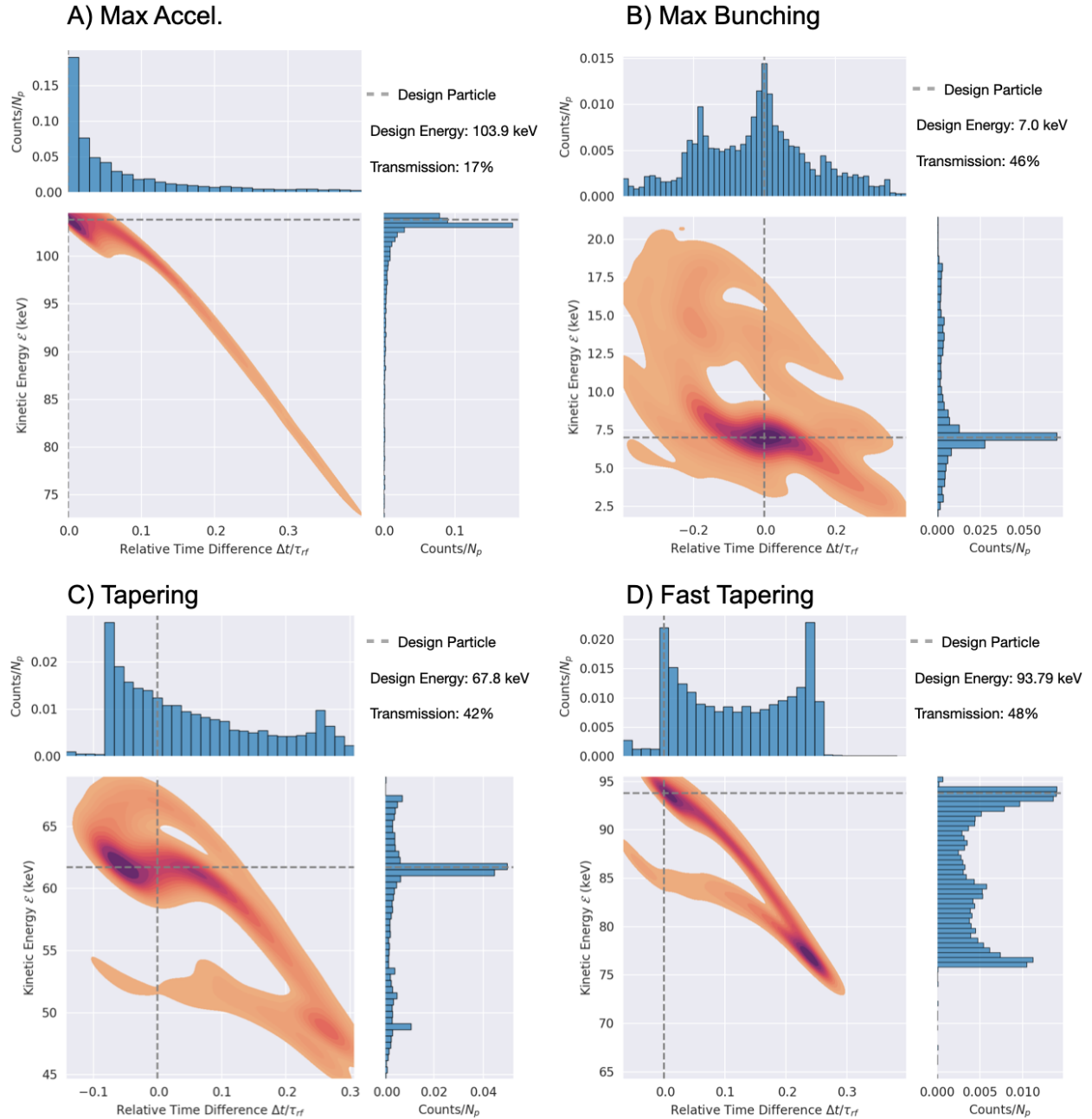


Figure 6.5: Analogous to Fig. 6.4, reduced physics longitudinal model simulations for an additional case for a 16 gap system. Panel D has been included for a design where the gaps are spaced for tapered bunching the first four gaps before transitioning to maximum acceleration.

This design is inspired from Panel C in Fig. 6.4 where, although there is excellent particle capture, the tapering is slow and the final energy is about half of the maximum acceleration case. Therefore, in this case we try a longer lattice where tapering is done for a few early

gaps and then we switch to maximum acceleration. Gaps are spaced for synchronous particle phasing is $\phi_s = (-\pi/2, -\pi/3, -\pi/6, -\pi/8, 120)$. Here, we introduce a short hand notation of $^k\phi$ which means that k -gaps are operating at the same phase ϕ . The longer lattice exacerbates the phase space morphing shown in Panels A, B, and C of Fig. 6.5. In Panel A, the “golf club” structure is clearly present and we see significant particle losses out of the selection window with clumping near the design particle as before and a longer band where these clumped particles eventually fall into after so many gaps. Panel B now shows a spiral shape, which is indicative of non-linear effects. Still, the majority of particles are concentrated near the design particle with high transmission as expected. Panel C is almost the same as before with some additional structure due to the longer evolution at max acceleration. The results of Panel D are surprising. Note the particle capture is better in Panel B. This shows that the phase choice of $\phi_s = -\pi/2$ in the first four gaps does not guarantee maximum particle capture when non-linearities are present. Additionally, after the first four gaps, the lattice is operating at maximum acceleration and we should expect a significant loss in particles. This effect is not an artifact of the reduced physics model and also occurs in the self consistent Warp simulations (see Ch. 8). This suggests that the system is sensitive to longitudinal over focusing. This fast tapering can inspire further design optimizations in future studies using the rapid running 1D model. However, we will see that the gaps also have strong transverse focusing effects (see Ch. 7) and this tapering design will have to be modified (see Ch. 8) to account for this. Nonetheless, we have an efficient code that allows us to probe this reduced parameter space and gain insight into the combined non-linear focusing and strong acceleration effects present. These results also illustrate that the complex physics involved that occur in a reduced physics model are only a small portion of the complexities that arise in the transverse motion and the full 3D self consistent simulations. Generally, one must

be cautious when tuning parameters for optimization of these systems using intuition based on conventional perturbative treatments as reviewed in Ch. 2. The rapid running code also enables application of machine learning methods, such as described in Part II of this thesis in a different project to optimize designs for specific applications.

6.4 Summary and Concluding Remarks

This chapter discussed the development of a rapid running reduced physics longitudinal code that uses the extracted on-axis electric field from Warp 3D field solvers of the acceleration gap arrays to model the acceleration of particles along z in 1D (no transverse motion). Good agreement with 3D Warp simulations gives confidence to the reduced physics model and we can expect that the model provides a good approximation to the longitudinal dynamics in a self-consistent simulation when particles are near the axis. Since the simulation advances particles spatially, we used filtering in time and energy to remove particles that drift too far from the design particle and would be lost in 3D simulations or in lab when the finite aperture is accounted for. We showed design cases for maximum acceleration, maximum bunching, and simultaneous acceleration and bunching. These results showed the nuances involved when designing in the non perturbative regime of this system and is a caution against using the conventional intuitions reviewed in Ch. 2.

Additional improvement to the reduced physics model would be to create an envelope model using the longitudinal Neuffer distribution [33]. The presented model does not include space charge. Creating an envelope model from the Neuffer distribution would allow inclusion of longitudinal space charge effects and be analogous to the transverse work done in Ch. 7. Thus, both the longitudinal and transverse motion would be modeled with a similar approach

creating a nice complementary simulation suite to model the dynamics of the system. In addition to an envelope model, better testing could be done to fine tune choices to guide experiment. Questions like, what time filtering should be employed to best fit experiment and/or desired outputs? Are there a better filtering criteria for considering particles to be lost? Lastly, more rigorous optimization schemes would be very beneficial. The optimization is currently done by hand using intuition developed by examining numerous plots and testing various designs. This is impractical and may lead to suboptimal conclusion. Optimization results depend strongly on defining what is best. Although transmission and energy are two important and intuitive metrics, we have seen how intuition can lead astray and the optimum found here and may not generally translate to optimums in the transverse motion or in full 3D simulations using Warp. These issues are illustrated in the source to target evaluations of optimized systems presented in Ch. 8.

Chapter 7. Transverse Dynamics in RF-MEMS

In order to efficiently transport the extracted ion beam array emerging from the injector, we wish to pre-shape the beam using a matching quadrupole section in order to rms-match the transverse beam envelope array to the acceleration lattice. The lattice is aperiodic due to the strong acceleration experienced by the beam in the acceleration gaps. This cannot be fully compensated for by adjusting ESQ focusing and is strong in the early stages of acceleration. This will prevent us from remaining perfectly matched as the beam accelerates. However, pre-shaping the beam to *almost* match will improve radial confinement the beam and limit beam loss from radial excursions produced by the defocusing space charge and thermal forces arising from the distribution of a beam of charged particles.

7.1 Geometric and Beam Parameters

The ESQs have geometric parameters: aperture radius r_p , rod radius R , physical length ℓ_q , and effective length ℓ_{eff} . Sec.5.4 found the optimal geometry had $R = 1.304 r_p = 0.717$ mm when choosing $r_p = 0.55$ mm and $\ell_q = 0.695$ mm. These values are used in the matching section unless otherwise specified. As shown previously, the ESQ fringe fields did not overlap when ESQ pairs were held at a distance (ESQ edge to ESQ edge) of $d > 2.5$ mm. We choose $d = 3$ mm to ensure that successive quadrupole fringe fields will not overlap. The injected beam parameters from the ion source were presented in Section 5.2 and are used here. These values are conveniently repeated in Table 7.1.

Geometric Parameters	
r_p	0.55 (mm)
R	0.717 (mm)
ℓ_q	0.695 (mm)
Beam Parameters	
\mathcal{E}_b	7 (keV)
I	10 (μA)
T_b	0.1 (eV)
Q	6.99×10^{-5}
ϵ	1.34 (mm-mrad)

Table 7.1: Geometric and initial-beam parameter settings used in the matching section simulations at injection energy.

7.2 Numerical Integration of KV-Envelope Equations

We apply the Kapchinskij-Vladimirskij (KV) envelope equations (discussed in Sec. 3) to model the DC beam evolution in the matching section. Ignoring acceleration and denoting the transverse rms-edge radii of the envelope as r_x and r_y , the KV-envelope equations are:

$$\begin{aligned}
r_x'' + \kappa_x(z)r_x - \frac{2Q}{r_x + r_y} - \frac{\epsilon^2}{r_x^3} &= 0, \\
r_y'' + \kappa_y(z)r_y - \frac{2Q}{r_x + r_y} - \frac{\epsilon^2}{r_y^3} &= 0
\end{aligned} \tag{7.1}$$

with generalized perveance Q , rms-edge emittance ϵ , and focusing strength in the transverse directions $\kappa_x(z)$ and $\kappa_y(z)$. For the nonrelativistic beam, Q can be expressed in terms of the beam current I , species mass m , and free-space permittivity ϵ_0 as

$$Q = \frac{qIm^{1/2}}{2^{5/2}\pi\epsilon_0\mathcal{E}_b^{3/2}}, \tag{7.2}$$

giving the value shown in Table 7.1. Similarly, the rms-edge emittance can be related to non-relativistic injected beam parameters as

$$\epsilon = 2r_b \left(\frac{k_B T_b}{\mathcal{E}_b} \right)^{1/2}, \quad (7.3)$$

where r_b is the beam radius, and T_b the beam temperature (see Sec. 3). Unless otherwise specified, we assume that the focusing strength is equal and opposite in the two planes such that $\kappa_x(z) = -\kappa_y(z) = \kappa(z)$ with $\kappa(z)$ given by Eq. 3.11.

These coupled equations can be numerically integrated in the usual way by transforming them into a set of coupled first-order differential equations. We first define the following variables:

$$\begin{aligned} u_x &= r_x, & v_x &= u'_x = r'_x, \\ u_y &= r_y, & v_y &= u'_y = r'_y. \end{aligned}$$

Using the newly defined variables Eq. 7.1 transforms to:

$$\begin{aligned} v'_x &= \kappa(z)u_x - \frac{2Q}{u_x + u_y} - \frac{\epsilon^2}{u_x^3}, \\ v'_y &= -\kappa(z)u_y - \frac{2Q}{u_x + u_y} - \frac{\epsilon^2}{u_y^3}. \end{aligned}$$

Discretizing the above formulas and rearranging results in the update equations for a forward

differencing scheme:

$$v_x^{n+1} = v_x^n + \Delta z \left(\kappa(z^n) u_x^n - \frac{2Q}{u_x^n + u_y^n} - \frac{\epsilon^2}{(u_x^n)^3} \right),$$

$$v_y^{n+1} = v_y^n - \Delta z \left(\kappa(z^n) u_y^n + \frac{2Q}{u_x^n + u_y^n} + \frac{\epsilon^2}{(u_y^n)^3} \right).$$

The above equations can be written more compactly if the terms in the parentheses are wrapped into a function $f_x^n(\kappa(z^n), u_x^n, u_y^n)$, with a similar expression for f_y^n . The coupled first order differential equations can then be written as

$$v_x^{n+1} = v_x^n + \Delta z f_x^n,$$

$$u_x^{n+1} = u_x^n + \Delta z v_x^{n+1},$$

with a similar expression for v_y and u_y . Note that v_x updated first and then used to update u_x which gives increased accuracy compared to a forward/backward Euler algorithm at no cost to computational complexity. This scheme is known as the Euler-Cromer or Semi-Implicit Euler scheme [34] and is used to solve the KV-envelope equations.

7.3 Modeling the RMS-Edge Envelope Evolution Without Acceleration

We construct a beam array matching section from four ESQs. As discussed in Sec. 5.4, we use $d = 3$ mm to prevent the ESQ array fringe fields from overlapping. An example of a matching section lattice simulation is shown in Fig. 7.1. The extracted $\kappa(z)$ is shown in blue

with the hard-edge equivalent model overlaid in green. The $\kappa(z)$ amplitude is shown set for typical biases of a matching solution by scaling the extracted fringe field of the ESQ arrays. At the top of the plot the physical ESQ extents are represented by black rectangles and the geometric variables shown. An ESQ strength normalization factor $\hat{\kappa}$ is calculated from the 2D equation using the half-breakdown voltage limit $V_q = 500$ V and the values listed in Table 7.1:

$$\hat{\kappa} = \frac{V_q}{\mathcal{E}_b r_p^2} \Big|_{V_q=500\text{V}} = 4.72 \times 10^5 \text{ m}^{-2}.$$

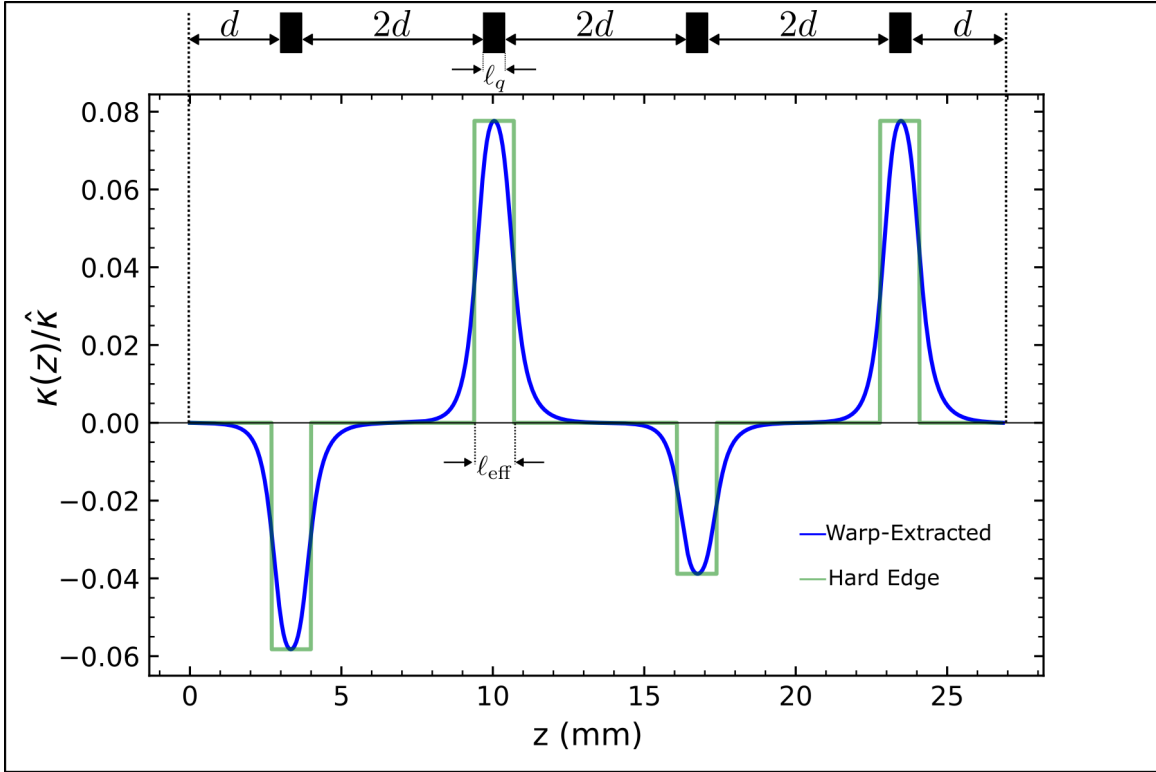


Figure 7.1: The extracted $\kappa(z)$ in blue and the hard-edge equivalent in green are both normalized by $\hat{\kappa}$, the focusing strength calculated in the 2D case using the initial beam parameters and the half breakdown limit for V_q . At the top of the figure is a schematic showing the ESQ array locations in the lattice. The case shown has ESQ bias voltage settings (from left to right) of $(V_1, V_2, V_3, V_4) = (-59.97, 79.94, -39.98, 79.97)$ V. Note these are small fractions of the half breakdown voltage safety limit.

7.3.1 Benchmarking with Warp

Using the placement scheme shown in Fig. 7.1, the ESQ array conductors are loaded onto a 3D mesh in Warp with periodic transverse boundary conditions for an array of beams and Dirichlet boundary conditions at longitudinal ends of the simulation box. The Warp simulations include beam space charge. A continuous beam of Ar^+ ions is injected with the parameters listed in Table 7.1 along with zero initial envelope angle to model the beam exiting the source aperture $r_s = 0.25$ mm. Particles are injected uniformly and continuously at each time step to create a uniformly filled cylindrical beam that extends the length of the simulation box. The grid is discretized with $dx = dy = 30 \mu\text{m}$, $dz = 63 \mu\text{m}$, and a $dt = 0.24$ ns time step is employed. The grid resolution was set so that higher resolution runs resulted in differences within numerical noise. To calculate the statistical evolution of the beam envelope in the full 3D beam simulation, the rms-edge statistics are calculated at each grid cell using data for particles within the selected grid cell. Fig. 7.2 shows an example of the discretized with an injected x - z beam projection. Within each grid-cell, the beam moments are calculated and stored for that z -location. At the end of the simulation when particles fill the mesh, there will be 10^6 particles injected. Particles whose transverse position exceed r_p are absorbed and not included the statistical moments calculations. This can influence the statistical measure calculations. However, voltages were kept relatively low for sufficiently low focusing strengths that the envelope excursions are well within r_p and particle losses were negligible and did not impact the calculated statistical measures.

We compare the statistical evolution of the beam edge in a full 3D simulation with the evolution calculated by integrating the KV-envelope equations in Eq. 7.1. The statistical moments of interest are the rms envelope edge radii r_x and r_y , envelope angles r'_x and r'_y

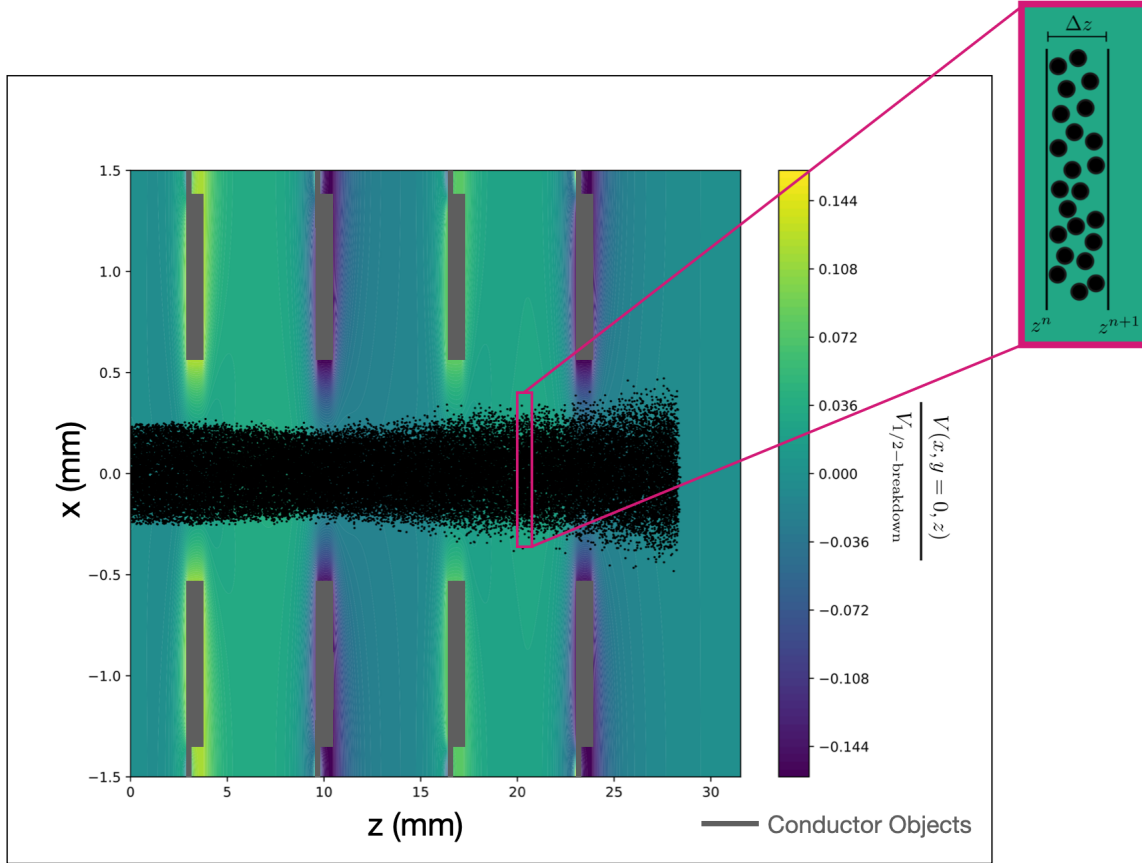


Figure 7.2: Warp simulation example of a continuously injected beam and the discretized grid for which the moments are calculated. The potential is normalized by the half-breakdown limit $V_{1/2\text{-breakdown}} = 500$ V and is generated by the ESQ conductors in gray. A random selection of 10,000 macro-particles are plotted as black dots. The outer panel is a close-up view of the discretized grid. Macro-particles within Δz of the two grid points are used to calculate the statistical moments which are then used to evaluate the statistical envelope evolution in the four ESQ matching section.

given by Eq. 3.20. Each statistical measure is extracted from Warp macro-particle data using the correct call found in the “Z_moments dump” section in the top.v file. For example, $2\sqrt{\langle x^2 \rangle_{\perp}}$ can be extracted from the `warp.top.xrmsz` array after depositing moments in Warp.

Using the aforementioned tools and schemes, we analyze the rms-edge evolution using the KV envelope model for an extracted Warp $\kappa(z)$ from ESQ array field solves, a hard-edge model calculated using the equivalency procedure given in Sec. 3, and a full 3D Warp simulation that advances particles in a potential field generated by ESQ conductors while including full space charge forces consistent with boundary conditions. The results for each model and the statistical measures in Eq. 3.20 are shown in Fig. 7.3. There is excellent agreement between the extracted $\kappa(z)$ model (dashed line) and the full 3D Warp simulation (solid line). This suggests that rapid testing can be done with the extracted $\kappa(z)$ and the results used without having to adjust for model differences and fine-tune with a full 3D simulation. The rapidity with the solver model will be especially advantageous when optimizing the voltages to shape the beam for a target matched envelope condition. The hard-edge model shows reasonable agreement but is slightly offset as is shown by the dotted line. Note that because the acceleration is non perturbative, the successive lattice periods will be significantly different resulting in asymmetric successive periods. Therefore, it will not be possible to perfectly match the beam to successive periods of the acceleration lattice. In this regard, the discrepancies of the hard-edge model may not be so detrimental or noticeable after a few lattice periods. The hard-edge model is the most computationally efficient since it does not require an initial solve of the ESQs to extract the gradient and therefore be easier to apply in some cases. However, if the beam can be perfectly matched, then the small error could result in a mismatched beam and an inefficient transport early on in the lattice. Lastly, the disagreements are more apparent early in the simulation and in the x -

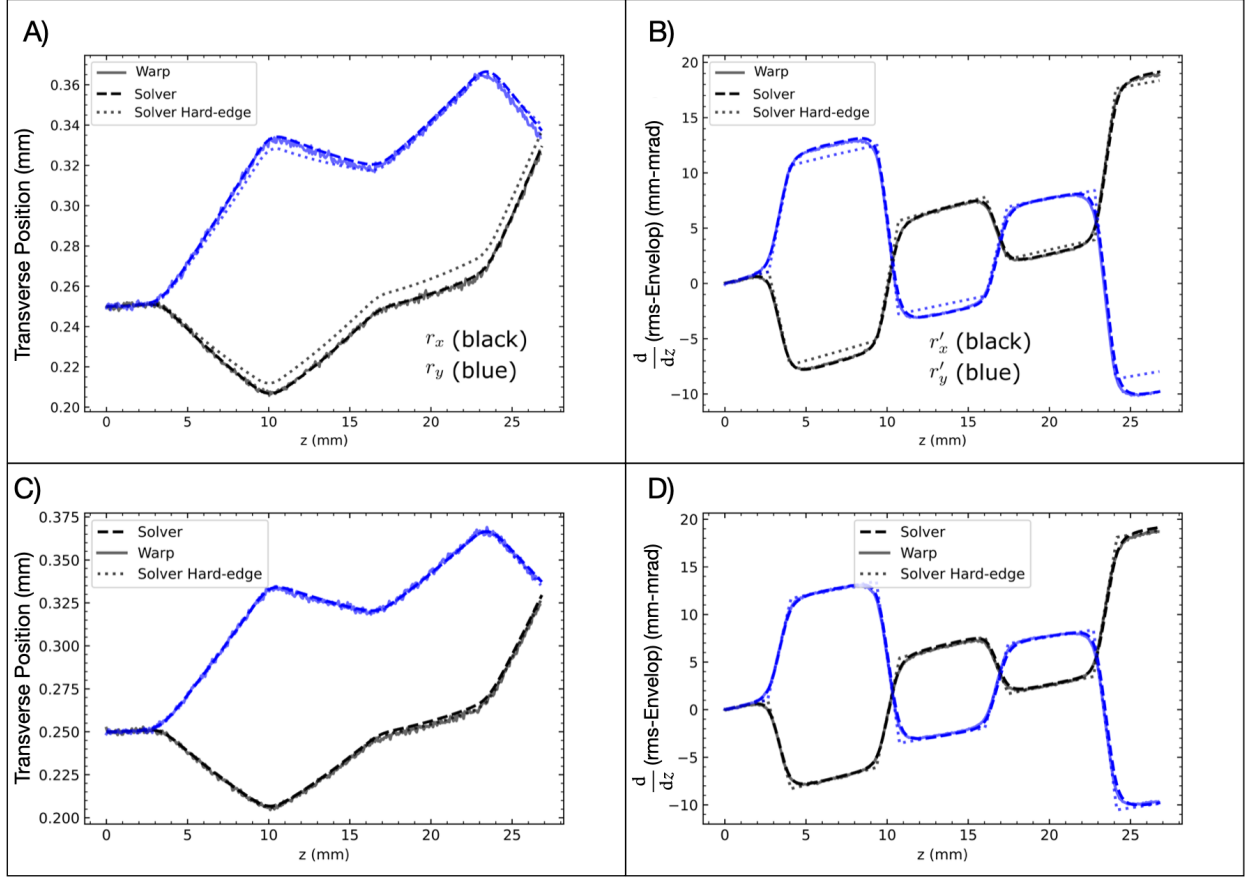


Figure 7.3: The rms-edge of the beam envelope boundary shown in Panel A and Panel C respectively, and the envelope angles r'_x and r'_y of the beam edge shown in Panel B and Panel D respectively. The x -coordinate is in black and y -coordinate in blue. Each model is shown on the plot: a full 3D simulation with space charge done with Warp (solid line), the KV model using the extracted $\kappa(z)$ (dashed), and the KV model using a hard-edge model (dotted). The full simulation and the extracted $\kappa(z)$ overlay showing excellent agreement and suggesting that the extracted $\kappa(z)$ using the KV model can serve in-place of an expensive full 3D simulation. The bottom row is a result of adjusting the effective length in the hard-edge model for more focusing. With a 10% increase in $\kappa(z)$, the hard-edge model is brought into agreement with the other two models.

direction. It is not fully understood why this is the case, but a potential reason is that the hard-edge model is not sufficiently capturing the focusing encountered in the extracted $\kappa(z)$. To test this hypothesis, ℓ_{eff} was increased by 10% to provide more focusing and the bottom row of Fig. 7.3 shows results obtained. With the increased focusing, the hard-edge model was in excellent agreement with the other two and suggests that the hypothesis was

correct. If computation time was an optimization criterion, these results suggest that one could first compare the extracted $\kappa(z)$ model with a full simulation. From there, the hard-edge model could be fine-tuned with the extracted model and then used from then on to save computational time. In this case, the computational time is not an issue and the extracted model is preferred since it captures full fringe field effects for analysis.

7.3.2 Finding Acceleration Lattice Matched Conditions Using the Nelder-Mead Search Algorithm

With a tool capable of rapidly integrating the KV-envelope equations, we now wish to extend the functionality to providing a matching condition for a given lattice period with ESQ focusing and acceleration. The four ESQ matching section can then be used to match the beam emerging from the injector into the lattice period. A previously developed Mathematica algorithm was used to benchmark [35] our algorithm developed in Python. The Mathematica algorithm assumes that the lattice is periodic and, given various input parameters, the user can extract the matched solution among other quantities. Additionally, $\kappa(z)$ can be given directly as an array of values along with the z -mesh and a match condition for r_x , r_y , r'_x , and r'_y will be found if one exists. Unfortunately, the Mathematica tool does not allow for providing acceleration in the lattice and thus we are limited to comparing lattices with only ESQ focusing and constant Q and ϵ . Fig. 7.4 shows an example output for a doublet configuration operating at $V_q = 400$ V for the Mathematica algorithm. The top panel A) shows $\kappa(s)$ where $s \equiv z$. Note that $\kappa_x = -\kappa_y = \kappa$. The bottom panels shows r_x (black) and r_y (red) in Panel B). In Panel C) are r'_x (black) and r'_y (red).

The matching was done for the first lattice period with the energy gain of the two

	r_x (mm)	r_y (mm)	r'_x (mrad)	r'_y (mrad)
Initial Condition	0.250	0.250	0	0
Match Solution	0.155	0.280	0.847	-11.146

Table 7.2: Initial conditions (envelope edge radii and envelope edge angles) used for initial condition is shown in the first row. The matched coordinates are shown in the second row.

gaps in the period nulled and a beam energy of $\mathcal{E}_b = 7 \text{ keV}$. Using maximum acceleration design for $V_g = 5 \text{ kiloV}$, the resulting lattice period of the first period has length length $L_p = 25.148 \text{ mm}$ (calculated for $V_g = 5 \text{ kV}$). A voltage of $V_q = \pm 200 \text{ V}$ was used for the focusing doublet. Using the values for $Q = 6.99 \times 10^{-5}$ and $\epsilon = 1.34 \text{ mm-mrad}$, the algorithm found the matched solutions listed in Table 7.2. Fig. 7.4 shows $\kappa(z)$ along the lattice period in Panel A, and the matched envelope evolution for the transverse envelope radii and angles in Panel B and Panel C respectively. With the matched condition constructed, we now compare results with simulation.

The algorithmic procedure is shown schematically in Fig. 7.5. We wish to find the matched coordinates given in Table 7.2 given the symmetric doublet voltages $V_q = \pm 200 \text{ V}$. Thus, the algorithm will be voltage fixed and the initial coordinates varied until matched condition is achieved. Writing the 4D coordinates in a compact vector form as $\vec{r} = (r_x, r_y, r'_x, r'_y)$, we can define the final coordinates extracted from integrating Eq. 7.1 as

$\vec{r}_f = (r_{x,f}, r_{y,f}, r'_{x,f}, r'_{y,f})$ and a set of target coordinates we wish to match as

$\hat{\vec{r}} = (\hat{r}_x, \hat{r}_y, \hat{r}'_x, \hat{r}'_y)$. A matched solution requires $\vec{r}_i = \vec{r}_f$ where \vec{r}_i are the initial coordinates at the start of the lattice period. For a matched condition, we therefore have $\vec{r}_i = \vec{r}_j = \hat{\vec{r}}$.

A cost function can then be defined as:

$$J(\vec{r}_f, \hat{\vec{r}}, \vec{S}; \vec{V}) = \frac{1}{2} \sum_{j=1}^4 \frac{(r_{j,f} - \hat{r}_j)^2}{S_j^2} \quad (7.4)$$

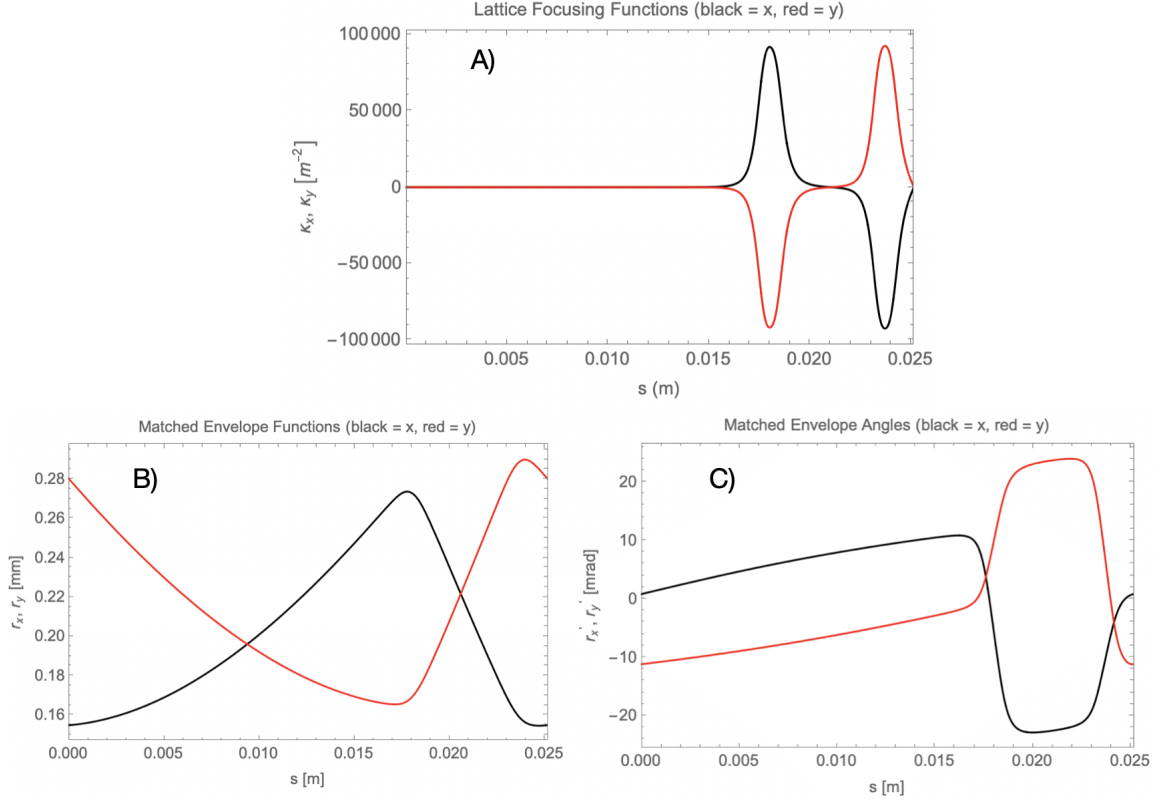
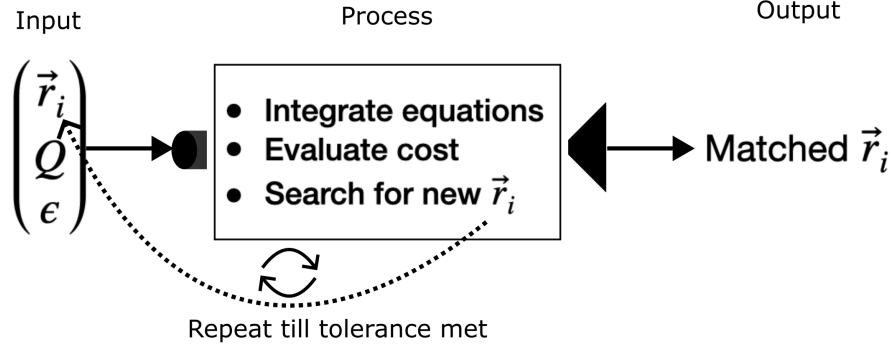


Figure 7.4: Matched solutions given by a previously developed Mathematica algorithm [35]. The software assumes a periodic lattice and does not account for acceleration. Panel A shows the focusing strength from the ESQ conductor κ_q in the transverse direction (x in black, y in red) for single lattice period. Panel B shows the evolution of the matched rms-edge radii r_x and r_y in black and red respectively over a single lattice period. Panel C shows the evolution of the matched rms-edge angle r'_x and r'_y in black and red respectively over a single lattice period.

where the index j runs over the components of the vectors \vec{r}_f , $\hat{\vec{r}}$, and \vec{S} . The vector \vec{S} is a set of normalization coefficients chosen so the algorithm can better converge to accurate matched coordinates and angles. We made the natural choice of r_p to be the normalization factor for the position and through simulation we found a maximum envelope excursion value of $r'_{\max} = 21$ mrad. With this we can define the vector of normalization values to be $\hat{\vec{S}} = (r_p, r_p, r'_{\max}, r'_{\max})$. We used the Nelder-Mead search algorithm to minimize the cost function [36, 37]. This is a robust and derivative-free searching algorithm that is advantageous here because of the low computational burden [38]. With an initial guess at \vec{r}_i , the

Voltage Fixed

$$\vec{r}_i = (r_{x,i}, r_{y,i}, r'_{x,i}, r'_{y,i})$$



Coordinate Fixed

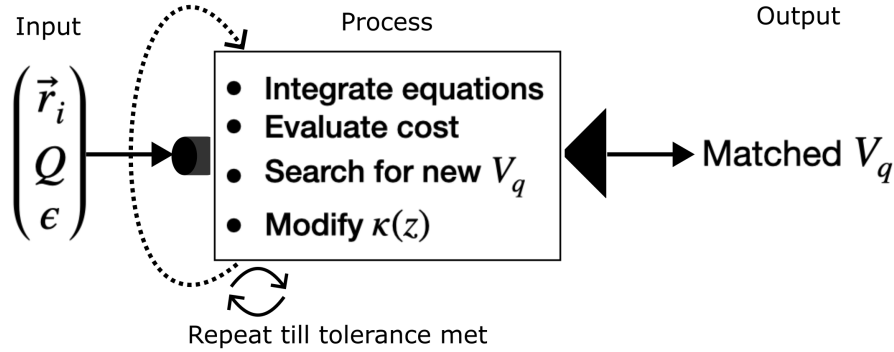


Figure 7.5: Flow diagram of the optimization algorithm for finding a matched condition. There are two types of matching that can be done. In voltage fixed matching, we seek a matched solution given the ESQ voltages and thus $\kappa(z)$. The algorithm will then iteratively perform the optimization process and vary the input coordinates \vec{r}_i until a set tolerance is met. In coordinate fixed matching, we seek a matched solution given \vec{r}_i and \vec{r}_f . The algorithm will iteratively perform the optimization process, vary the voltages, and recompute $\kappa(z)$ until a set tolerance is met. Both schemes use the Nelder-Mead search algorithm and rely on a created cost function to guide the searching.

algorithm will integrate the KV-equations with the set $\kappa(z)$, Q , and ϵ . Once integrated, the cost function will be computed using the extracted \vec{r}_f . This cost function informs the search algorithm on how to pick the next set of coordinates for minimizing the cost. This procedure will be repeated until the change in the cost function is below a set tolerance. Executing the algorithm produced the matched coordinates to within numerical noise verifying the algorithm.

Additionally, we modified the algorithm to also search for the required V_q that yield a matched solution \vec{r}_i giving a coordinate fixed algorithm. We used the matched coordinates found above to seek the required voltage and reproduced the result with $V_q = \pm 200$ V. We also tested different parameter settings then compared to the Mathematica algorithm to increase confidence.

7.4 Modeling the RMS-Edge Envelope Evolution with Acceleration

7.4.1 Consideration for Creating a Benchmark with Warp

In the previous section Warp was used as a benchmark to establish credibility in the modeling of ESQ effects in the transverse evolution with the KV envelope equations. We wish to do the same in modeling the focusing with acceleration. Two acceleration gaps are simulated in order to get the correct fields from the periodic array geometry (see Sec. 5.3) and the gaps are placed symmetrical about the midpoint of $z = 0$ with enough spacing between the simulation box-ends and the midpoint to ensure there is no interference or fringe overlap. Instead of an RF-varying electric field, the time variation is turned off giving a static DC-field which provides a good model if the particle's gap traversal time is small compared to τ_{rf} . Comparing the traversal time in the worst case at injection energy gives:

$$\frac{\tau_{rf}}{\Delta t_{\text{gap}}} \approx 0.1,$$

The transit-time factor (see Sec. 2) can also be used as a gauge. Using the hard-edge approximation, the transit time factor at injection energy is:

$$T = \frac{\sin\left(\frac{\pi g}{\beta\lambda}\right)}{\left(\frac{\pi g}{\beta\lambda}\right)} = 0.964,$$

which shows that the DC-assumption is valid. As before, a stream of particles is continuously injected into the system until the beam-head reaches $z = 0$. As described in Sec. 7.3.1, the particle statistics can be extracted. The simulation extends from $[-z_{\max}, z_{\max}]$, so the extraction of Eq. 3.20 is carried out from $[-z_{\max}, 0]$. However, when plotting, the z -axis bounds are shifted to $[0, z_{\max}]$ in order to avoid any confusion that may arise from a negative distance.

7.4.2 Thin-gap KV Envelope Model for Transverse Acceleration Effects from Acceleration

Here we analyze matching for prototypical lattice periods with ESQ focusing and pairs of acceleration gaps. The KV-envelope in Eq. 7.1 can also model the acceleration of the beam which changes the equations to:

$$\begin{aligned} r_x'' + \frac{(\gamma_b\beta_b)'}{(\gamma_b\beta_b)}r_x' + \kappa(z)r_x - \frac{2Q}{r_x + r_y} - \frac{\epsilon^2}{r_x^3} &= 0, \\ r_y'' + \frac{(\gamma_b\beta_b)'}{(\gamma_b\beta_b)}r_y' - \kappa(z)r_y - \frac{2Q}{r_x + r_y} - \frac{\epsilon^2}{r_y^3} &= 0. \end{aligned} \quad (7.5)$$

Here, $\beta_b = v_b/c$ and the usual relativistic factor $\gamma_b = (1 - \beta_b^2)^{-1/2}$, which satisfies $\gamma_b \approx 1$ for our nonrelativistic system. In earlier sections Q and ϵ were assumed constant as the beam progresses through the lattice. However, acceleration will change the beam energy and

will therefore change the value of these parameters according to Eqs. 7.2 and 7.3. Scaling equations for these parameters can be derived if a thin gap model is taken resulting in abrupt gains in the energy at gap locations. In this approximation, the beam radius does not change, but the angle and energy are abruptly changed. Here, we denote a parameter value as χ^- for the value of χ before the gap, and χ^+ for the value after the gap.

Starting with Q , the values before and after the gaps are

$$Q^\pm = \frac{qIm^{1/2}}{2^{5/2}\pi\epsilon_0(\mathcal{E}_b^\pm)^{3/2}}.$$

The energy gain can be expressed as $\mathcal{E}_b^+ = \mathcal{E}_b^- + \Delta\mathcal{E}_g$, where $\Delta\mathcal{E}_g$ is the energy change from the acceleration gap. Dividing Q^+ by Q^- and substituting for \mathcal{E}_b^+ gives:

$$Q^+ = \frac{Q^-}{(1 + \Delta\mathcal{E}_g/\mathcal{E}_b^-)^{3/2}}. \quad (7.6)$$

From the derivation of ϵ in Sec. 3, one sees that for unchanged T_b and r_b , $\epsilon \propto \mathcal{E}_b^{-1/2}$. Using steps for ϵ^+ and ϵ^- with conservation of normalized emittance $\beta\epsilon = \text{const}$, the change in rms-edge emittance through the gap is:

$$\epsilon^+ = \frac{\epsilon^-}{(1 + \Delta\mathcal{E}_g/\mathcal{E}_b^-)^{1/2}}. \quad (7.7)$$

Lastly, the particle angle x'^+ can be calculated:

$$\begin{aligned}
x'^+ &\equiv \frac{v_x^+}{v_z^+}, \\
&= \frac{v_x^-}{v_z^- + \Delta v_z}, \\
&= \frac{v_x^- v_z^-}{1 + \Delta v_z / v_z^-}, \\
&= \frac{x'^-}{1 + \frac{\sqrt{\mathcal{E}_b^- + \Delta \mathcal{E}_b} - \sqrt{\mathcal{E}_b^-}}{\sqrt{\mathcal{E}_b^-}}}, \\
\therefore x'^+ &= \frac{x'^-}{\sqrt{1 + \Delta \mathcal{E}_g / \mathcal{E}_b^-}}.
\end{aligned}$$

The formula for y'^+ will be of the same form. Eq. 3.20 gives $r'_x = 2\langle xx' \rangle / \sqrt{\langle x^2 \rangle}$. Since $x^+ = x^-$ and $y^+ = y^-$ holds for the thin gap, using these changes in the statistical envelope angle formula and similar steps tkane for Q and ϵ gives:

$$r'_x{}^+ = \frac{r'_x{}^-}{\sqrt{1 + \Delta \mathcal{E}_g / \mathcal{E}_b^-}}, \quad (7.8)$$

with a similar expression for r'_y . Eqs. 7.6, 7.7, and 7.8 can be used to update Q , ϵ , r'_x and r'_y respectively at the center of each gap when integrating the KV equations to take into account acceleration. The thin gap KV model was benchmarked against 3D Warp simulations for a single gap. The initial envelope edge radii and edge angle were set to $r_x = r_y = 0.25$ mm and $r'_x = r'_y = 2.7$ mrad and gap model results were compared with self consistent 3D Warp simulations of the thick gap. Comparisons are shown in Fig. 7.6. With equal initial conditions the envelope evolution of r_x and r_y will be identical and therefore only r_x and r'_x plots are shown. The gap edges are marked by thick vertical dashed lines. Clearly, the thin gap model

does not capture the physics as can be seen by the monstrous disagreement between the two models. Due to these poor results, a new model is developed in Sec. 7.4.3 to account for the strong gap acceleration effects evident in the Warp simulations that are not captured in the thin gap model of the KV envelope equations. Note that the strong acceleration of the gap is inducing large envelope angle changes and related envelope angle evolution. These large effects must be properly modeled in viable a reduced physics formulation.

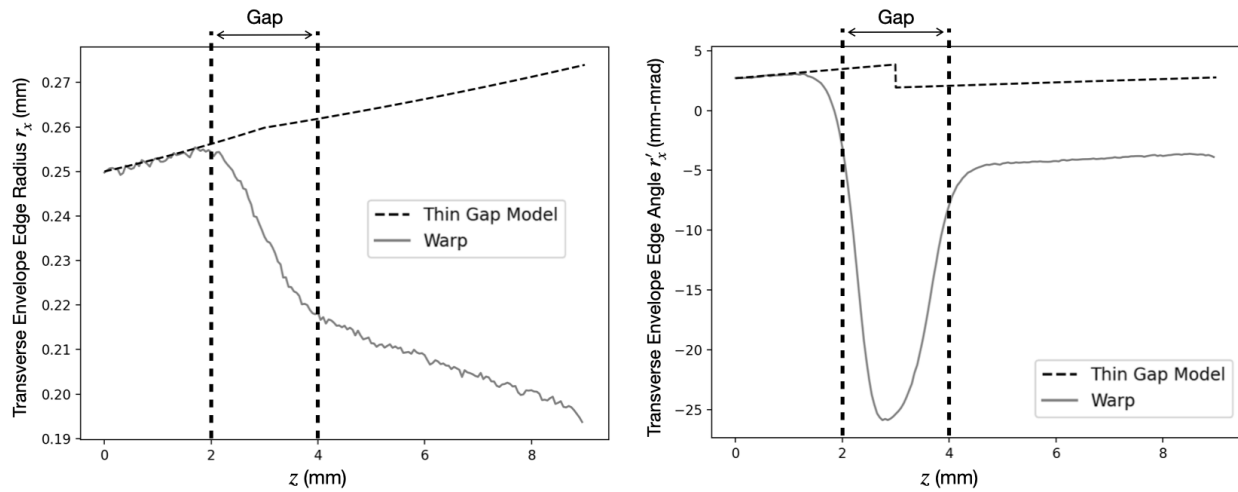


Figure 7.6: Comparing the envelope rms-edge radius and angle evolution through a single DC acceleration gap between 3D Warp simulations and a thin-gap KV model. On the left is the rms-edge envelope radii and on the right the rms-edge envelope angle. The thin-gap envelope model result is given by dashed curves and Warp results are solid curves. Dashed vertical lines mark the axial extent of the gap edges.

7.4.3 Thick-gap KV Envelope Model to Capture Acceleration Effects

In the thin-gap model Eqs. 7.6 and 7.7 are related to \mathcal{E}_b . In the electrostatic approximation, the energy changes as a function of z by

$$\mathcal{E}_b(z) = \mathcal{E}_{b,i} - qV_0(z) \cos(\phi_s)$$

where $V_0(z)$ is the on-axis electric potential. As the beam traverses the acceleration gap the beam line charge density λ will remain constant. Rewriting Q in terms of the charge density gives:

$$Q = \frac{q\lambda}{2\pi\epsilon_0 m \gamma_b^3 \beta_b^2 c^2}.$$

In the nonrelativistic limit, $\gamma_b = 1$ and using the constancy of λ , a scaling relation can be found for the variation of Q in the gap. To do so, define the initial perveance before the gap in the non-relativistic limit as

$$Q_i = \frac{q\lambda}{2\pi\epsilon_0 m \beta_{b,i}^2 c^2}.$$

Then take similar steps as before by finding Q/Q_i to give:

$$\begin{aligned} Q &= \frac{Q_i}{\beta_b^2 / \beta_{b,i}^2}, \\ &= \frac{Q_i}{1 + \frac{qV_0(z) \cos(\phi_s)}{\mathcal{E}_{b,i}}}. \end{aligned}$$

where, the last expression comes from using $\beta_b^2 = 2\mathcal{E}_b/(mc^2)$ and the equation for \mathcal{E}_b above.

It is useful to define a normalized potential as $\tilde{V}_0(z) = V_0(z)/V_g$. Rewriting the perveance

formula with $\tilde{V}_0(z)$ gives:

$$Q(z) = \frac{Q_i}{1 + \frac{Vg \cos \phi_s}{\mathcal{E}_{b,i}} \tilde{V}_0(z)}. \quad (7.9)$$

The gap evolution of the rms-edge emittance ϵ can be found in a similar fashion. We start from the constance of the normalized rms-edge emittance $\epsilon_n = (\gamma_b \beta_b) \epsilon$ in the gap. Once again, the non-relativistic limit is used and using $\epsilon_i = \epsilon_{n,i} / \beta_{b,i}$ to take a ratio gives:

$$\begin{aligned} \epsilon(z) &= \frac{\epsilon_i}{\beta_b / \beta_{b,i}}, \\ &= \frac{\epsilon_i}{\left(1 + \frac{qV_0(z) \cos \phi_s}{\mathcal{E}_{b,i}}\right)^{1/2}}. \end{aligned}$$

Rewriting this expression with $\tilde{V}_0(z)$ gives:

$$\epsilon^2(z) = \frac{\epsilon_i^2}{1 + \frac{Vg \cos \phi_s}{\mathcal{E}_{b,i}} \tilde{V}_0(z)}, \quad (7.10)$$

where the square was taken since it is ϵ^2 in the KV envelope equation.

Finally, the acceleration term in the KV envelope equation must be treated. We examine the term

$$\frac{(\gamma_b \beta_b)'}{\gamma_b \beta_b} r'_x,$$

and analogous results hold in the y -plane. In the non-relativistic limit, we have

$$\frac{(\gamma_b \beta_b)'}{\gamma_b \beta_b} r'_x \approx \frac{\beta_b'}{\beta_b} r'_x.$$

Writing β_b in terms of \mathcal{E}_b gives

$$\begin{aligned}\beta_b(z) &= \sqrt{2\mathcal{E}_b(z)/(mc^2)}, \\ &= \sqrt{\frac{2}{mc^2}(\mathcal{E}_{b,i} + qV_0(z) \cos(\phi_s))}^{1/2}.\end{aligned}$$

Taking the derivative with respect to z gives:

$$\beta'_b(z) = \sqrt{\frac{2}{mc^2} \frac{(q/2)V'_0(z) \cos(\phi_s)}{(\mathcal{E}_{b,i} + qV_0(z) \cos(\phi_s))} }^{1/2}.$$

The acceleration term can now be written as:

$$\frac{\beta'_b}{\beta_b} r'_x = \frac{(q/2)V'_0(z) \cos(\phi_s)}{\mathcal{E}_{b,i} + qV_0(z) \cos(\phi_s)} r'_x.$$

Lastly, rewriting in terms of the scaled gap potential \tilde{V}_0 gives:

$$\frac{\beta'_b}{\beta_b} r'_x = \frac{\frac{Vg \cos \phi_s}{2\mathcal{E}_{b,i}} \tilde{V}'_0(z)}{1 + \frac{Vg \cos \phi_s}{\mathcal{E}_{b,i}} \tilde{V}_0(z)} r'_x. \quad (7.11)$$

So far, all that has been done is to account for the longitudinal energy changes in a thick-gap formulation. An additional term not yet considered results from the radial component of the electric field. Here, the steps in Ref. [7] will be paralleled to derive the transverse focusing field of the applied gap field. Approximate the applied electric field as static ($\partial/\partial t = 0$). The field also is rotationally symmetric with $\partial/\partial\theta = 0$, $\partial/\partial\phi = 0$. Additionally, using the static and source free Maxwell's equations in vacuum, the applied electric field can be written as the gradient of a scalar potential $f^a(r, z)$ which obeys Laplace's equation. In spherical

coordinates, Laplace's equation is

$$\nabla^2 f^a(r, z) = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial f^a}{\partial r} \right) + \frac{\partial^2 f^a}{\partial z^2} = 0.$$

To solve this equation, we take a power series solution of the form

$$f^a(r, z) = \sum_{n=0}^{\infty} f_{2n}^a(z) r^{2n}, \quad \text{with } n = 0, 1, 2, \dots$$

An even series is assumed to ensure that the radial component of the electric field is zero on-axis $E_r^a(r = 0, z) = 0$. To see this, note that taking the gradient of the series solution will give the electric field. If there is a linear r -term, then taking the gradient with respect to r will give a constant field meaning that there will be a non-zero field at $r = 0$. Denoting $' = \partial/\partial z$ and plugging the solution into Laplace's equation gives

$$\sum_{n=0}^{\infty} (2n)^2 f_{2n} r^{2(n-1)} + \sum_{n=0}^{\infty} f_{2n}'' r^{2n} = 0,$$

where the superscript a has been dropped reduce clutter. The first term on the left hand side is 0 for $n = 0$ and the index can be advanced by 1 giving

$$\sum_{n=1}^{\infty} (2n)^2 f_{2n} r^{2(n-1)} + \sum_{n=0}^{\infty} f_{2n}'' r^{2n} = 0.$$

Reindexing the first term by letting $m = n - 1$ gives

$$\sum_{m=0}^{\infty} 4(m+1)^2 f_{2m+2} r^{2m} + \sum_{n=0}^{\infty} f_{2n}'' r^{2n} = 0.$$

Since n and m are dummy indices, both sums can be written in terms of the same index m .

Doing so gives

$$\sum_{m=0}^{\infty} 4(m+1)^2 f_{2m+2} r^{2m} + \sum_{m=0}^{\infty} f_{2m}'' r^{2m} = 0.$$

For $r \neq 0$ the solution requires coefficients of each radial power equate to zero which establishes the recursive relationship:

$$f_{2m+2} = -\frac{f_{2m}''}{4(m+1)^2}.$$

The first few terms in the series are:

$$f(r, z) = f_0(z) - \frac{f_0''(z)}{4} r^2 + \frac{f_0^{(4)}(z)}{64} r^4 - \mathcal{O}(r^6),$$

where $f_0(z) = f(r=0, z)$ is the on-axis potential. The transverse electric field can now be extracted by taking the gradient. Doing so and truncating at the linear term in r gives:

$$E_r = -\nabla_r f(r, z) = \frac{f_0''}{2} r = -\frac{r}{2} \frac{\partial E_z}{\partial z}.$$

Converting to cartesian coordinates the transverse electric field components are:

$$E_x^a = \frac{1}{2} f''|_{r=0} x, \quad E_y^a = \frac{1}{2} f''|_{r=0} y. \quad (7.12)$$

This result applies to any rotationally symmetric potential that meets the assumptions.

Thus, substituting V_0 for f_0 gives:

$$E_x = \frac{V_g \cos(\phi_s) \tilde{V}_0''(z)}{2} x, \quad E_y = \frac{V_g \cos(\phi_s) \tilde{V}_0''(z)}{2} y. \quad (7.13)$$

In Sec. 3 the applied ESQ focusing term was derived. Comparing the ESQ focusing term with Eq. 7.13 shows that another focusing term will be introduced from the rapid axial variation of the applied gap field. Following the same steps, a gap focusing function $\kappa_{x,g}(z)$ can be found. The gradient is given by

$$G_g(z) = -\left. \frac{\partial E_x}{\partial x} \right|_{x=0} = -\frac{V_g \cos(\phi_s)}{2} \frac{\partial^2 \tilde{V}_0(z)}{\partial z^2}.$$

Plugging the gradient in to the optic equation gives:

$$\kappa_{x,g}(z) = \frac{qG_{g,x}(z)}{2\mathcal{E}_b(z)} = -\frac{qV_g \cos(\phi_s)}{4\mathcal{E}_b} \frac{\partial^2 \tilde{V}_0(z)}{\partial z^2}.$$

Lastly, expanding \mathcal{E}_b gives the focusing optic for the acceleration gap, we have:

$$\kappa_{x,g} \equiv \kappa_{y,g} \equiv \kappa_g(z) = \frac{\frac{-qV_g \cos(\phi_s)}{4\mathcal{E}_{b,i}} \frac{\partial^2 \tilde{V}_0(z)}{\partial z^2}}{1 + \frac{qV_g \cos(\phi_s)}{\mathcal{E}_{b,i}} \tilde{V}_0(z)}. \quad (7.14)$$

It is useful to define coefficients to collect all the new terms for the thick-gap model.

Using the following definitions:

$$\begin{aligned}
C_{x,2} &= \frac{\frac{V_g \cos \phi_s}{2\mathcal{E}_{b,i}} \tilde{V}'_0(z)}{1 + \frac{V_g \cos \phi_s}{\mathcal{E}_{b,i}} \tilde{V}_0(z)} r'_x, \\
C_{x,2} = \kappa_q(z) + \kappa_g(z) &= \frac{\frac{q}{2\mathcal{E}_{b,i}} G(z) \Big|_{\text{quad}}}{1 + \frac{qV_g \cos(\phi_s)}{\mathcal{E}_{b,i}} \tilde{V}_0(z)} - \frac{\frac{qV_g \cos(\phi_s)}{4\mathcal{E}_{b,i}} \frac{\partial^2 \tilde{V}_0(z)}{\partial z^2} \Big|_{\text{gap}}}{1 + \frac{qV_g \cos(\phi_s)}{\mathcal{E}_{b,i}} \tilde{V}_0(z)}, \\
C_{x,3} = 2Q(z) &= \frac{2Q_i}{1 + \frac{V_g \cos \phi_s}{\mathcal{E}_{b,i}} \tilde{V}_0(z)}, \\
C_{x,4} = \epsilon^2(z) &= \frac{\epsilon_i^2}{1 + \frac{V_g \cos \phi_s}{\mathcal{E}_{b,i}} \tilde{V}_0(z)}. \tag{7.15}
\end{aligned}$$

The coefficients in y are almost identical except that $C_{y,2}$ will have a $-\kappa_q(z)$. The KV envelope equations (Eq. 7.5) can then be written in terms of the thick gap model coefficients as:

$$\begin{aligned}
r''_x + C_{x,1}(z)r'_x + C_{x,2}(z)r_x - \frac{C_{x,3}(z)}{r_x + r_y} - \frac{C_{x,4}(z)}{r_x^3} &= 0, \\
r''_y + C_{y,1}(z)r'_y + C_{y,2}(z)r_x - \frac{C_{y,3}(z)}{r_x + r_y} - \frac{C_{y,4}(z)}{r_y^3} &= 0. \tag{7.16}
\end{aligned}$$

We apply this model for the envelope to numerically solve for the transverse beam evolution in thick gaps with consistent energy gain.

We repeat benchmark tests carried out for the thin gap model with a single thick gap. Integrations of the new thick gap model for the KV envelope equations are compared with 3D Warp simulations. Fig. 7.7 shows r_x and r'_x obtained from integrating Eq. 7.16 (dashed-lined) and extracted from 3D Warp simulations (solid line) in the top row analogously to as

presented in Fig. 7.6. The lower panel shows $\epsilon(z)$ from the model (dashed-line) and from Warp (solid line). Note that space charge is included in both the Warp simulation and KV envelope model. The rapid running thick gap KV model provides excellent agreement. Note that the thick gap model accurately captures the strong gap focusing evident in the 3D Warp simulations while remaining a highly efficient and rapid running model. This opens the door to carrying out rapid optimization runs for improved approaches. Inclusion

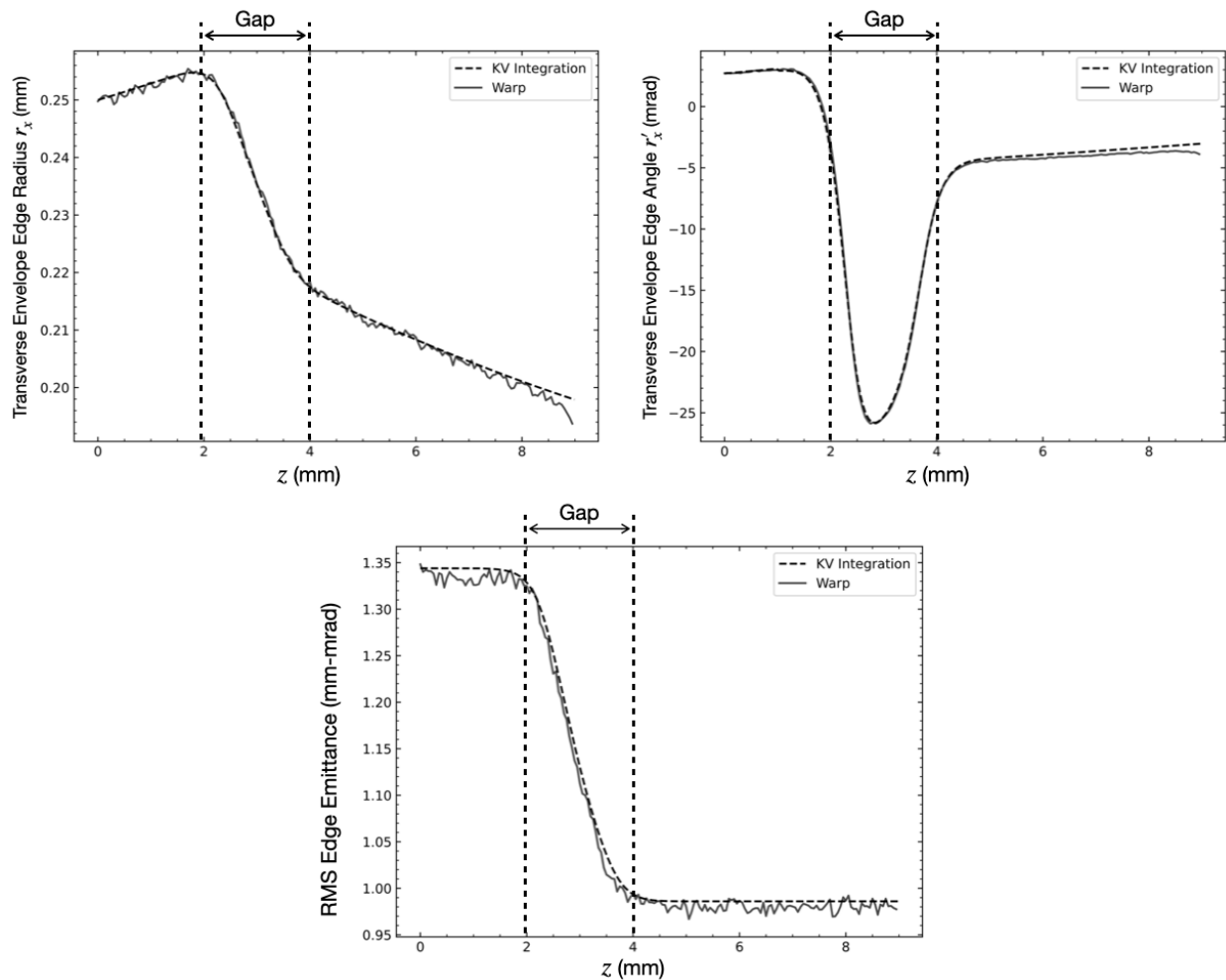


Figure 7.7: Comparisons of the envelope rms-edge evolution envelope model through a single DC acceleration gap with $V_g = 7$ kV. 3D Warp simulations are compared to the thick-gap model. (Top row): On the left is the rms edge envelope radii and on the right the rms-edge envelope angle. (Bottom row): Rms edge emittance $\epsilon(z)$. The thick-gap model results are dashed curves and 3D Warp simulations solid curves. Dashed vertical lines mark the edges of the RF gap.

of envelope parameter variations from strong acceleration together with gap focusing have produced a high fidelity yet efficient model. Recall the discussion on the KV distribution from Ch. 3. There, we discussed the validity in applying the KV envelope equations to a beam that does not have a KV distribution. Although the beam distribution is not KV form, we can form an rms equivalent beam and apply the KV envelope equations as a low order approximation to the real beam evolution so long that the normalized emittances are nearly constant. We check that the normalized rms edge emittance $\epsilon_{n,x}$ is nearly conserved in the Warp simulations. Fig. 7.8 shows $\epsilon_{n,x}$ through a single gap with source conditions and $V_g = 7$ kV. Such strong acceleration is where the conditions for conserved $\epsilon_{n,x}$ will be least likely. As shown in the figure, we have nearly constant $\epsilon_{n,x}$ with fluctuations below 1%. Therefore, we can feel confident that using the KV envelope equations is valid as an approximation to the lab beam.

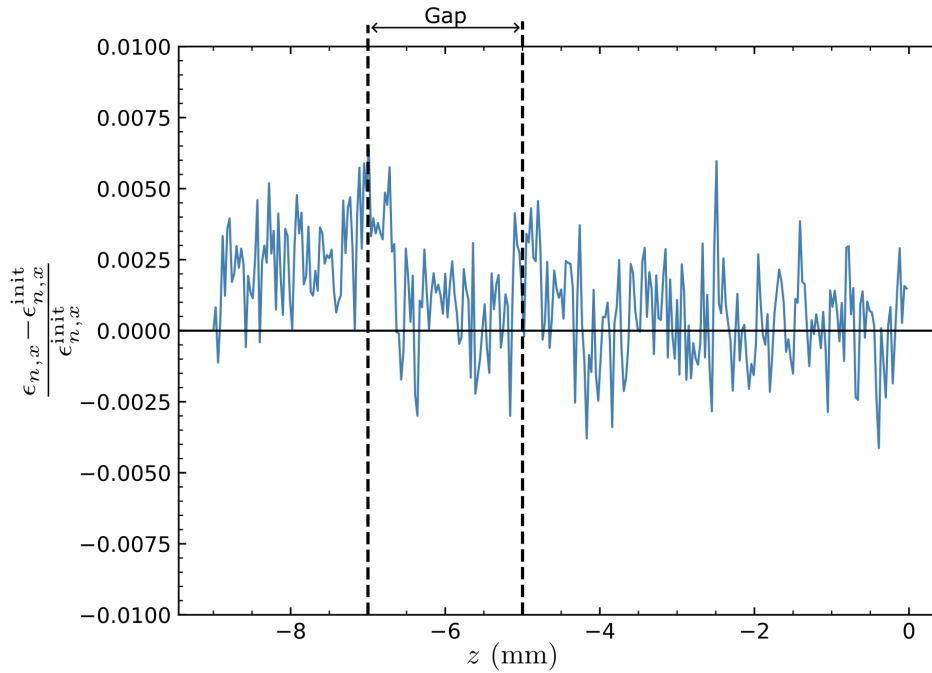


Figure 7.8: Fractional changes in normalized rms edge emittance calculated at each grid-point normalized by the initial value. The gap is at 7 kV with an injected 7 keV beam representing the initial beam conditions from the source and where the acceleration is the strongest. Here, the normalized emittance fluctuates less than 1% from the initial value and can be considered conserved.

7.5 Lattice Design For Compact RMS Edge Envelope

The previous two sections demonstrated an efficient model for capturing the physics of the beam's rms-edge envelope evolution in the presence of strong acceleration in a thick gap and a search algorithm to find a matched condition by providing the required voltage (voltage fixed) or required coordinates (coordinate fixed). With these two tools available, a thorough investigation into the lattice design can now be carried out. In the following sections, we create a point design for a 14-gap system that will accelerate a beam with a kinetic energy of 7 keV out the source to 81 keV. The system can be separated into three parts: 1) a Gap-only acceleration section comprised of only acceleration gaps, 2) a matching section comprised of four ESQs, and 3) a Gap+ESQ section comprised of regular lattice of acceleration gaps (two per period) and ESQ doublet focusing. Due to the many possible design choices the presented option is not proven to be a global optimum. Furthermore, the optimum will depend on the envisioned application where tradeoffs can be made in, for example, beam current vs beam kinetic energy during the design. What constitutes the best case is up to the user and these sections are meant to show how one may tune the lattice and quantify improvement.

7.5.1 The Gap-only Acceleration Section

From Fig. 7.7 we see that the acceleration gaps provide strong axisymmetric focusing in r_x and r_y . This suggests a lattice design that uses only acceleration gaps with no ESQ focusing in the pre-matching acceleration section until the energy rises high enough to where the effect becomes too weak to control beam expansion. The focusing effect from the gap does not continue to high energies as depicted in Fig. 7.9. Here, the r_x and r_y are shown on the

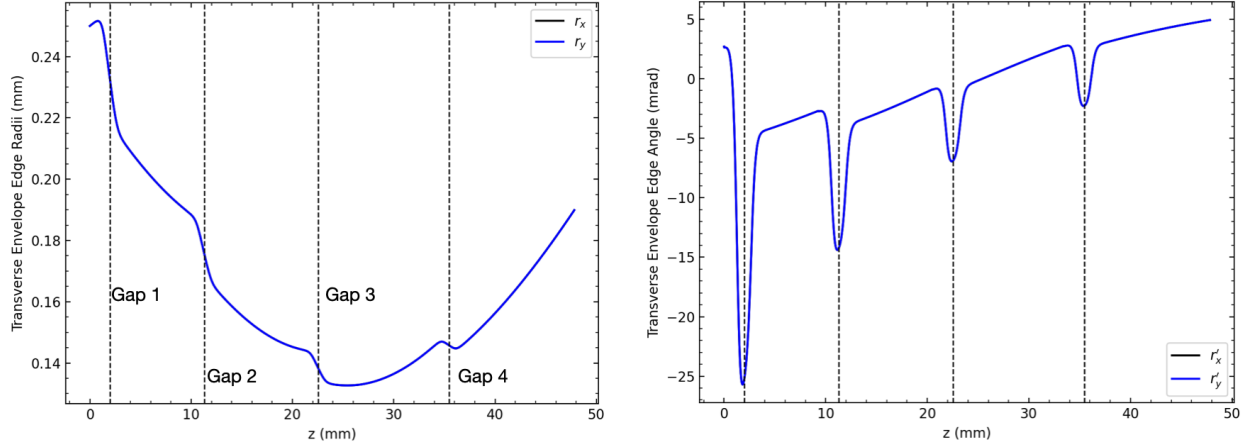


Figure 7.9: Transverse envelope edge radius and angle in x (black) and y (blue) through a six gap lattice at maximum acceleration ($\phi_s = 0$). The gap centers are marked by vertical dashed lines. Due to the injection and gap symmetry ($r_x = r_y$) the blue curve is directly overlaid on the black curve. Transverse focusing from the gap is strong but after about four gaps begins to over focus and the beam begins to expand while the effect becomes weaker. Additionally, the gap acts to decrease the angle of the beam gap-to-gap until the beam expansion after gap 4.

left panel in black and blue respectively, while r'_x and r'_y are shown on the right panel. With symmetric initial conditions from the source, the evolution in r_x is identical to that in r_y and similarly for r'_x and r'_y . Vertically dashed lines denote locations of the axial gap centers which are operating at maximum acceleration ($\phi_s = 0$). We observe the strong focusing effect up to the fourth gap (corresponding to ≈ 31 keV) where the beam begins to expand. This prompts the question: how long can we maximally accelerate until we can expect significant scraping of the beam in the $r_p = 0.55$ mm apertures? To answer this question, the envelope code was used to iteratively calculate the $r_x = r_y$ evolution for the first 16 lattice periods. Recall that each lattice period is comprised of two gaps, so this results in 32 gaps total. For each lattice design, the average r_x over the lattice was calculated, which we denote as \bar{r}_x , and the results for each lattice design is shown in Fig. 7.10. We see that, after about 7 lattice periods (14 gaps and ≈ 100 keV beam energy) $\bar{r}_x > r_p$ and we can expect significant scraping. At 16 lattice periods (32 gaps and about ≈ 200 keV beam energy) $\bar{r}_x \approx 2r_p$ and

there little chance of getting significant beam delivered to target. It appears that 4-6 gaps can be safely used before transitioning to an ESQ matching section to interface to a lattice with ESQ focusing doublets and two RF gaps per period.

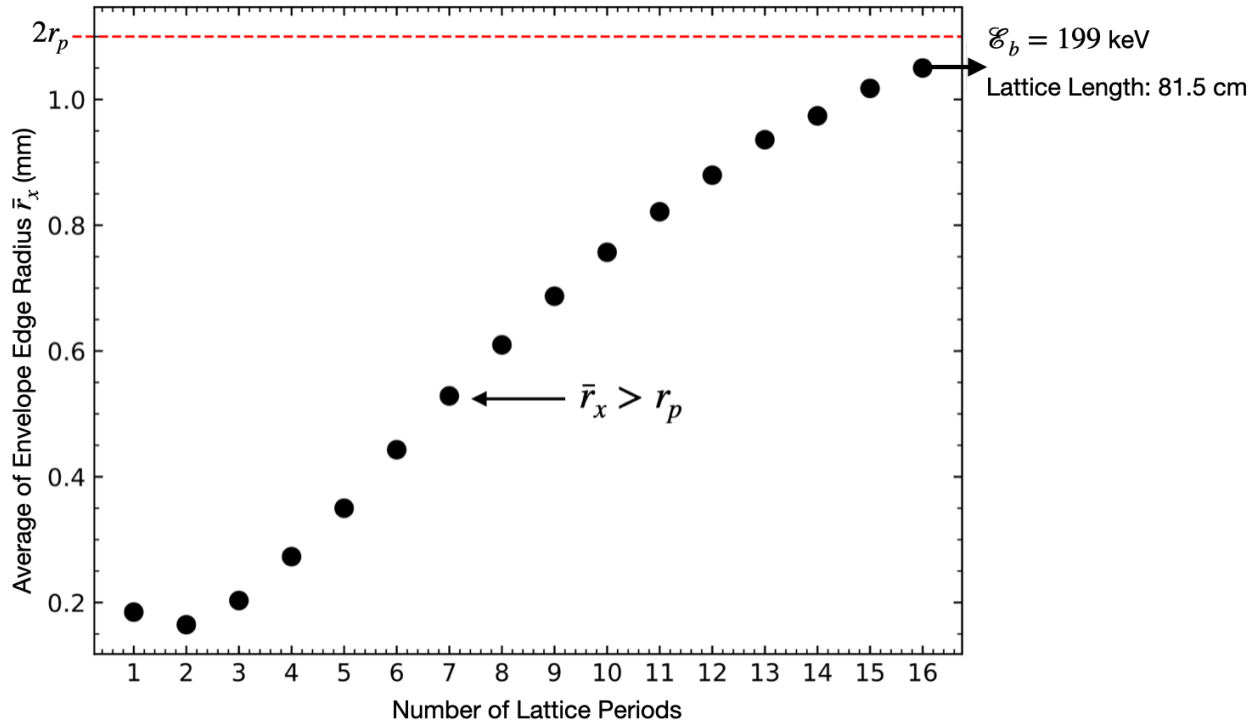


Figure 7.10: Period average \bar{r}_x ($r_x = r_y$) by lattice periods for gap only lattice with maximum acceleration design ($\phi_s = 0$). After 7 lattice periods, $\bar{r}_x > r_p$ and significant scraping is expected.

We seek to find a design to both accelerate, longitudinally focus the bunch, and radially confine beam. The design was found use both the rapid running 1D longitudinal model and transverse envelope model by hand-tuning the parameters to find a working case. Using the 1D longitudinal script, the design phase was chosen to be $\phi_s = (-\pi/2, -\pi/3, -\pi/6, \quad {}^30)$, where the notation ${}^k\phi$ denotes an operating phase of ϕ for k successive gaps. The resulting transverse dynamics are shown in Fig. 7.11. With this design choice, the necessary parameters can be extracted and are shown in Table 7.3 under the Gap-only column. Parameter values Q_f and ϵ_f are necessary to track in order to correctly model the Gap-only beam in the first lattice period of the Gap+ESQ section. Once modeled correctly, the final coordinates of the Gap-only section can be “stitched” to the initial coordinates of the Gap+ESQ section via the matching section.

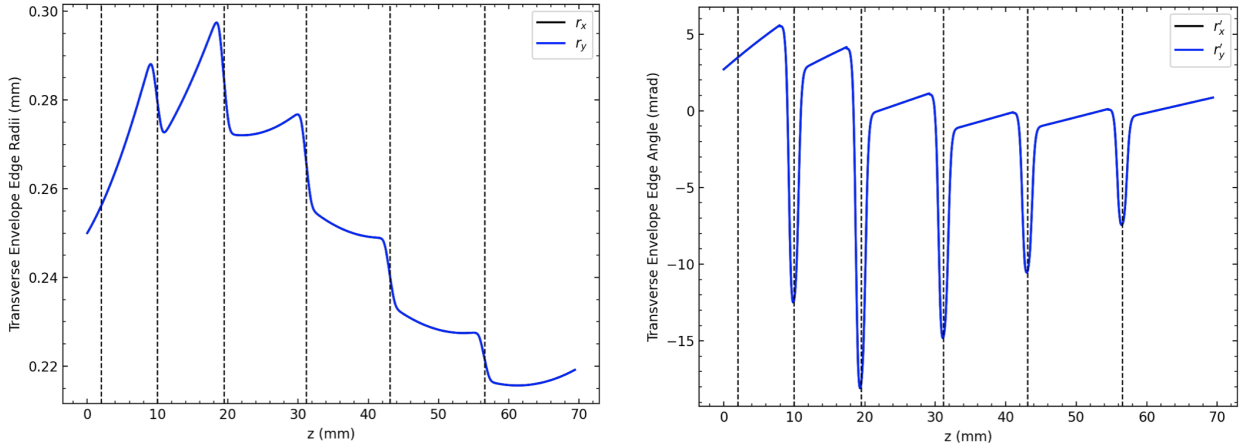


Figure 7.11: Transverse envelope edge radii (left) and angle (right) for x (black) and y (blue). Since the initial conditions are symmetric the two curves are overlaid. The successive gaps are chosen to operate with design phase $\phi_s = (-\pi/2, -\pi/3, -\pi/6, \quad {}^30)$ to provide acceleration, longitudinal focusing, and radial confinement/focusing.

7.5.2 The Regular Acceleration and ESQ Doublet Focusing

(Gap+ESQ) Section

With the Gap-only section after the injector designed, we record the parameter values of \mathcal{E}_b , Q and ϵ of the emerging beam. These parameters are now used to find a match condition for the Gap+ESQ section and the associated beam. There are two approaches using our developed search algorithm. We can either seek the required ESQ biases and provide the desired coordinates (fixed coordinate). In this approach, the ESQ biases will be opposite but not equal (asymmetric voltages). Or, we can seek the coordinates and provide the desired ESQ biases (voltage fixed). We chose the latter approach for multiple reasons. This choice allows us to pick equal and opposite ESQ voltages (ESQ doublet focusing). Also, the acceleration in the system is still non perturbative for many more lattice periods, and thus perfectly matching successive lattice periods would be complicated. The continuous ESQ voltage parameter space yields an infinite set of matched solutions. We therefore must establish a metric to determine what solution is “good.” It is not immediately obvious what this metric should be, only that we should remain with the safe operating limit of the ESQ voltage half breakdown limit (see Sec.5.4). For a metric, one could use:

- Flutter: $\frac{\text{Max}[r_x] - \text{Min}[r_x]}{\langle r_x \rangle}$; minimal better,
- Maximum rms-edge angle: $\text{Max}[r_x]$; minimal in aperture,
- Phase advance per lattice period: σ ; smaller better,
- Average rms-edge angle: \bar{r}_x ; smaller within aperture.

We decided to use the average rms-edge angle \bar{r}_x . To do so, an ESQ bias voltage is set and the KV equations integrated. Then \bar{r}_x and \bar{r}_y are calculated from the envelope evolution.

We tuned ESQ bias voltages to give $\bar{r}_x \approx \bar{r}_y \approx 0.28$ mm. This method could be automated but is easy enough to quickly iterate by hand.

Fig. 7.12 shows match solution for the first period after the four ESQ matching section in the Gap+ESQ section. The solution satisfies $\bar{r}_x \approx \bar{r}_y \approx 0.28$ mm. In this period, the gaps are operating at maximum acceleration ($\phi_s = 0$). ESQ axial extents centers are marked by black squares at the top and a vertical dotted line marks the ESQ axial centers. On the left is the first focusing ESQ with bias $V_1 = 193$ V, and the second ESQ has a symmetric $V_2 = -V_1$ bias. Initial conditions are marked by stars at the end of the period to show that the final coordinates match the initial coordinates. We extract the match coordinates (Table 7.3 under Gap+ESQ) and use the matching section to shape the beam so that the final coordinates of the Gap-only section equal the initial coordinates of the Gap+ESQ section. Note that the matched beam radial excursion within the $r_p = 0.55$ mm apertures are very modest. Period variations of r_x and r_y are also small. This is characteristic of very low focusing strength which should lead to enhanced stability and insensitive to residual mismatch in subsequent periods. Estimating the phase advance in the x -plane σ_x with the accelerating beam using

$$\sigma_x = \int_{\text{period}} \frac{\epsilon_x}{r_x^2} dz, \quad (7.17)$$

and similar expression for σ_y , gives $\sigma_x \approx \sigma_y \approx 13^\circ/\text{period}$. This is exceptionally low compared to typical strong focusing systems where $\sigma \sim 60^\circ - 120^\circ$ per period is common.

7.5.3 The Matching Section

Using the values in Table 7.3, the search algorithm can be used to match the beam emerging from the Gap-only section to the Gap+ESQ section by finding the necessary voltages. In

Parameter	Gap-Only	ESQ+Doublet
$r_{x,f}$	0.219 mm	0.28 mm
$r_{y,f}$	0.219 mm	0.30 mm
$r'_{x,f}$	0.859 mrad	1.91 mrad
$r'_{y,f}$	0.859 mrad	-0.36 mrad
Q_f	1.472×10^{-5}	1.081×10^{-5}
ϵ_f	0.617 mm-mrad	0.529 mm-mrad

Table 7.3: Final parameter values for the Gap-only and Gap+ESQ sections. The matching section will have to “stitch” \vec{r}_f from the Gap-only section to the \vec{r}_i from the Gap+ESQ section. Q_f and ϵ_f from the pre-match section are used to correctly model the beam in the Gap+ESQ section before the two acceleration gaps and find match coordinates listed using the search algorithm.

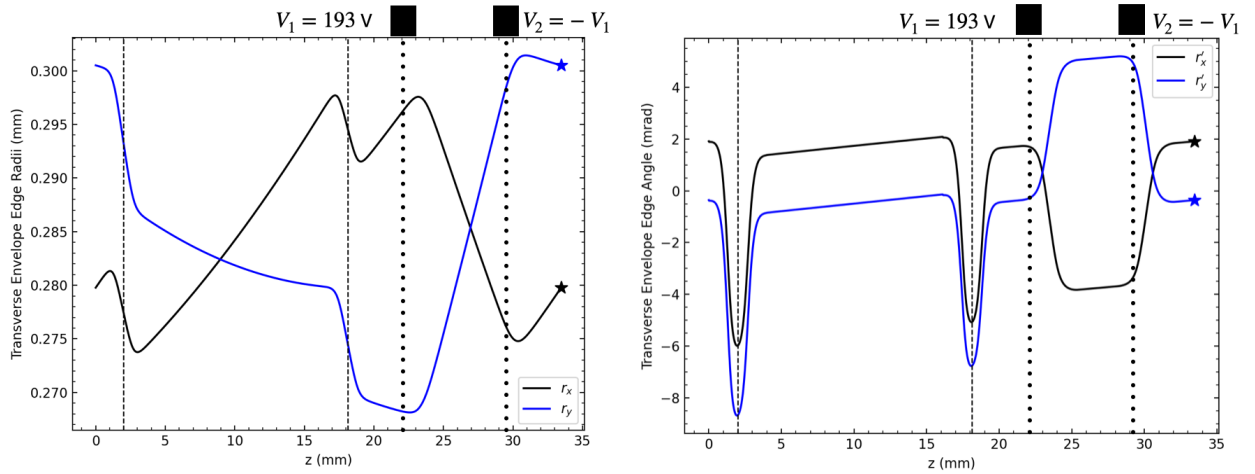


Figure 7.12: A matched solution for the first lattice period after the matching section in the Gap+ESQ section. Acceleration gaps operating a $\phi_s = 0$ with $V_g = 6$ kV are shown as vertical dashed lines act to increase the beam energy from 33 keV to 55 keV at the end of the period. ESQs centers are shown as vertical dotted lines with a black rectangle at the top showing axial extents and along with the applied bias voltage. Matched parameters from the exit of the Gap-only section are used as the initial parameters in this Gap+ESQ section. The voltage setting was then iterated until $\bar{r}_x \approx \bar{r}_y \approx 0.28$ mm was achieved to keep the beam well confined within the aperture. Stars mark the initial envelope coordinate values to show that the match condition is reached and it is also clear that the final coordinates are coincident with the marked initial values.

favor of compactness, while having full adjustability (four voltages to match four parameters), we chose a matching section of four ESQs. Sec.5.4.2 gave an approximate 3 mm separation distance between ESQ ends so that fringe fields did not overlap. Additional 3 mm of space

was added to be safe for a total of 6 mm spacing between successive ESQ centers and from the grounded acceleration planes on either end of the matching section. This distance was used to set a minimum bound on the placement. Additionally, the distance between the final acceleration gap in the Gap-only section and the initial gap in the Gap+ESQ section must be spaced in order to maintain RF synchronism. Therefore, if the final and initial gap are operating at the same ϕ_s , the distance must be $\beta_b \lambda_{\text{rf}}(n + 1/2)$ with integer $n = 1, 2, \dots$, in order to maintain synchronism. Ensuring that the gaps maintain RF synchronism along with sufficient spacing between neighboring ESQs and the grounded gap planes requires a matching section length of approximate 60 mm which corresponds to $n = 4$ for the resonance condition. The envelop in the matching section is shown in Fig. 7.13 on the bottom row. Since there are only ESQs present in this section, the dotted lines mark the ESQ centers. The top image shows $\kappa(z)$ with the corresponding bias for matching. Note that the matched envelop excursions are modest within $r_p = 0.55$ mm ESQ apertures and ESQ bias voltages are modest in spite of the higher beam energy after the initial six gaps.

7.5.4 Gap+ESQ Lattice

With the Gap-only section and first lattice period of the Gap+ESQ section successfully interfaced via the matching section, we proceed to tune successive lattice periods in the Gap+ESQ section after the initial lattice period directly after the matching section. We consider two approaches:

1. Match neighboring lattice periods one-by-one successively for the entire lattice by providing the match coordinates and searching for the required ESQ bias voltages to maintain sufficient (nearly fixed) radial confinement with energy gain,

- Examine single lattice periods and adjust the ESQ bias voltage for the beam parameters arriving at that lattice period to find a match condition using a metric as a guide.

Due to the same reasons before, we proceed with the second approach and continue using \bar{r}_x and \bar{r}_y as a metric. Also, because the doublet lattice has exceptionally low phase advance with $\sigma_x \approx \sigma_y \approx 13^\circ/\text{period}$, we expect the lattice to be insensitive to slight mismatch that will occur in the following periods under this approach. This is later verified. Table 7.4 gives parameter values for the first four lattice periods in the Gap+ESQ section following

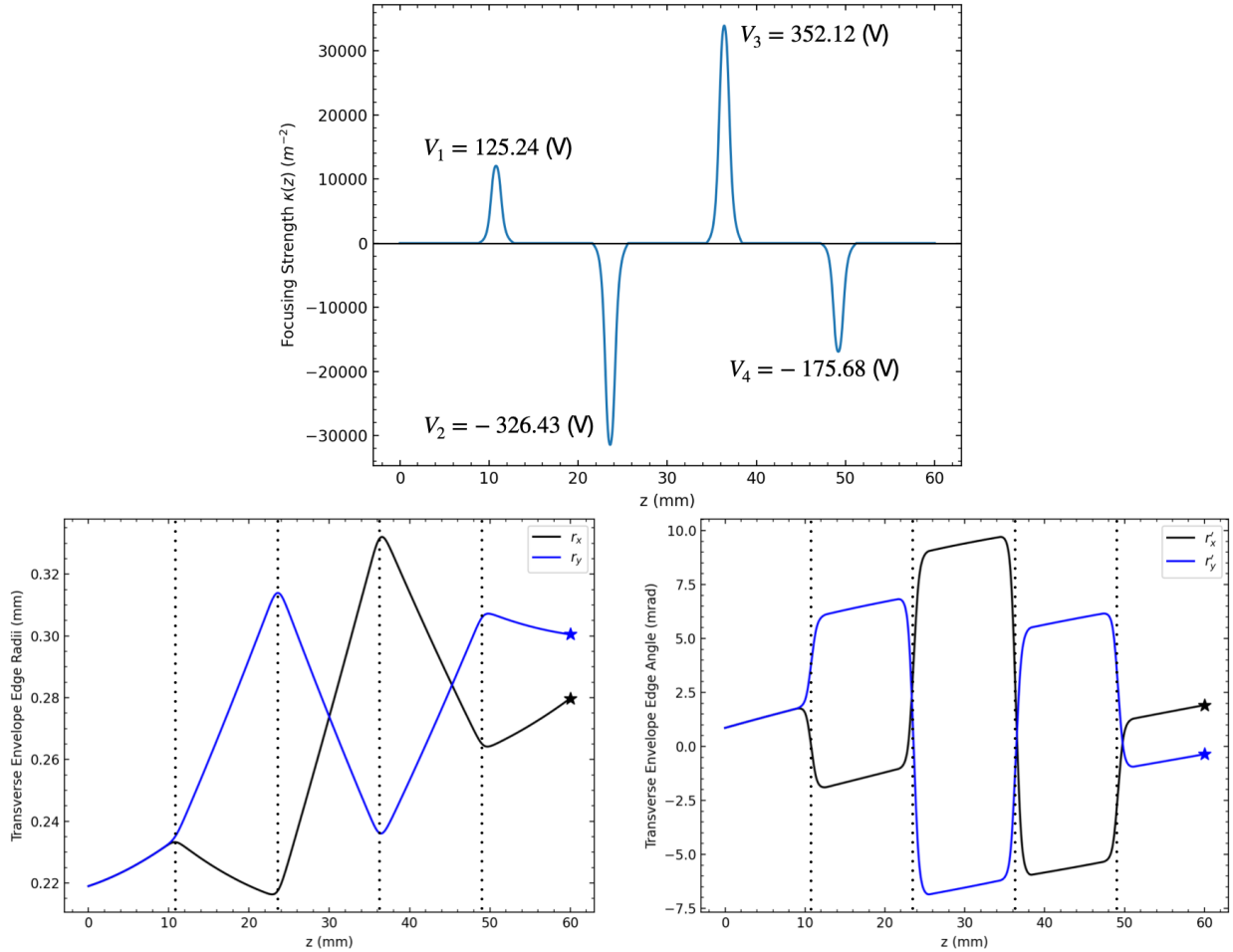


Figure 7.13: Transverse envelope edge radii (left) and angle (right) for x (black) and y (blue) for the matching section where there is no acceleration (bottom row). The four ESQ centers are shown as vertical dotted lines. Star markers are placed at the desired initial coordinates for the post matching section. The top figure shows the focusing strength $\kappa(z)$ along the matching section and each ESQ voltage is labeled.

the matching section. Each row represents a lattice period starting in the Gap+ESQ section. Using values from the previous lattice period \mathcal{E}_f , Q_f , and ϵ_f the beam can be modeled in the current lattice period and the ESQ biases are adjusted to keep $\bar{r}_x \approx \bar{r}_y \approx 0.25$ mm, corresponding to a well confined beam with the $r_p = 0.55$ mm aperture. Each lattice period consists of two gaps operating at max acceleration. The transverse evolution of the beam for this tuned lattice is shown in Fig. 7.14. The beam envelope expands slightly which can be seen by comparing the terminal points of the curves with the star markers placed at the initial conditions. The stronger expansion in y is due to the mismatch being exacerbated as the beam is firstly defocused in y . This expansion can be possibly countered by reversing the doublet polarity so that the first ESQ in the doublet is focusing in y . Although there is mismatch, we observe that expectation of low lattice sensitivity to mismatch due to low σ is reasonable. This procedure can be repeated to get up to higher energies.

We calculated additional matched envelope data using the envelope code and exported to a csv file to scrutinize focusing properties with energy gain. Fig. 7.15 is a snap shot of part of the data where, due to the numerous columns, multiple snapshots had to be taken.

r_x (mm)	r_y (mm)	r'_x (mrad)	r'_y (mrad)	\mathcal{E}_i (keV)	\mathcal{E}_f (keV)	Q_i	ϵ_i (mm-mrad)	V_1 (V)
0.2798	0.3005	1.9114	-0.3615	33.22	45.232	1.4729×10^{-5}	0.6169	193
0.2745	0.3064	1.780	-1.1752	45.232	57.244	1.0811×10^{-5}	0.5287	310
0.270	0.3056	1.581	-1.398	57.244	69.256	0.8542×10^{-5}	0.5484	375
0.271	0.308	1.431	-1.437	69.256	81.268	0.706×10^{-5}	0.4986	415

Table 7.4: Matched envelope parameter values for tuned regular lattice periods in the Gap+ESQ section that follows the matching section. The tuning was done for each lattice period (each row) by extracting the necessary parameters \mathcal{E}_f , Q_f and ϵ_f from the previous lattice period to model the beam in the current lattice period and tune. The doublet voltage biases are symmetric with $V_2 = -V_1$ with amplitudes hand tuned so that $\bar{r}_x \approx \bar{r}_y \approx 0.28$ mm to maintain nearly constant fill. Note that only slight changes in initial matched envelope values occur. These variation will generate small mismatches between successive lattice periods. However, due to low phase advance, we expect the lattice to be relatively insensitive and the beam to remain well confined.

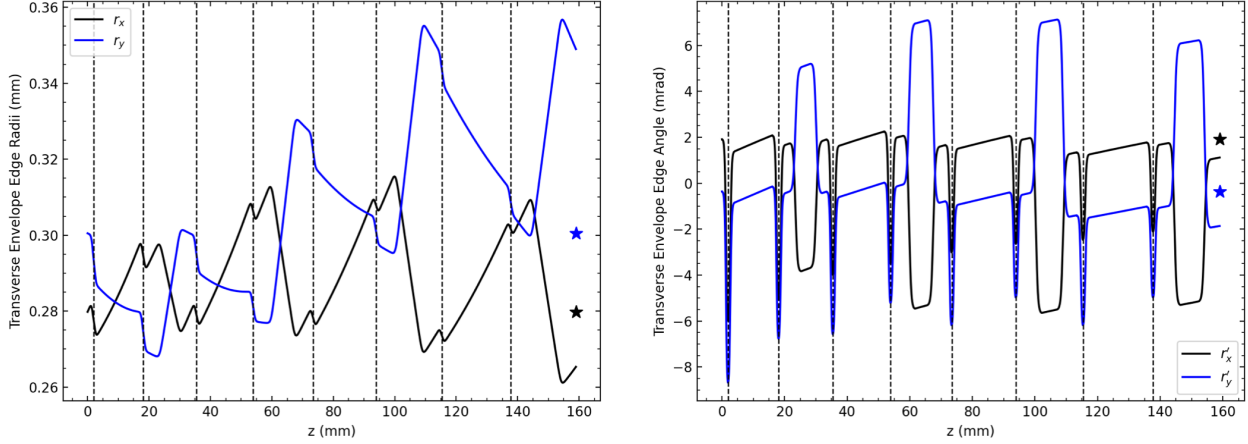


Figure 7.14: Transverse envelope edge radii (left) and angle (right) for x (black) and y (blue) for a regular lattice periods in the Gap+ESQ section following the matching section. The acceleration gap centers are marked by vertical dashed lines. All acceleration gaps are spaced for $\phi_s = 0$. After every two gaps, there is an ESQ doublet to provide focusing (not shown in the graphic). Each period (two gaps) was tuned using the ESQ bias voltages to keep $\bar{r}_x \approx \bar{r}_y \approx 0.28$ mm. Star markers are placed at the initial conditions. Due to low lattice phase advance σ , the beam is well confined despite slight mismatch between lattice periods.

This data was collected by performing the same tuning procedure and adjusting ESQ bias voltages to keep $\bar{r}_x \approx \bar{r}_y \approx 0.28$ mm. When the focusing voltage approaching the half-breakdown limit, another “stack” of ESQ wafers was added to increase focusing strength and decrease bias voltage requirements. This tuning was done up to 1 MeV, only a subset of data is shown in Fig. 7.15. A summary of ESQ focusing bias voltage scaling for higher beam kinetic energies is shown in Fig. 7.16. A red vertically dotted line shows the half breakdown limit and we can see that for beam energies above 200 keV, we will need to start incorporating ESQ stacks to increase the focusing provided while maintaining biases well below the breakdown limit. The data collected goes up to 1 MeV using up to five stacks of wafers. The difficulty in deciding that constitutes a good metric is immediately apparent. Multiple parameters can be calculated that are equally informative. Future studies may benefit approaches by finding the best metrics for guiding lattice design choices. Finally, it should be noted, this presentation is predicated on a fixed RF frequency. Due to this, the gap

max rx (mm)	min rx (mm)	max ry (mm)	min ry (mm)	max rxp (mrad)	min rxp (mrad)	max ryp (mrad)	min ryp (mrad)					
0.2965	0.2768	0.2988	0.2727	1.8727	-7.0230	4.6880	-8.9999					
0.3001	0.2707	0.3032	0.2660	2.0215	-5.2032	6.2610	-7.6829					
0.3028	0.2674	0.3056	0.2641	1.9322	-5.6454	6.4259	-6.6383					
0.3039	rx (mm)	ry (mm)	rxp (mrad)	ryp (mrad)	avg. rx (mm)	avg. ry (mm)	avg. rxp (mrad)	avg. ryp (mrad)				
0.3046	0.2845	0.2988	1.8202	0.1566	0.2865	0.2848	-0.0813	0.1110				
0.3061	0.2762	0.3007	1.6889	-0.7237	0.2845	0.2832	-0.1269	0.1343				
0.3082	0.2730	0.3014	1.4927	-1.0208	0.2838	0.2831	-0.1264	0.1246				
0.3104	0.2706	0.3010	1.3390	-1.1489	0.2829	0.2824	-0.1202	0.1140				
0.3107	0.2691	0.3006	1.2180	-1.1978	0.2823	0.2820	-0.1127	0.1053				
0.3114	0.2693	0.3014	1.1211	-1.1997	0.2830	0.2828	-0.1005	0.0930				
0.3102	0.2704	0.3028	1.0445	-1.1855	0.2845	0.2844	-0.0937	0.0856				
0.3058	0.2720	0.3046	0.9803	-1.1595	0.2864	0.2864	-0.0863	0.0785				
0.3133	0.2720	0.3046	0.9226	-1.1308	0.2865	0.2866	-0.0803	0.0730				
0.3108	0.2719	init E (keV)	final E (keV)	Lattice Period (mm)	V1 (V)	V2 (V)	Nstack	Vg (kV)	init Q	init emit (mm-mrad)	sigmax (deg/per)	sigmay (deg/per)
0.3111	0.2701	30.0	42.0120	32.1530	150.0	-150.0	1	6.042	1.629E-05	0.649	12.9461	13.1051
0.3132	0.2648	40.0	52.0121	36.0442	275.0	-275.0	1	6.042	1.222E-05	0.562	13.1166	13.2421
0.3100	0.2726	50.0	62.0120	39.5530	345.0	-345.0	1	6.042	9.76E-06	0.503	13.1757	13.2500
0.3139	0.2697	60.0	72.0121	42.7743	400.0	-400.0	1	6.042	8.13E-06	0.459	13.2539	13.3072
0.3133	0.2697	70.0	82.0121	45.7690	445.0	-445.0	1	6.042	6.97E-06	0.425	13.3049	13.3431
0.3142	0.2715	80.0	92.0121	48.5792	480.0	-480.0	1	6.042	6.1E-06	0.398	13.2519	13.2712
	0.2676	90.0	102.0121	51.2353	510.0	-510.0	1	6.042	5.42E-06	0.375	13.0978	13.1136
	0.2717	100.0	112.0120	53.7603	535.0	-535.0	1	6.042	4.88E-06	0.356	12.9308	12.9358
	0.2708	110.0	122.0121	56.1717	560.0	-560.0	1	6.042	4.4E-06	0.339	12.9016	12.8963
	0.2717	120.0	132.0121	58.4838	590.0	-590.0	1	6.042	4.07E-06	0.325	12.8995	12.9019
		130.0	142.0121	60.7078	620.0	-620.0	1	6.042	3.76E-06	0.312	13.0302	13.0312
		130.0	142.0121	60.7078	420.0	-420.0	2	6.042	3.76E-06	0.312	13.4930	13.4938
		140.0	152.0121	62.8531	420.0	-420.0	2	6.042	3.49E-06	0.3	12.7429	12.7425
		150.0	162.0121	64.9276	440.0	-440.0	2	6.042	3.25E-06	0.29	12.9970	12.9917
		160.0	172.0121	66.9378	455.0	-455.0	2	6.042	3.05E-06	0.281	12.9982	12.9954
		180.0	192.0120	70.7870	480.0	-480.0	2	6.042	2.74E-06	0.265	12.8296	12.8204
		200.0	212.0121	74.4375	515.0	-515.0	2	6.042	2.47E-06	0.251	13.1445	13.1397
		220.0	232.0121	77.9170	530.0	-530.0	2	6.042	2.25E-06	0.239	12.7586	12.7509
		240.0	252.0121	81.2477	555.0	-555.0	2	6.042	2.06E-06	0.229	12.8393	12.8317
		260.0	272.0121	84.4471	575.0	-575.0	2	6.042	1.9E-06	0.22	12.7578	12.7525

Figure 7.15: An example of the various parameter calculations that can be saved in a csv file to efficiently evaluate focusing performance when using the envelope code. The tuning method discussed in this section was used to hand-tune lattice periods until an energy of 1 MeV was reached using approximately 6 kV gap voltages. Each line represents a period of the initial energy specified.

to gap separation continues to increase with beam kinetic energy gain allowing more space for ESQ focusing stacks to stay below the breakdown limit. If frequency is “jumped” in integer multiples to keep the system compact at higher kinetic energies, the gap to gap spacing ($\beta_b \lambda_{rf}/2$ for $\phi_s = 0$ max acceleration or slightly shifted for required ϕ_s) will shrink along with the available space for the ESQ doublets. This additional optimization would need to be accounted for in longer, high energy systems. More space can also be generated between gaps by spacing for additional RF wavelengths with gap to gap spacings ($\beta_b \lambda_{rf}(n + 1/2)$ with integer n). However, this will add space which presenting a design tradeoff between a compact system and sufficient doublet focusing.

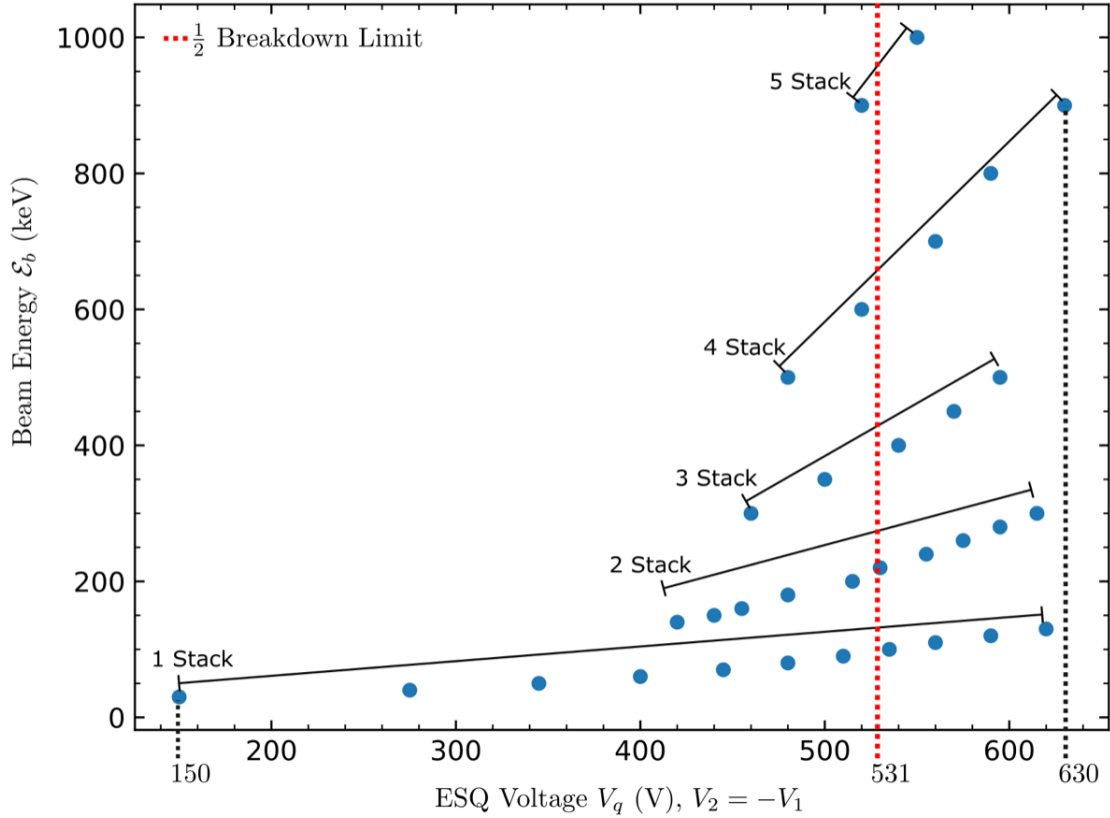


Figure 7.16: ESQ bias voltage scaling for increasing ESQ stacks to provide more focusing. As the beam kinetic energy increases, the required ESQ bias voltage will need to increase to maintain radial confinement of the beam while also staying safely below the breakdown limit. The ESQ focusing provided can be increased while limiting bias voltages by making the ESQ length longer. Instead of machining an ESQ with a different length, the system design allows ESQ wafers to be stacked to increase the ESQ length in discretized “stack” steps of the fundamental length. The red vertical dashed line marks the half breakdown bias. Results summarized provide a guide for ESQ lengths needed to maintain radial beam confinement at higher kinetic energies up to 1 MeV. Lastly, the black vertical dashed line marks the highest ESQ voltage bias used for a four stack doublet at 630 V.

7.6 Summary and Concluding Remarks

This chapter detailed the development of a transverse envelope model to efficiently model the evolution of the transverse beam envelope through a lattice comprised of accelerating gaps and ESQs. Inspired by the 1D longitudinal model, the near-axis gradient of ESQ arrays was extracted from 3D Warp field solves of the arrays and used to build a focusing function

consistent with short ESQ array fields. We benchmarked the improved KV model against 3D Warp simulations in a gap-free design where only the ESQ effects were compared. The agreement was excellent as depicted in Fig. 7.3. We then used a Mathematica code model for periods without acceleration [13] to guide construction of a search algorithm that would find matched conditions for a considered lattice design. A Nelder-Mead algorithm [36] was applied to search and allows the user to find a match condition in one of two ways. One way is to hold the voltages fixed and the algorithm will find the 4D coordinates (transverse position and angle) that will match to the provided $\kappa(z)$. The second way is to provide the 4D coordinate and the algorithm will find the necessary voltages. We then augmented the envelop model to incorporate the focusing effects from the acceleration gaps, by using a typical thin gap model where the coordinates r_x and r_y were unaffected after the gap, but the parameters r'_x , r'_y , Q , and ϵ would all receive a “kick” and change values after the gap. Benchmarks with 3D Warp simulations showed this thin gap model performed poorly forcing a reformulation to developing a thick gap model leading to Eq. 7.16 which takes into account strong acceleration and gap focusing. Benchmarks with 3D Warp simulations showed excellent agreement as illustrated in Fig. 7.7. This resulted in a viable model to efficiently guide optimization of the system.

With established confidence in our abilities to model the transverse effects from ESQ focusing and the acceleration gaps along with matching algorithm, we pressed on to develop lattice design concepts for a compact envelope. We found that there is strong transverse focusing provided by the acceleration gaps up to a certain point before the beam begins to expand due to over focusing. Gap only focusing was found to be sufficiently strong so the beam only expands with $\bar{r}_x > r_p$ and $\bar{r}_y > r_p$ at around seven lattice periods ($N_g = 14$) as shown in Fig. 7.10. Strong focusing from the gaps inspired a system design where the

first six gaps are acceleration gaps only, followed by a four ESQ matching section, and then subsequent lattice periods with two acceleration gaps and ESQ doublet focusing up to target. The 1D longitudinal model developed in Ch. 6 was used to find a candidate design that would provide longitudinal bunching and acceleration in the first gap only section of the lattice (Fig. 7.11). The matching algorithm was then applied to match the emerging beam from the Gap-only to the first lattice period of the downstream Gap+ESQ section comprised of two acceleration gaps at maximum acceleration and ESQ doublet transverse focusing (Fig 7.13). The lattice was optimized to keep $\bar{r}_x \approx \bar{r}_y \approx 0.28$ mm and resulted in the ESQ bias voltage values listed in Table 7.4. This approach necessarily leads to a slight mismatch period-to-period since the acceleration lattice is aperiodic due to the strong acceleration in the system. However, we find that the lattice is relatively insensitive to mismatch due to the small phase advance with modest envelope excursion within the 0.55 mm aperture.

The envelope model could be improved by including time-dependence and the finite pulse length. Recall that to benchmark the model, a static acceleration gap was used. This was justified since the RF-wavelength is much greater than the portion of beam we wish to model, i.e., we sought to model the physics of a portion of the beam rather than the entire bunched beam evolving from the DC injection. It is not obvious how to best include time-dependence. One approach may be to first use the 1D script to advance a particle through a single gap with the design choice of ϕ_s . For this design, the potential profile $\tilde{V}_0(z)$ can be calculated. Note the coefficients in Eq.7.16 are functions of z through $\tilde{V}_0(z)$ and its derivatives only (see Eq. 7.15). Once $\tilde{V}_0(z)$ and its corresponding derivatives are computed, they can be used to integrate Eq. 7.16. This will only represent a limited fraction of the overall beam since this was done for a single ϕ_s . The process can be computed for a range of RF gap phases to

represent a pulse of a continuously injected beam and then combined to give a time-varying profile of the rms-edge envelope. Another difficulty comes from how one would apply such a model and testing this approach against 3D Warp simulations. Such a model would require substantial further development beyond the scope of this dissertation. In our case, we apply full 3D Warp simulations from the source to target to better evaluate expected performance improvements. This is the topic of Ch. 8.

Chapter 8. 3D Self-Consistent Particle-In-Cell Simulations of RF-MEMS

In this chapter, we present 3D Warp simulations of fully systems to better evaluate integrated system performance of optimizations to better evaluate integrated system performance of optimizations found with the reduced physics models. These simulations include fully self consistent space charge with array conductor structures loaded on the mesh. Warp is an extensive code and is briefly overviewed in Ch. 4.

8.1 Aside: Panel Plots and Simulation Snapshots

Before detailing various parts of the Warp simulations, it is beneficial to thoroughly explain Fig. 8.1 since similar plots will be used frequently without extensive labeling. Warp can output snapshot particle-plots with various projections to a computer graphics metafile (cgm). These cgms are collected and can be paged through to visualize the evolution of the beam. We created a separate tool to extract these frames and create an mp4 movie of the evolution. This video file can be saved for later viewing and screenshots can be taken of the video. A typical example can be seen in Fig. 8.1. The panels are numbered starting at the top left corner with Panel 1 and progressing to the right, column by column in a three column plot, before continuing at the left column again on the second row at Panel 4. Each panel indicates a contour plot for the electric field $E_z(x, y = 0, z, t)$ and each employs the same scale in the color bars. Although the x -axis range is the same for each plot (and will be that way for all subsequent panel plots of this kind), the z -axis changes since the simulation window moves

with the design particle (red dot in the plots). Additionally, we used this small two-gap simulations to highlight various objects that are in the simulation, but not are normally not visible. In Panel 1, the injection boundary is shown at $z = 0$ ranging from $\pm r_p$. The horizontal lines mark an absorption boundary that acts as a “scraper” in the simulation—particles with $r \geq r_p$ are considered lost. Acceleration gaps are labeled G_n with G_1 being shown in Panel 1. All panel plots will have conductors labeled G_n for gaps, MQ_n for a matching ESQ, and Q_n for ESQs in lattice periods with $n = 1, 2, 3$ and 4 . Panel 2 and Panel 3 show the injected beam. The injection is continuous for a whole RF period and the full L_b is shown in Panel 3 where $t \approx \tau_{rf}$. Panel 3 shows where a numerical beam position monitor (BPM) for particle diagnostics is placed in the simulations at the midway point between G_2 and G_1 . Additional numerical BPMs, which are not labeled to avoid confusion, are placed right at the injection point, and right before the final Faraday cup (Fcup) diagnostic. Lastly, the Fcup is labeled in Panel 5. Particles are absorbed (Panel 6) and the simulation continues to run for a specified time after the design particle reaches the Fcup.

The next several sections will detail different parts of integrated 3D Warp simulations and make frequent reference to plots analogous to Fig. 8.1. Subsequent sections will use plots of identical layout, but with reduced labeling.

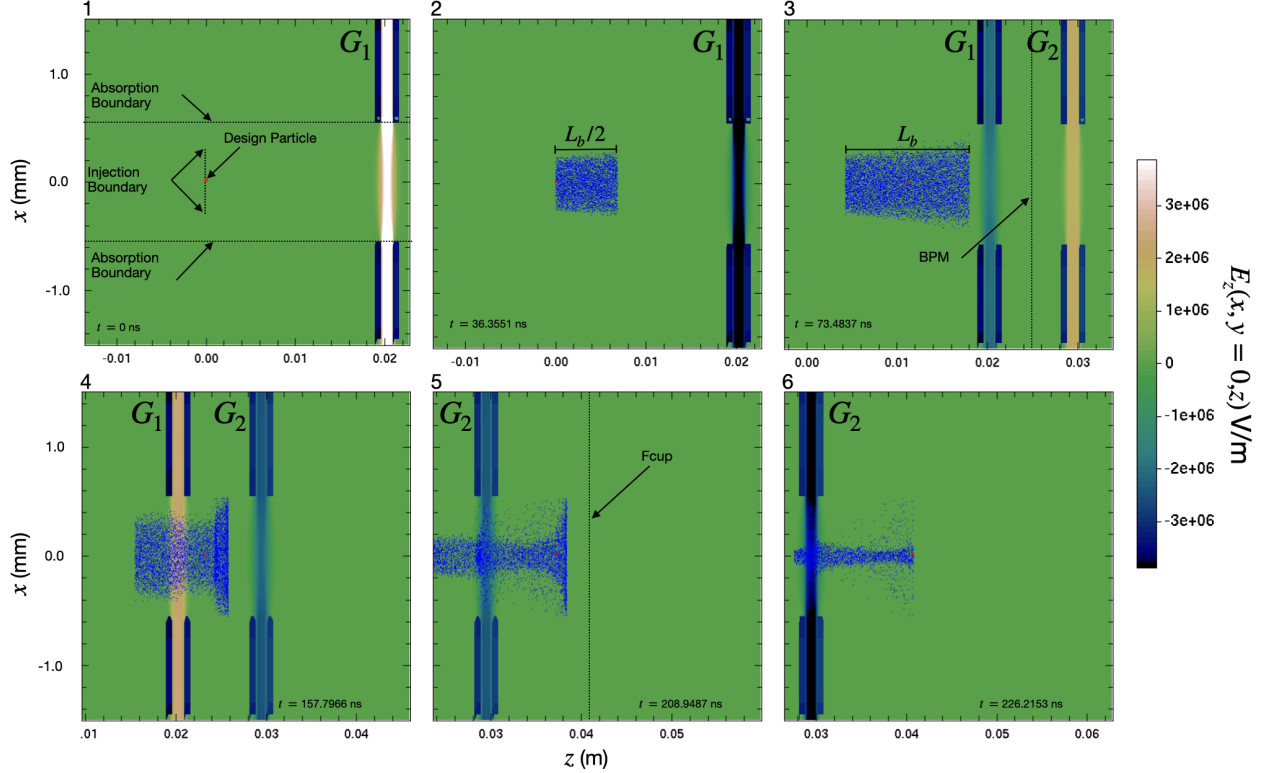


Figure 8.1: Self consistent 3D Warp simulation of a single RF period beam injected into a four RF gap system. Labeled snapshots of the beam distribution in x and z from a continuous injection simulation at different times labeled from 1 to 6 starting at the top left. Panel 1 is at $t = 0$. The design particle is shown as a red dot that starts at the injection plane but is “disabled” until half the beam has been injected (Panel 2). A radial absorption boundary is placed at $r = 0.55$ mm and is labeled in Panel 1 with horizontal dashed lines. Acceleration gaps are labeled G_n and Panel 1 shows G_1 which starts at a maximum positive z -field indicated by the color bar to the right of the panels. Panel 3 shows the full length of the beam, and the design particle begins to actively advance. Note the change in the z values as the simulation window is moving about the design particle as it gains in kinetic energy. Between G_1 and G_2 is a numerical BPM monitor that will record particle parameters. Lastly, Panel 5 shows the location of the Fcup and the particles are being absorbed at the Fcup as seen in Panel 6.

8.2 Beam Creation and Loading

A uniform density cylindrical beam with circular cross-section, flat longitudinal edges, and a Gaussian distributed thermal spread is injected (see Panel 2 of Fig 8.1). Source parameters are given in Table 7.1. We represent a continuous DC beam injection by injecting each time

step till a pulse duration of one RF period (τ_{rf}) is injected. This corresponds to an axial beam length of $L_b = \beta_b c \tau_{\text{rf}}$. The injection is done symmetrically about the design particle such that there is a beam of particles of length $L_b/2$ ahead and behind the design particle. Using the Warp method `add_uniform_cylinder`, the beam loading implemented with inputs shown in Listing 8.1. We first determine how many macro-particles to put in the simulation ($N_{p,\text{max}}$), then the first argument `np=Np_inject` is calculated as $N_{p,\text{inject}} = N_{p,\text{max}} dt / L_b$ where dt is the simulation time step. This will inject (integer rounded) $N_{p,\text{inject}}$ macro particles every time step during injection. The radius of the injected beam is `rmax`. Each time step injects an axial slice of the beam of axial length `[zmin, zmax]`. Here, `zmin=0` is the emission plane in the simulation (i.e., `wp.top.zinject[0] = 0`). The physical length of the injected slice `zmax`, which is calculated using the initial design velocity and simulation time step. This setting ensures that there is no overlap of injected particles. Arguments `vthx`, `vthy`, `vthz` specify the Gaussian distributed thermal velocity spread to apply (see Table 7.1 in Ch. 5). This low temperature corresponds to a cold beam. The design velocity is given by the argument `vzmean` which we set by the source bias potential. Out of the source, we assume an initial beam envelope expansion of 3.78 mrad at the source plane. Lastly, the angle of the beam is given by `vrmx` and has a radial scaling applied via Warp.

```

1 beam.add_uniform_cylinder(
2     np = Np_inject,
3     rmax = 0.25mm,
4     zmin = wp.top.zinject[0],
5     zmax = wp.top.zinject[0] + beam.vbeam * wp.top.dt,
6     vthx = vthx,
7     vthy = vthy,
8     vthz = vthz,

```

```

9     vzmean = beam.vbeam ,
10    vxmean = 0.0 ,
11    vymeans = 0.0 ,
12    vrmax = beam.vbeam * div_angle ,
13 )

```

Listing 8.1: Sample Warp input parameters for specifying a uniform cylindrical injection.

To correctly represent the number of physical particles in the system with $N_{p,\max}$ macro particles, the species weight w_p must be correctly specified. We use the desired source current I_s and calculate the w_p using:

$$w_p = \frac{I_s dt}{ZqN_{p,\text{inj}}},$$

where Z is the charge state, which we set as $Z = 1$ for singly charged ions. This weight sets the charge of each macro particle such that at each time step the slice of beam injected has current I_s .

8.3 Simulation Mesh

For higher final beam kinetic energy, the length of the linac increases to allow more acceleration gaps. In order to preserve the simulation quality and keep conductors resolved, the number of mesh points would need to increase as length is added to the system if the mesh remained stationary. This quickly becomes computational intractable. To limit mesh size and related storage needs, a moving simulation window is used. Warp allows one to specify a simulation box that can move at a specified speed. Conductors loaded onto the mesh only need to be placed within the simulation window. This significantly reduces computational

cost since the Poisson field solve for conductors does not need to be done in its entirety of the system on every time step. The window size is specified at initialization and is maintained throughout. Particles are set to be absorbed at the mesh boundaries (as well as on conductors), so it is important to set the upstream and downstream mesh window sufficiently so as not to immediately begin absorbing particles. The downstream window must be far enough away to allow particle acceleration and evolution without being particles being absorbed at the window edge. Recall that in Sec. 5.3, it was shown that in order to get the correct field structure, two gaps must be in the simulation window. With these considerations, the upstream simulation window extent is set to be $-\beta_b c \tau_{\text{rf}}$. For the downstream window extent, particle absorption can be avoided by specifying the symmetric extent $\beta_b c \tau_{\text{rf}}$. To ensure that there will always be two gaps within the window, the largest gap-to-gap distance is used which is the distance between the final two gaps within the accelerator lattice. Denoting this distance D , the upstream window is set as $D + \beta_b c \tau_{\text{rf}}$. This gives a total window size of $D + 2\beta_b c \tau_{\text{rf}}$, with the range being $z = [-\beta_b c \tau_{\text{rf}}, D + \beta_b c \tau_{\text{rf}}]$. Lastly, mesh refinement is used to increase resolution in the proximity of biased conductors. At each conductor (RF gaps and ESQ focusing elements are used) the transverse mesh is refined to give three times more grid points in x and y within the aperture. Longitudinally, the mesh is refined to give four times more grid points and extends over the physical longitudinal length of the conductor object. Using the source parameters in Table 7.1, the window size is set to approximately 3.8 cm. We use periodic boundary conditions to correctly model the array geometry with grid spacing $dx = dy = 50 \mu\text{m}$. The axial grid spacing is set to $dz = 0.42 \text{ mm}$ with Dirichlet boundary conditions for the upstream and downstream end of the simulation window. Near the conductor objects, the simulation grid is enhanced by a factor of three in both transverse and axial directions for higher resolution. An initial time step of $dt \approx 1 \text{ ns}$ which is adjusted

as the beam gains kinetic energy as discussed in Sec. 8.4.

8.4 Simulation Control

The simulation is controlled by five while-loops. The first loop injects particles each time step with the injection scheme previously discussed. Once the time step is equal to or greater than $\tau_b/2$, where $\tau_b = L_b/(\beta_b c)$ is the beam pulse duration, the synchronous design particle is activated and the simulation window is set to move (see Panels 2 and 3 in Fig. 8.1) with the speed of the design particle. While moving with the design particle, another while loop continues injecting particles until the simulation time is equal to or greater than τ_b and injection turns off. A third while-loop is then entered which covers most of the simulation. This while-loop execution is wrapped in a for-loop. The while-loop executes until the design particle reaches a gap center. After reaching a gap center (or pass in the for-loop), the time-step is recalculated based on the maximally attainable velocity in order to satisfy the Courant-Friedrichs-Lewy (CFL) condition condition [39]:

$$dt = 0.7 \frac{dz}{\beta_{\max} c}. \quad (8.1)$$

This effectively shortens the time step consistent with energy gain as the beam kinetic energy increases. Once the time-step is recalculated, the for-loop increments to the next gap-center and the while-loop executes until particles advance through the next gap. After the design particle passes each gap, the next while-loop executes until the design particle reaches the final diagnostic Fcup and is absorbed (see Panels 5 and 6 in Fig. 8.1). Finally, the last while-loop executes for an RF period to ensure that slower particles are can reach the Fcup or are

absorbed on the edges. Particles that arrive later are ignored at the end of the simulation. The time to ensure absorption can be decreased to speed up performance.

8.5 Diagnostic and Data Management

Data collection for diagnostics is done through Warp’s `zCrossingParticles` class. This tool is analogous to a laboratory BPM monitor and we refer to it as a numerical BPM. The user specifies the z -location of the device and the longitudinal width. As particles cross the grid cells where the numerical BPM is located, the macro particle data is stored. Data for the numerical BPM can be accessed through the getter attributes. We chose to extract transverse coordinates (x , y), transverse angle (xp , yp), transverse and longitudinal velocities (vx , vy , vz), and the time of crossing (τ). Data can be accessed through getter attributes for the numerical BPMs by including the prefix `get`, i.e., `getx`. These diagnostics are placed at the midpoint between each two neighboring gaps (see Panel 3 in Fig. 8.1). If an ESQ matching section is included, diagnostics are placed at the midpoint between two neighboring ESQs. An additional diagnostic is placed a few grid cells from the injection point to characterize the injected beam, and a final diagnostic is placed just before the F_{cup} to capture the final delivered beam. In general, there are $(N_g - 1) + 2$ BPMs not including additional diagnostics placed for the ESQ matching section which depends on the number of ESQs used. Once the particle advance is complete (see Sec. 8.4), the diagnostics are looped through and particle data arrays are stored in a dictionary since the data is generally inhomogeneous with different size data arrays at different numerical BPMS due to particle losses throughout the simulation. The dictionary is then written to a pickle file for storage. Data is used to calculate distribution statistics, make distribution projection plots, etc., and then saved into

a single PDF file. These data routines can be easily modified by the user for desired metrics and rerun without repeating the full simulation to allow rapid refinement of diagnostics as understanding increases.

8.6 Short Pulsed Beam Simulations

First we present self-consistent simulations of a short pulse beam in order to better observe the dynamics. A pulse duration corresponding to 20° of the RF period is injected about the design particle with initial energy $\mathcal{E}_i = 7 \text{ keV}$, and an initial envelope angle of 3.78 mrad in both transverse directions (diverging beam). The macro particle weight was set to ensure a current of $10 \mu\text{A}$ when injecting 10^5 macro particles over the pulse. To ensure that the on axis potential gain is 6 kV , V_g is slightly scaled up to 6.19 kV (see Ch. 6). In short lattice simulations, this scaling is not as important but becomes important in simulations of systems with many gaps due to “slippage” in the design particle synchronism if not accounted for.

We used a four gap lattice operating at maximum acceleration $\phi_s = {}^40$, where the notation ${}^n\phi$ indicates n gaps at operating phase ϕ , to establish a baseline. This design is designated Case 1. A Case 2 simulation incorporates tapered longitudinal focusing with gaps

Design Parameters	
Beam	
\mathcal{E}_b	7 keV
Injection Type	Pulsed
L_b	$\tau_{rf}/18$
Lattice	
N_g	4
V_g	6.19 kV
ϕ_s	$\phi_s = {}^40$

Table 8.1: Notable parameter values for Case 1. This short pulse design was for four gaps at max acceleration as noted by ϕ_s and forms the baseline for comparing other design choices.

spaced for $\phi_s = (-\pi/2, -\pi/3, -\pi/6, 0)$. Finally, Case 3 simulation is a modification of Case 2 with $\phi_s = (0, -\pi/6, -\pi/6, -\pi/2)$ for reasons that will be discussed. In comparing designs, we decided to monitor three attributes of the delivered beam at the F_{cup} :

1. Transmission: Fractional number of particles, relative to injection, delivered on target,
2. Current: $I = N_{\text{bin}} w_p q / \Delta t_{\text{bin}}$ calculated by binning the transmitted particles into equal time interval bins Δt_{bin} across simulations and then scaling by the macro-particle charge $w_p q$,
3. Radial distribution: $r_i = \sqrt{x_i^2 + y_i^2}$ calculated for every i -particle at the target location and then binned in radius $r = \sqrt{x^2 + y^2}$.

Fig. 8.2 gives plots for the radial distributions (left) and the calculated currents (right) of the three simulation cases. Each plot gives histograms: Case 1 in blue, Case 2 in orange, and Case 3 in green.

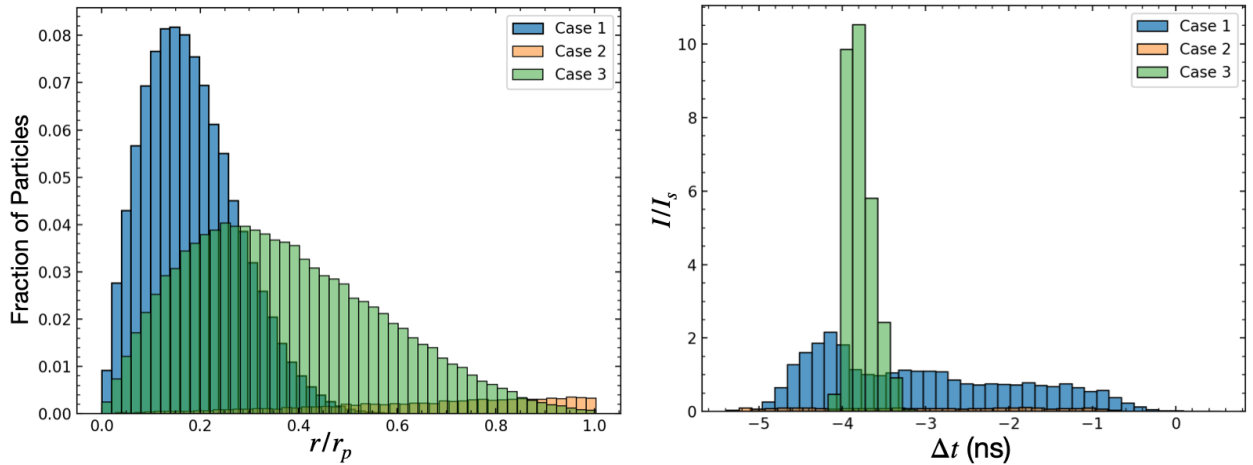


Figure 8.2: Simulation results for a four gap ($N_g = 4$) system. Snapshots of a binned radial beam profile as a function of radial distance normalized by the aperture r/r_p (left) and binned current normalized by the source current I_s (right) as a function of time relative to the synchronous particle. The bin widths in both plots are the same for the all three cases.

An immediate question is what happened in Case 2 for such poor performance? The answer is twofold. When the beam is compressed longitudinally, it expands transversely due to the increased space charge. Additionally, in Ch. 7 we showed that the first few acceleration gaps provide very strong transverse focusing and thus will also give strong transverse defocusing when the field is negative. At $\phi_s = -\pi/2$ particles that arrive in the gap early will experience strong defocusing. Furthermore, the RF synchronism corresponds to when the design particle reaches the center of the gap, and so particles lagging the design particle will also experience this defocusing. This was unaccounted for in the pure transverse model of Ch. 7. To mitigate this effect, the RF synchronism should be moved away from $\phi_s = -\pi/2$. so that the head of the beam does not get defocused. With this in mind, Case 3 was formulated with $\phi_s = -\pi/6$ being sufficient so that the head of the beam is not defocused, and still longitudinal bunching is maintained. In Case 3, the last gap was dedicated to maximally focusing longitudinally to produce a compressed beam just before the Fcup. The intuition in Case 3 succeeded: although there is more radial spread than in Case 1, the beam is radially well-confined and the calculated current on target five times greater relative to Case 1. Corresponding to peak current and transmission factors on target for the three cases are given in Table 8.2. Note for Case 3 there was only a 1% particle loss.

Case	Peak Current (μA)	Transmission
1	2.1	1.0
2	0.1	0.09
3	10.4	0.99

Table 8.2: Peak current and transmission fraction for different design cases using pulsed beam duration width of 20° of the RF period about the design particle for four acceleration gaps.

Snapshots of the beam evolution for the three cases (Case 1 in Fig. 8.3, Case 2 in Fig. 8.4, and Case 3 in Fig. 8.5). Panel 2 and 3 in Fig. 8.4 shows clearly the effects of the first gap set to maximally bunch. Initially, the particles are not lost but significant losses occur after gap 3. We can see from Fig. 8.5 how reversing Case 2 and employing a weaker bunching design keeps the beam well confined until the final gap maximally bunches the beam (panel 5). However, due to the previous gap radially focusing the beam, the expansion is not deleterious to the beam over the drift distance to the Fcup, and thus we achieve a good balance between radial and longitudinal focusing.

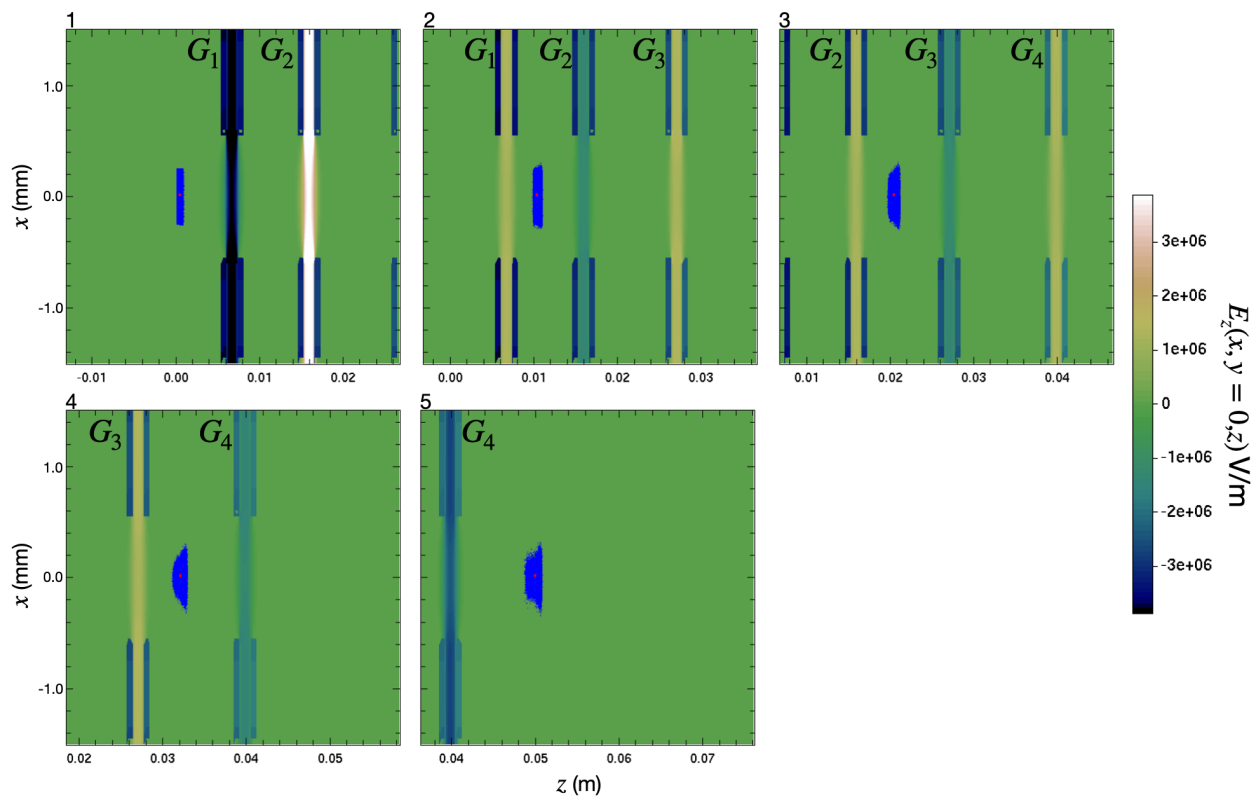


Figure 8.3: Short pulse injection into a 4-gap lattice operating at maximum acceleration with synchronous phase $\phi_s = 40^\circ$. No ESQ focusing.

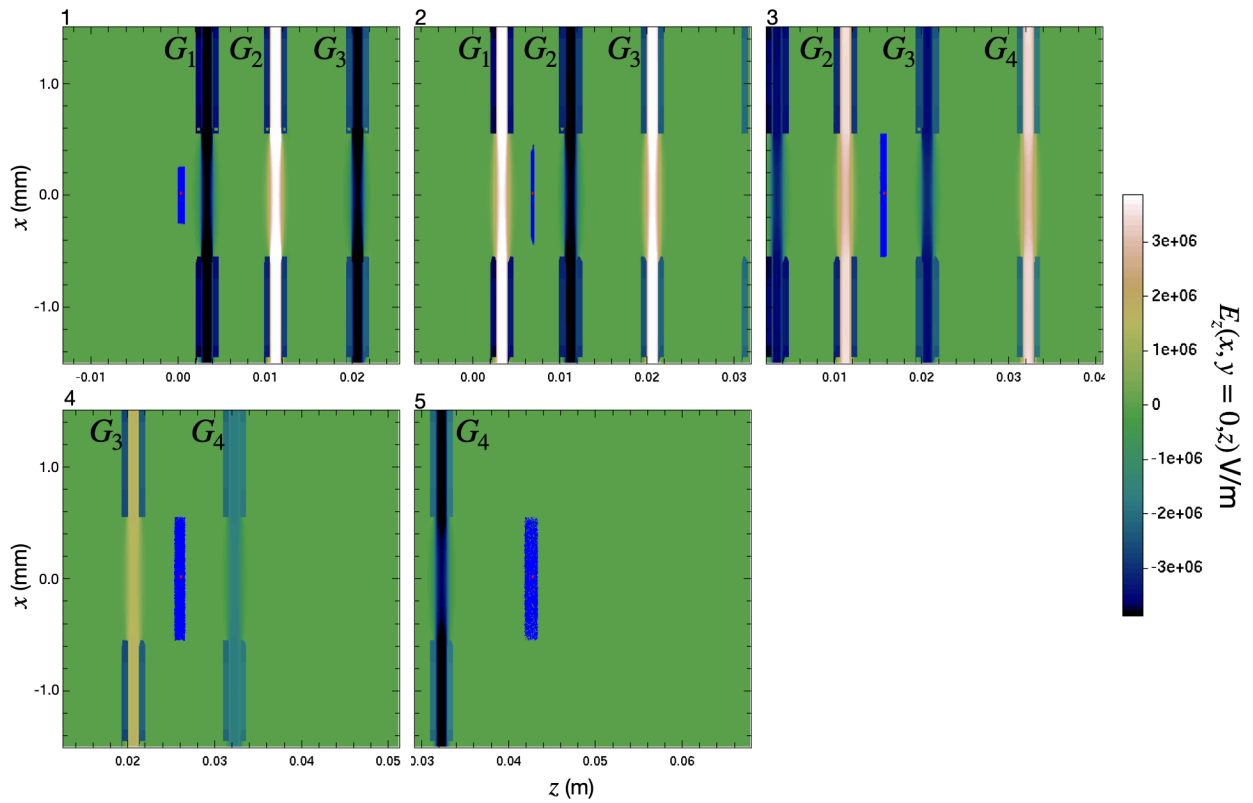


Figure 8.4: Short pulse injection into a 4-gap lattice with tapered synchronous phase for simultaneous acceleration and longitudinal focusing with $\phi_s = (-\pi/2, -\pi/3, -\pi/6, 0)$. No ESQ focusing.

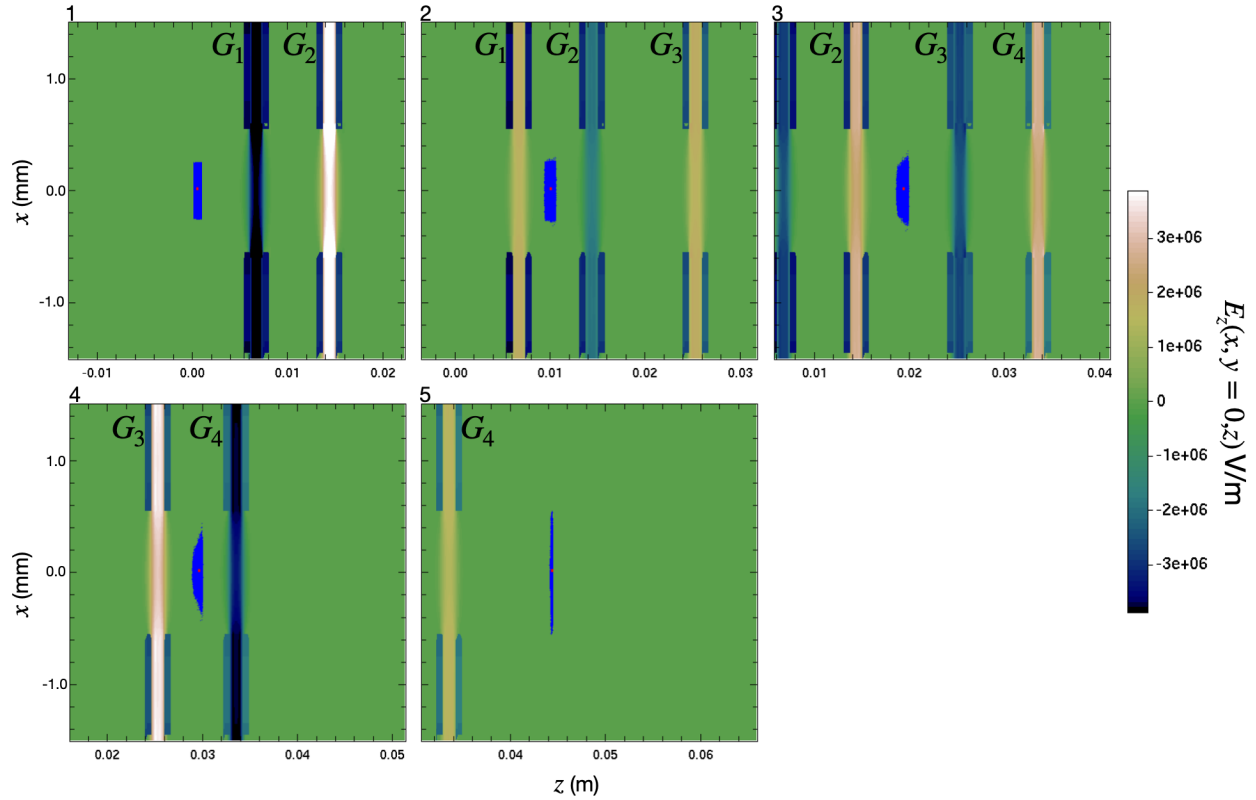


Figure 8.5: Short pulse injection into a 4-gap lattice tapered synchronous phase for simultaneous acceleration and longitudinal focusing with $\phi_s = (0, -\pi/6, -\pi/6, -\pi/2)$. Initial phasing is designed for max acceleration to avoid defocusing and maximum longitudinal focusing is designed for the last gap with $\phi_s = -\pi/2$. No ESQ focusing.

To observe effects of the ESQ matching section, and beam dynamics for many gaps after the matching section, we repeated the simulations for $N_g = 16$ (two additional gaps were added in one case which will be explained) with a four ESQ matching section after gap 6 using the voltage and lattice design shown in Fig. 7.13. The length of the matching section is kept consistent with RF synchronism. We compared three cases:

1. Case 1: No matching section, no periodic doublet focusing, and max acceleration

$$\phi_s = 180,$$

2. Case 2: Matching with downstream periodic doublet ESQ focusing and initial tapered

$$\text{bunching with } \phi_s = (0 \quad 3-\pi/6 \quad 120),$$

3. Case 3: Matching with downstream periodic doublet ESQ focusing and max accelera-

$$\text{tion } \phi_s = 160.$$

Case 1 again represents a baseline for comparison. Two additional gaps were added to Case 1 to see if there was a persistent trend of particle loss. Case 2 represents the point design developed with the matching and lattice design obtained with the thick gap envelope model in Ch. 7. Case 3 is a speculative “see what happens” scenario merging aspects of the previous two cases. Because the system has low phase advance ($\sigma \approx 13^\circ/\text{period}$) so one expects Case 3 to yield preferable results—especially considering the radial confinement advantage discussed in the four gap analysis. Fig. 8.6 shows the radial distribution (left) and the normalized current (right) for the three cases (Case 1 in blue, Case 2 in orange, and Case 3 in green). The radial distribution in Case 1 being drastically different from the others is not surprising. Recall in Ch. 7 that, after some gaps, the beam begins to expand from over focusing. This is precisely the effect shown here since the beam has expanded and almost uniformly fills the aperture. Almost half the particles are lost after transport (see Table 8.3)

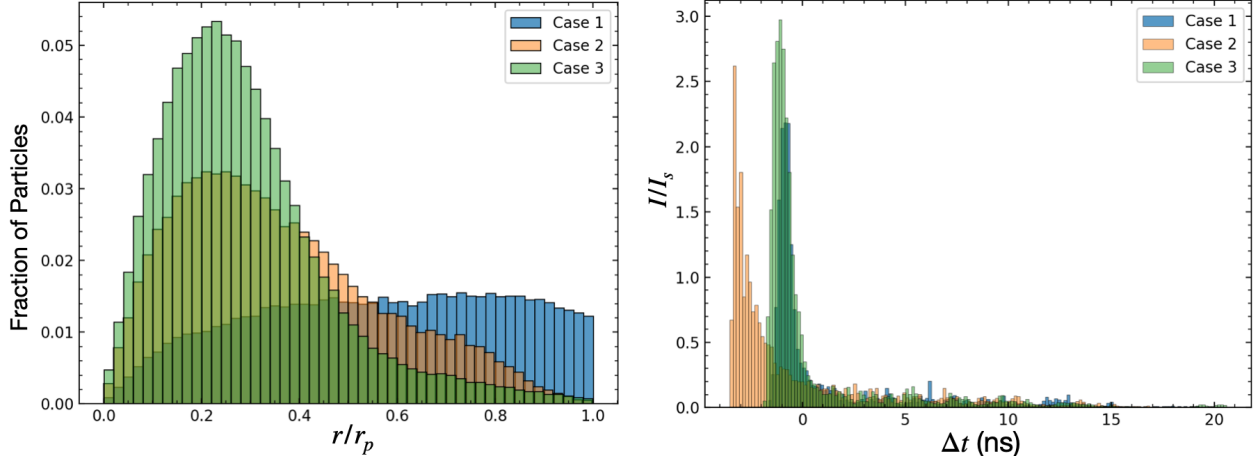


Figure 8.6: Simulation of results of a 16 gap ($N_g = 16$) system. Snapshots of the binned radial beam profile as a function of radial distance normalized by the aperture r/r_p (left) and binned current normalized by the source current I_s as a function of arrival time relative to the design particle (right) for three cases. Bin widths are the same for the all three cases.

and this trend in decreasing particles is persistent gap-to-gap after gap 8. It appears that the particle losses are beginning to plateau at gap 14. However, the additional two gaps showed this was not the case. Cases 2 and 3 are clearly superior to Case 1, but choosing a winner between Cases 2 and 3 is not obvious. At face-value, Case 3 appears superior. This design has better radial confinement, a higher peak current, and higher transmission. However, closer inspection of the current profile hints at a possible problem with Case 3. The peak, although highest, is also narrow compared to Case 2. Also, notice that the peak for Case 2 is well in the negative relative times meaning that most particles are arriving earlier. This hints at the possibility that a maximum bunching gap at the end of the lattice

Case	Peak Current (μA)	Transmission
1	2.1	0.62
2	2.6	0.79
3	2.9	0.94

Table 8.3: Values for the peak current and transmission for different design cases using a 20° pulsed beam and 16 acceleration gaps.

can significantly increase the peak current just as was shown in the 4-gap design. Since Case 3 has a thin peak that is only slightly in to negative relative times, it suggests that a maximum bunching gap would not address the issue and possibly even over focus the beam leading to transverse particle losses. Snapshots of the beam evolution for the three cases (Case 1 in Fig. 8.3, Case 2 in Fig. 8.4, and Case 3 in Fig. 8.5). The beam evolution snapshot Cases 2 and 3 give further insight when viewing the final panels for each case before the F_{cup} (panel 21). In Case 2, the beam is more longitudinally confined and the overall beam looks intact. Case 3 has a significant tail which will no doubt be absorbed if additional gaps were added. Also, the beam looks to be breaking apart as evident by the jagged outline of the x - z projection. This suggests that over the next few gaps, the beam will begin to lose cohesion in Case 3. To test this speculation, one would have to find the matching conditions for the next couple lattice periods as was done in Ch. 7. We leave this as an area for future investigation. It is evident that intricacies must be carefully scrutinized to weight relative merits of design choices. Applications may dictate the optimum.

The short pulse beam analysis provides insight on design choices for a continuously injected beam by allowing clearer visualization of the portion of the beam likely to transmit to target. When picking a design phases for a reference particle, bunching phases should be far enough from $\phi_s = -\pi/2$ so that the beam is not over focused. It will not be possible to prevent this effect over a continuous injection of a full RF period. Therefore we have a design tradeoff: we can prevent a larger section of the beam from being defocused by the RF gap by choosing synchronous phases close to $\phi_s = 0$ initially, but in doing so receive less longitudinal focusing. This tradeoff has not been thoroughly investigated and the best design strategy remains an open question. Furthermore, we should avoid strong bunching early in the lattice as this will drive beam expansion and potentially significant losses in

particles. Using the final gaps in the system to maximally compress the beam before the target worked well and suggests a strategy for design in cases where applications might need highly peaked flux on target within the RF periods. With the short pulse injection results in mind, we now turn to the case of continuous injection.

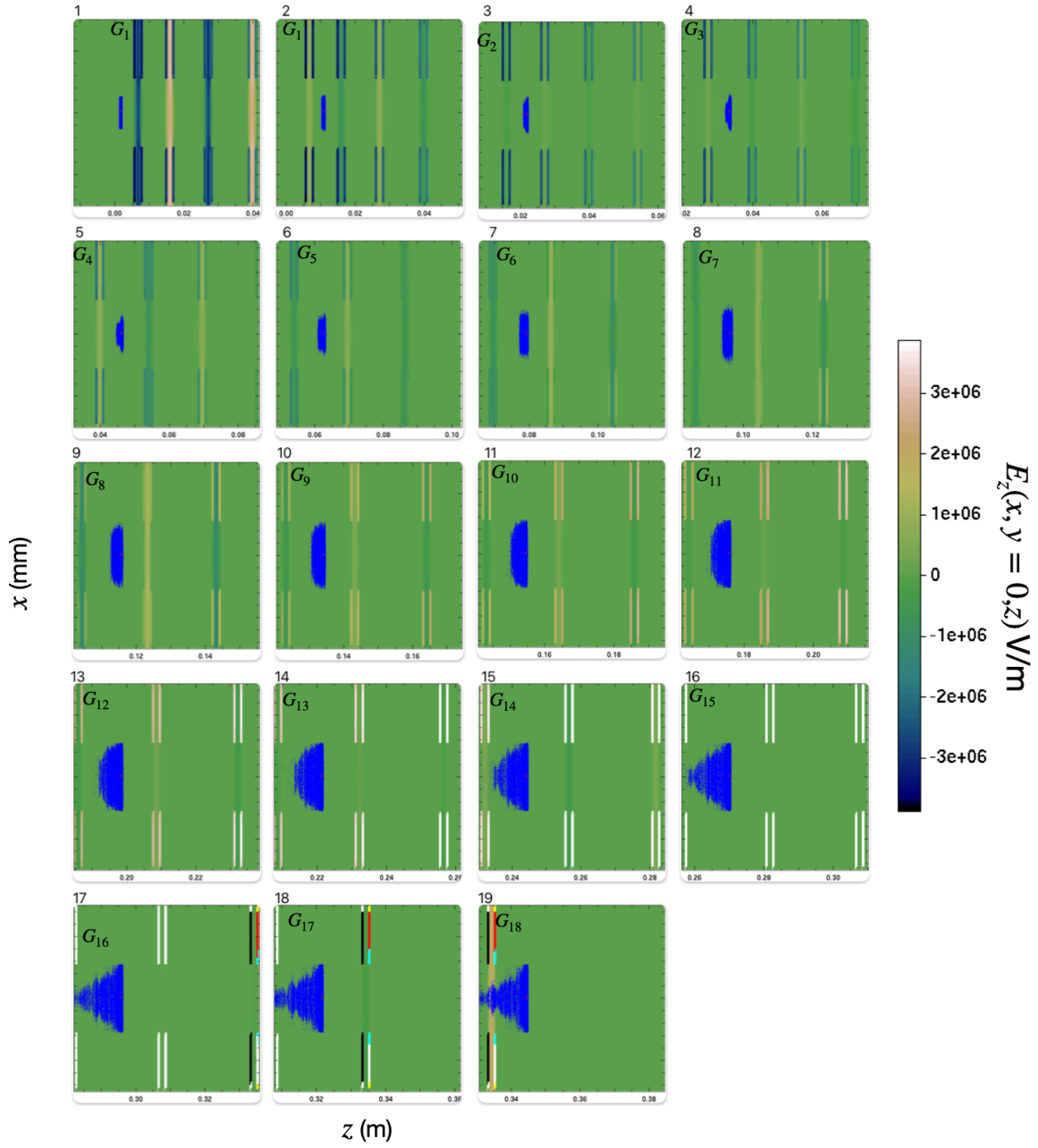


Figure 8.7: Short pulse injection into an 18-gap lattice with maximum acceleration with $\phi_s = 180^\circ$ at all gaps. To minimize clutter, only the first gap to appear in the frame is labeled and the vertical axis of each plot is the same as those in plots above with maximum extents of ± 1.5 mm. There is no ESQ focusing.

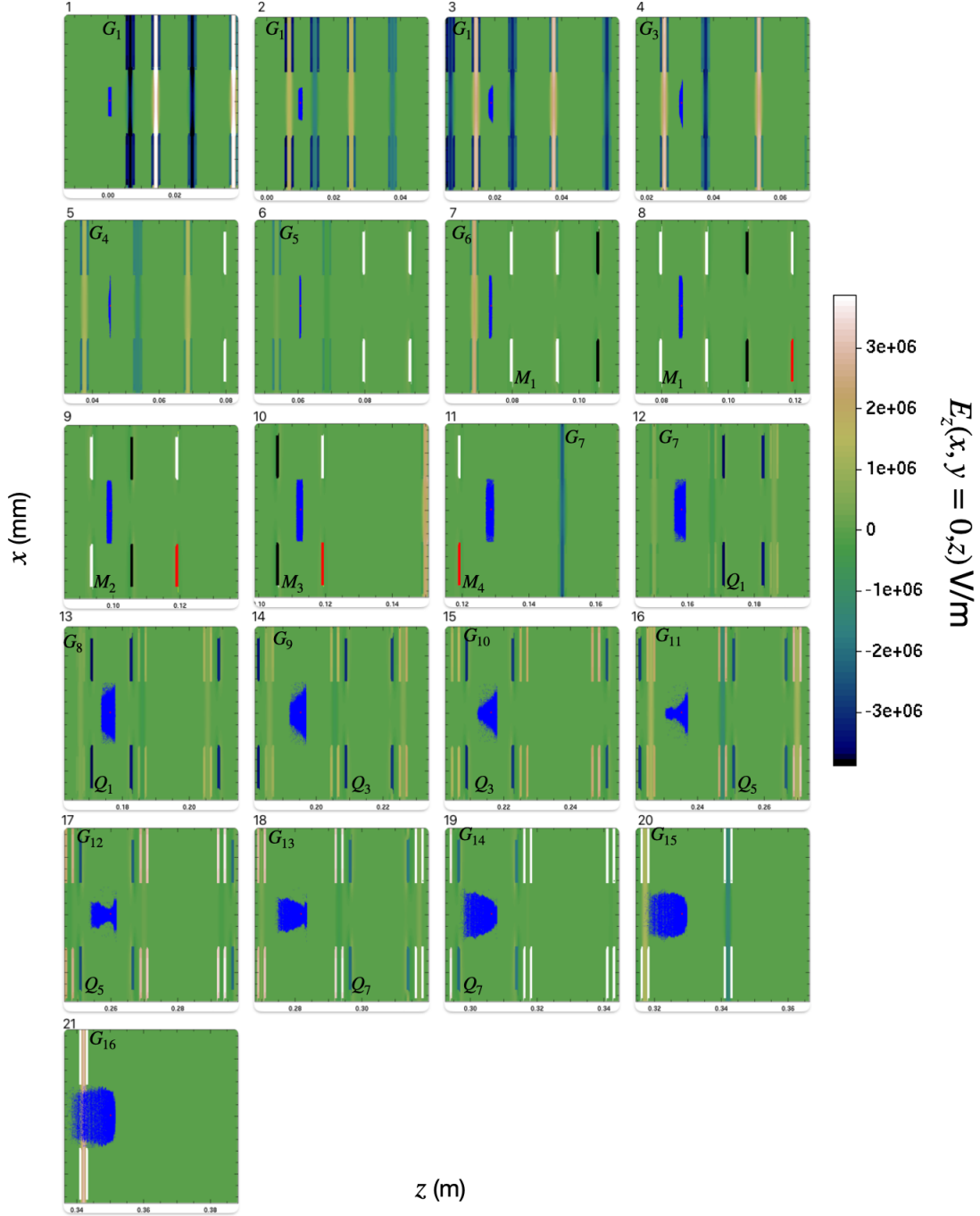


Figure 8.8: Short pulse injection into a 16-gap lattice designed for simultaneous longitudinal focusing and acceleration with $\phi_s = (0 \quad 3-\pi/3 \quad 120)$, a 4-ESQ matching section, and periodic ESQ doublet focusing and acceleration. The first six gaps are without ESQ focusing. Then there is a four ESQ matching section followed by five lattice periods (10 gaps) of acceleration and periodic doublet ESQ focusing. To minimize clutter, only the first gap to appear in the frame is labeled and the vertical axis of each plot extends ± 1.5 mm. Additionally, the first appearing ESQ is labeled either M_n for a matching ESQ or Q_n for a doublet ESQ. Note that each there are two ESQs for every Q_n -label. Due to complications with the cgm data writing, some conductors have off-coloring (red, teal). This has no physical significance.

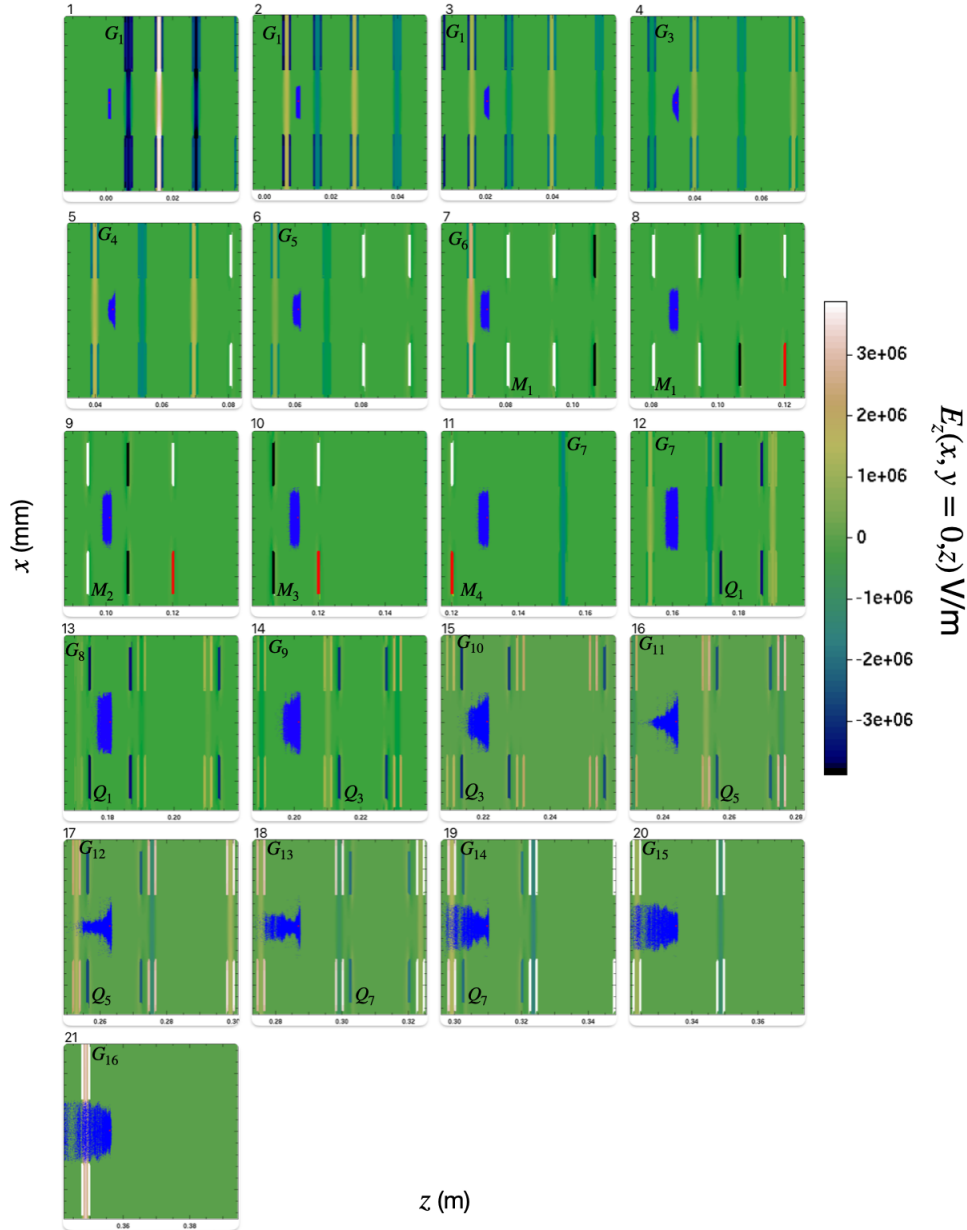


Figure 8.9: Short pulse injection into a 16-gap lattice designed for maximum acceleration with $\phi_s = {}^{16}0$, a 4-ESQ matching section, and periodic ESQ doublet focusing and acceleration. The first six gaps are without ESQ focusing. Then there is a four ESQ matching section followed by five lattice periods (10 gaps) of acceleration and periodic doublet ESQ focusing. To minimize clutter, only the first gap to appear in the frame is labeled and the vertical axis of each plot extends ± 1.5 mm. Additionally, the first appearing ESQ is labeled either M_n for a matching ESQ or Q_n for a doublet ESQ. Note that each there are two ESQs for every Q_n -label. Due to complications with the cgm data writing, some conductors have off-coloring (red, teal). This has no physical significance.

8.7 Continuous Beam Injection Simulations

When simulating a continuously injected (CW) beam, we inject for a full RF period about the design particle and then terminate the injection. This differs from a laboratory CW beam where there will be particles ahead of the beam (past the head) and particles behind the beam (before the tail). Particles from outside the RF period can influence beam dynamics and “jump” into the RF period modeled. Therefore, there is the potential that the CW injection scheme of the simulation is missing important physics. To test this, we injected particles for a full extra period both before and after the RF period about the synchronous particle and compared final outputs. No significant deviations were observed in these simulations relative to a single RF period about the design particle showing that the physics is accurately captured. Since this results in significant numerical savings we only simulate the primary RF period. As in the previous section, we test three design cases using a 16 gap lattice, all with an initial six gaps without ESQ focusing:

1. Case 1: No matching section or downstream ESQ periodic doublet focusing and max acceleration with $\phi_s = {}^{16}0$,
2. Case 2: Matching section with downstream ESQ periodic doublet focusing, and initial tapering acceleration with $\phi_s = (0 \quad {}^3-\pi/6 \quad {}^{12}0)$,
3. Case 3: Matching section, downstream ESQ periodic doublet focusing, and max acceleration with $\phi_s = {}^{16}0$.

Analogously to the preview cases, the radial distribution and calculated current at the Fcup are shown in Fig. 8.10 for each case along with snapshots of the beam evolution given in Fig. 8.11 for Case 1, Fig. 8.12 for Case 2, and Fig. 8.13 for Case 3. Table. 8.4 lists the peak

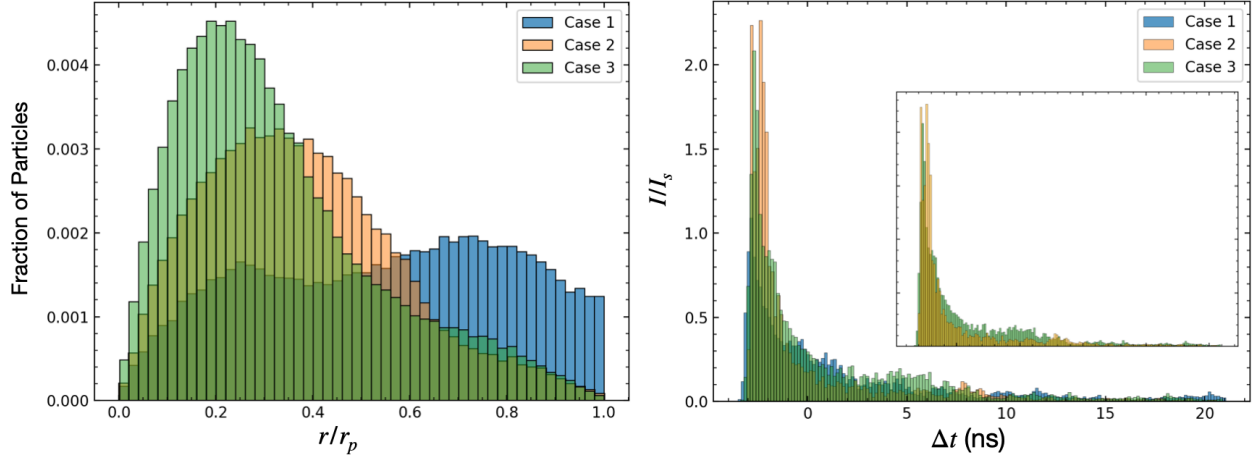


Figure 8.10: Simulation of results of a 16 acceleration gap ($N_g = 16$) system with continuous (full RF period) injection. Snapshots of the binned radial beam profile as a function of radial distance normalized by the aperture r/r_p (left) and binned current normalized by the source current I_s as a function of arrival time relative to the design particle (right) for three cases. Bin widths are the same for the all three cases. The inset plot shows Case 2 and Case 3 together to more clearly compare the current profiles.

current and transmission values for each case. Once again, the maximum acceleration in Case 1 is inferior to Cases 2 and 3—there is minimal radial confinement, low peak current, and low transmission. Again, the relative merits between Cases 2 and 3 are not obvious. Some differences should be noted compared to the corresponding previous short pulse injection cases. One can argue that the radial distribution for Case 2 is the same, if not better, than Case 3. The radial tail of the distributions are approximately the same and the beam is not so heavily concentrated as it is in Case 3. Additionally, the inset image on the current plot shows Cases 2 and 3 alone, showing that Case 2 has a slimmer peak. Also, Case 2 has less particles in the tail ends ($\Delta t > 0$), which suggests a better bunched beam compared to Case 3. It is difficult to discern which design performs better between Cases 2 and 3. However, and unsurprisingly, the matching section and ESQ doublet focusing yields for better performance.

The transmission values may be slightly alarming and raises the question: what trans-

Case	Peak Current (μA)	Transmission
1	1.1	0.073
2	2.3	0.082
3	2.1	0.096

Table 8.4: Peak current and transmission fraction for different design cases.

mission should we expect? For a uniform CW injection we expect there to be particle losses. In Ch. 2, longitudinal dynamics the RF linacs were reviewed. Recall that the region of $0 < \phi_s < -\pi/2$ simultaneously provided particle capture and acceleration. Since this region is 25% the RF period, we can roughly expect an upper bound for particle transmission of 0.25, i.e. we can at best capture 25% of a uniform CW beam¹. To establish a lower bound, one could iteratively simulate the beam for different pulse lengths using the desired design lattice. Starting with a pulse length of $\tau_b = \tau_{rf}$, one could decrease the pulse length until a transmission near 1.0 is attained giving a pseudo RF bucket. The end pulse-length $\tilde{\tau}_b$ will give the fraction of a uniform CW beam one could expect to fully capture and the expected transmission will be $\tilde{\tau}_b/\tau_b$. The transmission values in CW injections should fall between $\tilde{\tau}_b/\tau_b$ and 0.25. In the pulsed injection study, a beam pulse length corresponding to 20° RF phase extent about the design particle was used (5% of uniform CW beam) and the transmission was 0.96 (see Table 8.3). This comes close to the lower bound and assuming such, we can expect that in a uniform CW injection we should be able to attain at least 0.05 transmission factor and design for higher values. Indeed all three cases yield a higher transmission factor with Case 1 (with no longitudinal focusing) being close to the lower bound and Case 3 almost doubling it. These higher transmission results suggest that even when operating at max acceleration (which gives no particle capture for weak acceleration)

¹This upper bound assumes no particle transfer from areas outside the operating phase used. If the gap accelerations were perturbative then this would be a reasonable lower bound.

we are still capturing some particles in this pseudo RF bucket due to nonlinear dynamics and strong acceleration.

It is important to observe where in the lattice the significant particle losses occur. In all three cases, 50% of losses occur by the second gap. Before the matching section (or after gap 6 in these cases), 80-90% losses occur. Recall that Warp does not have the capabilities to simulate dielectric effects in a 3D simulation. The substantial loss of particles can charge on surrounding dielectric materials and possibly strongly perturb the beam with generated fields. The effects of this could be significant and warrant further investigation. Because Warp can track where particles are lost on material surfaces, it is an ideal tool to employ to support design of shielding and scrapers that should be placed early in the lattice for protection.

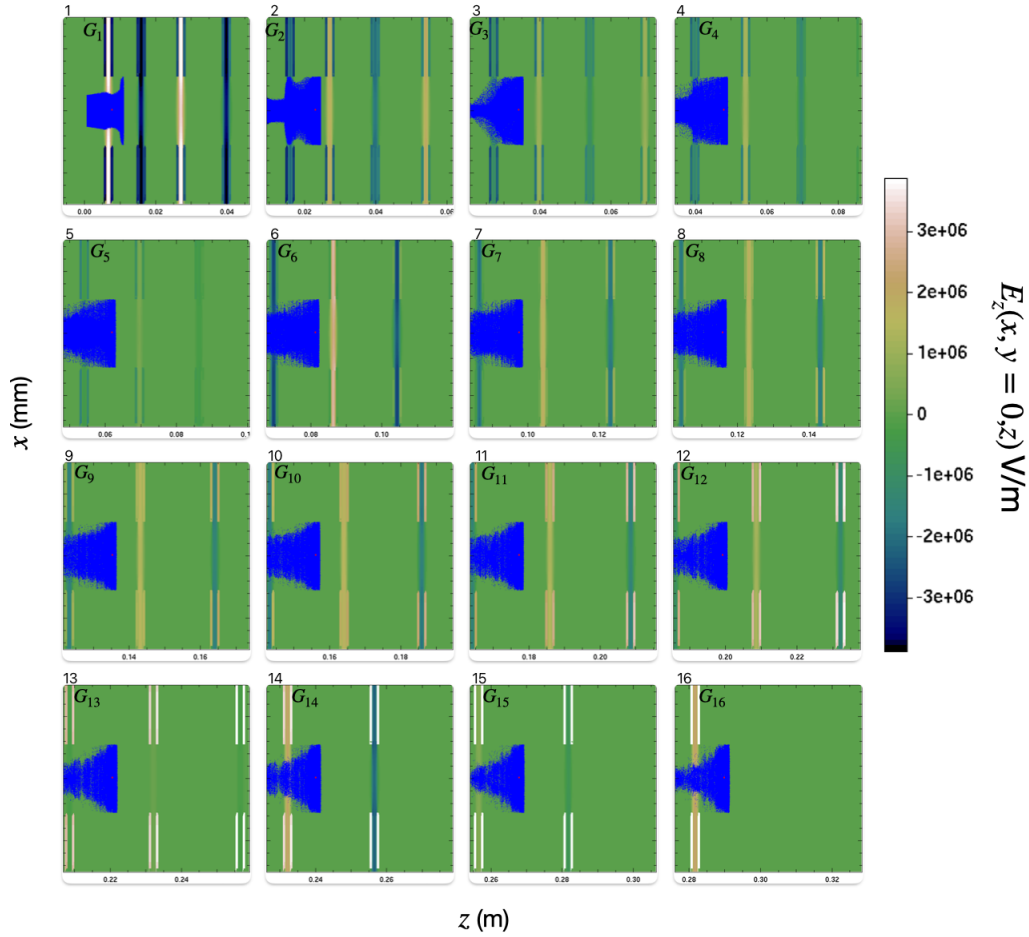


Figure 8.11: Continuous injection into a 16-gap lattice designed for max acceleration with $\phi_s = 16\theta$. The first six gaps are without ESQ focusing. Then there is a four ESQ matching section followed by five lattice periods (10 gaps) of acceleration and periodic doublet ESQ focusing. To minimize clutter, only the first gap to appear in the frame is labeled and the vertical axis of each plot extends ± 1.5 mm. Additionally, the first appearing ESQ is labeled either M_n for a matching ESQ or Q_n for a doublet ESQ. Note that each there are two ESQs for every Q_n -label. Panel 1 is a snapshot of the beam after approximately one RF cycle.

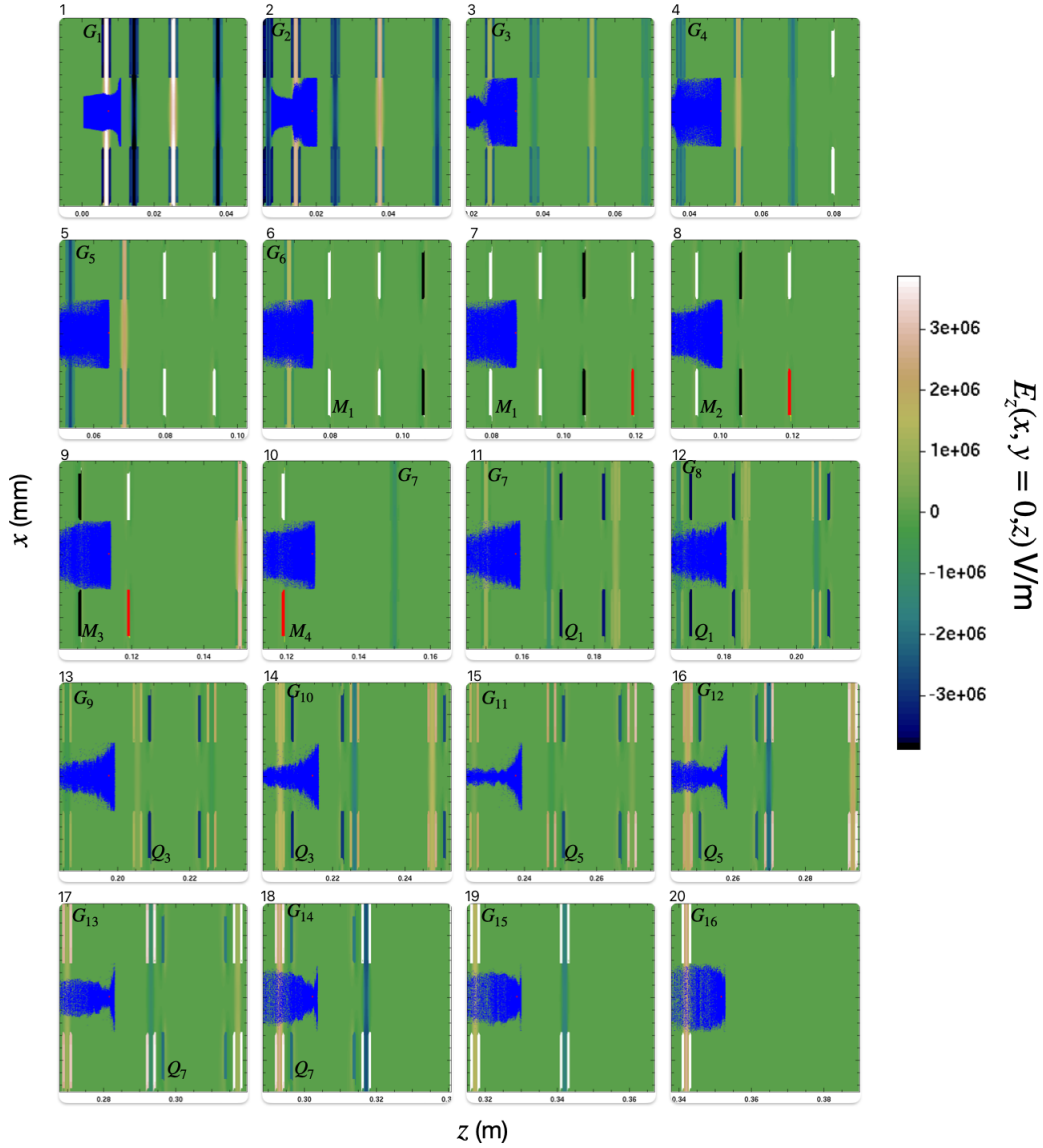


Figure 8.12: Continuous injection into 16-gap lattice designed for simultaneous acceleration and longitudinal focusing with $\phi_s = (0 \quad 3-\pi/3 \quad 120)$, a 4-ESQ matching section, and periodic ESQ doublet focusing and acceleration. The first six gaps are without ESQ focusing. Then there is a four ESQ matching section followed by five lattice periods (10 gaps) of acceleration and periodic doublet ESQ focusing. To minimize clutter, only the first gap to appear in the frame is labeled and the vertical axis of each plot extends ± 1.5 mm. Additionally, the first appearing ESQ is labeled either M_n for a matching ESQ or Q_n for a doublet ESQ. Note that each there are two ESQs for every Q_n -label. Due to complications with the cgm data writing, some conductors have off-coloring (red, teal). This has no physical significance. Panel 1 is a snapshot of the beam after approximately one RF cycle.

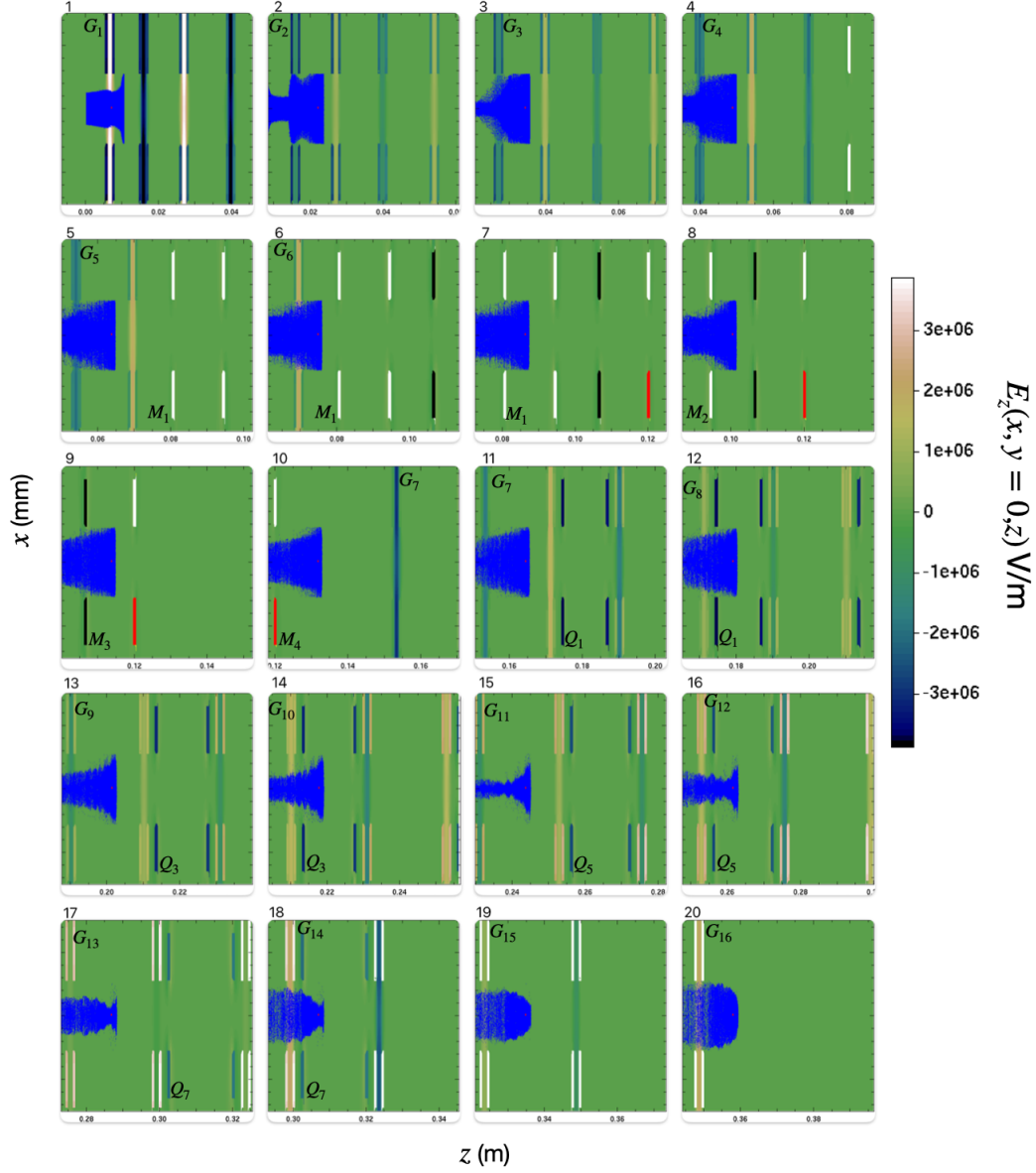


Figure 8.13: Continuous injection into 16-gap lattice designed for maximum acceleration with $\phi_s = 160$, a 4-ESQ matching section, and periodic ESQ doublet focusing and acceleration. The first six gaps are without ESQ focusing. Then there is a four ESQ matching section followed by five lattice periods (10 gaps) of acceleration and periodic doublet ESQ focusing. To minimize clutter, only the first gap to appear in the frame is labeled and the vertical axis of each plot extends ± 1.5 mm. Additionally, the first appearing ESQ is labeled either M_n for a matching ESQ or Q_n for a doublet ESQ. Note that each there are two ESQs for every Q_n label. Due to complications with the cgm data writing, some conductors have off-coloring (red, teal). This has no physical significance. Panel 1 is a snapshot of the beam after approximately one RF cycle.

As a final study, we wish to qualitatively analyze the strength of the space charge forces present. Using a similar continuous injection scheme as presented above, we modeled a six gap system with no ESQ focusing and acceleration gaps spaced for maximum acceleration with $\phi_s = 60$. The injected beam species weight was scaled to simulate a beam with no space charge forces (approximately zero species weight), and a species weight calculated from the Child-Langmuir current limit calculated using (see Ch. 5 in Ref. [40])

$$J_{\text{CL}} = 5.44 \times 10^{-8} \left(\frac{Z}{A} \right)^{1/2} \frac{V_0^{3/2}}{d^2}, \quad (8.2)$$

with charge state Z , atomic number A , and extraction voltage V_0 applied over a distance d . Applying an extraction bias voltage of $V_0 = 6$ kV over a distance of $d = 4$ mm, the Child-Langmuir (CL) current density for singly charged Ar^+ ions is

$$J_{\text{CL}} \approx 372 \frac{\mu\text{A}}{\text{mm}^2}.$$

Taking the source aperture to be $r_s = 0.25$ mm gives a CL current limit of $I_{\text{CL}} \approx 70$ μA . Using I_{CL} , the species weights are set accordingly and Fig. 8.14 shows the radial distribution normalized by the $r_p = 0.55$ mm aperture (left) and the distribution in arrival time at the Fcup relative to the synchronous particle (right) with the y -axis giving the fraction of transmitted particles from the injected beam. The 10 μA current has relatively weak space charge as can be seen by comparing the orange distribution (space charge of a 10 μA injected beam current) to the blue distribution (no space charge). With no space charge, only 5% more particles are transmitted to the Fcup compared to the standard 10 μA injected beam current. However, if one were to increase the extracted current up to the CL limit, the space

charge forces become more apparent as can be seen by the green distribution (CL injected beam current) which results in approximately 50% reduction in the particle transmission. The array structure helps avoid this issue by increasing the number of beamlets for higher current densities rather than increasing the extracted current. These results also suggest that since the $10 \mu\text{A}$ injected beam current suggests weak space charge, the extracted current per beamlet can be increased from this value. Lastly, because of the weak space charge, simulation time can be reduced in full 3D Warp simulations by turning space charge off to give an approximation of the modeled beam.

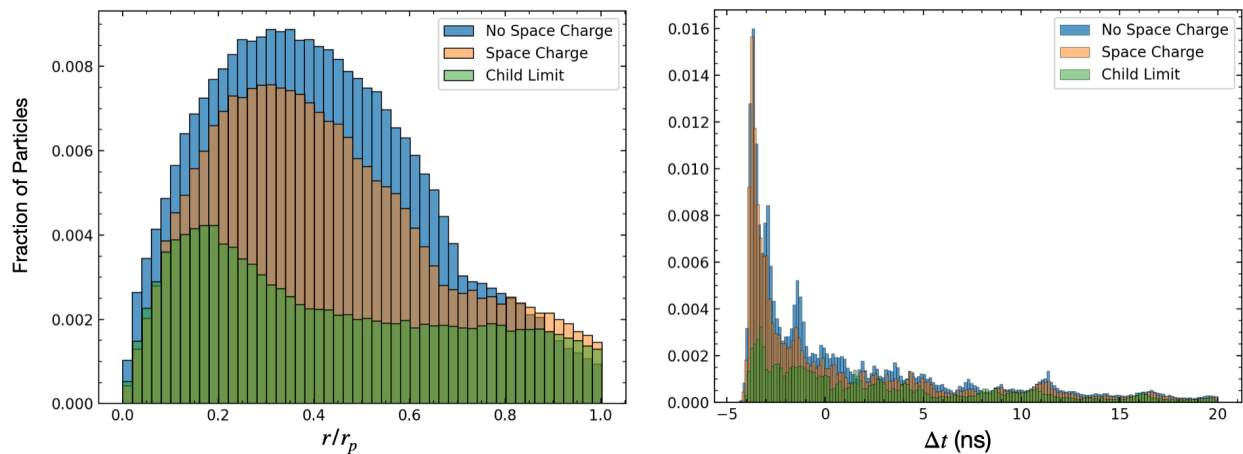


Figure 8.14: Simulation of results of a 6 acceleration gap ($N_g = 6$) system with continuous (full RF period) injection. Snapshots of the binned radial beam profile as a function of radial distance normalized by the aperture r/r_p (left) and binned arrival time profile as a function of arrival time relative to the design particle (right) for three cases. Bin widths are the same for the all three cases. The three cases shown are an injected beam current at the Child-Langmuir limit (green), a typical injected beam current of $10 \mu\text{A}$ (orange), and an injected beam with the self field modeling turned off (blue).

8.8 Summary and Concluding Remarks

This chapter presented self consistent 3D source to target Warp simulations to evaluate integrated performance of RF MEMS ion beam array. Reduced physics models presented in Chapters 6-7 were applied to efficiently formulate designs with vastly improved performance compared to our base case of a lattice designed for maximum acceleration with no ESQ focusing. Then 3D self consistent Warp simulations were applied to evaluate performance in a comprehensive sense. Using a combination of particle plots and data analysis, we observed better beam profiles, increased transmission, and higher beam currents delivered to target when evaluating our improved designs. Our current optimal design consisted of a 16 gap lattice producing an ≈ 100 keV Ar^+ ion beam array. The first six gaps were accelerator gap only, and several optimizations of design phase tapering were explored. A four ESQ matching section with voltages and spacing shown in Fig. 7.13 was used to match the beam downstream of the Gap-only section to the rest of lattice which consisted of acceleration gaps and periodic doublet ESQ focusing (Gap+ESQ section). The ESQ doublet bias voltages that were found with a thick gap envelope model are given in Table. 7.4 were used, and the remaining gaps were spaced for maximum acceleration ($\phi_s = 100$). We found that 80-90% of the beam was lost after the first six gaps with roughly 50% of the beam lost after the second gap structure suggesting that the system would benefit from shielding at these locations to prevent damage from lost beam particles and/or charging of insulators. Work presented here was only an initial start. We did not find a clear global optimum design and it is apparent that such a task requires considerably more work outside the scope of this dissertation project that was focused on overall approaches and development of tools to efficiently guide optimization strategies. The techniques discussed and simulation itself form the fundamental pieces for

taking such an endeavor into the next logical steps.

There are several areas of improvement that further investigations would benefit from to more efficiently evaluate design options. Simulation run time can be significantly reduced by importing gridded electric fields to eliminate needing 3D field solves of the conductor objects every time step. However, 3D field solves are needed to evaluate the beam self field. To do this, one would have to extract the 3D fields for structures on the mesh and scale the applied RF field components. This would require the current code to be restructured. The imported fields need to be correct for the given system, and thus any changes in conductor spacing where there is some field overlap, like near the ESQs around neighboring wafer plates in the periodic doublet focusing, would need to be correctly loaded on to the grid. Once this is completed, the simulation should be compared to experiment in controlled situations to test validity and establish confidence that the simulation is reliably describing the experimental system. From this point, there are many directions that can be taken to support laboratory work. Efficient simulations can be used to create synthetic experimental data which can then be fed into a neural network to seek significant design improvements for specific applications, acts as a surrogate model, etc.

Part III

Machine Learning and Optimization of an Ion Induction Linac

Chapter 9. Introduction and System Overview

Accelerators are highly complex machines with rich physics and can also produce large physical data sets over extend operating periods. This situation is ripe for machine learning (ML) applications such as fault detection, anomaly detection, modeling and optimization, and more [42]. The remaining chapters are dedicated to a short-lived ML project applied to an induction accelerator called the Neutral Drift Compression Experiment II (NDCX-II) [41, 43] at Lawrence Berkeley National Laboratory. We assume general knowledge of ML and neural network (NN) architectures. If this is not the case, we briefly list available

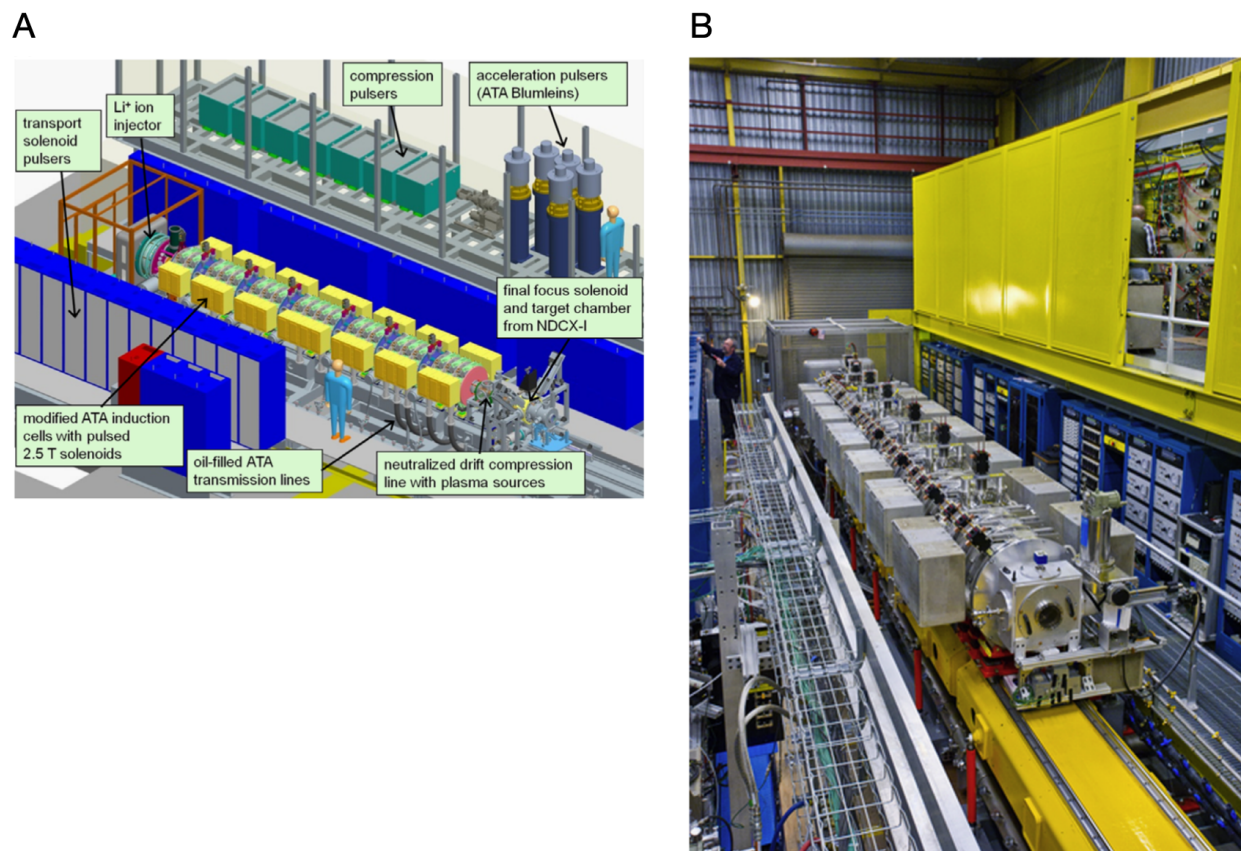


Figure 9.1: (A) CAD drawing of the NDCX-II with various components labeled. (B) Photo of the accelerator. Image borrowed from Ref. [41].

resources that were of particular use to the author. Of the many resources, Ref. [44] gives a practical overview directed at physicists and has accompanying Jupyter notebooks for practice. There is also a free textbook [45] that gives a more rigorous approach to the mathematics behind machine learning. There are relatively cheap courses, of which, Andrew Ng’s [46] and Jose Portilla’s [47] are highly recommended. For MSU students, Dr. Wade Fisher’s course on Data Analysis is a must take. Lastly, the two textbooks in Refs. [48, 49] are excellent sources for practical knowledge on machine learning techniques and both have accompanying Jupyter notebooks and downloadable Python scripts to practice with. A final textbook in Ref. [50] is a balanced approach to machine learning with practical guidance and theoretical underpinnings.

NDCX-II was originally designed to probe materials in the warm dense matter regime by delivering an energetic and compressed beam to a target [43]. The beam is intense and compresses both longitudinally and transversely with significant factors. The facility was commissioned in 2012. A CAD drawing shows the key components (Panel A) and an actual picture of the accelerator (Panel B) in Fig. 9.1. The accelerator is approximately ten meters long and consists of 27 induction cells with pulse forming lines that perform acceleration, compression, and diagnostics. Fig. 9.2 shows the nominal beamline configuration. From the source, an ion beam of ~ 600 ns in pulse duration with kinetic energy of ~ 100 keV is extracted. Each of the 27 cells have an independently adjustable solenoid to provide radial focusing. Active induction cells, labeled “A” in Fig. 9.2, provide carefully tailored compression and acceleration voltage waveforms to the beam. The waveforms are shown in Fig. 9.3 for the 12 active induction cells—note the last five active induction cells are powered by Blumleins to provide a higher applied maximum voltage and are labeled accordingly. Labeled “I” in Fig. 9.2 are the inactive induction cells that do not provide acceleration

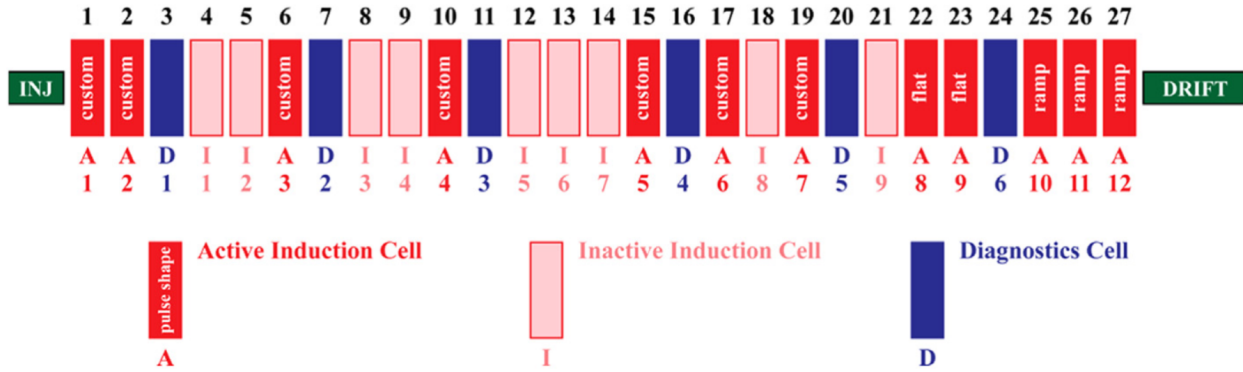


Figure 9.2: The nominal beamline configuration for NDCX-II [41]. Each of the 27 cells have solenoidal focusing. The dark red cells are active induction cells which accelerate and compress the beam, for which there are 12. The first seven have custom waveforms designed for optimal performance. Following are two flat waveform cells and then the last through apply a ramp waveform to longitudinal compress the beam before entering a neutral drift plasma region. In light red are inactive cells that only contribute solenoidal focusing. In dark blue are diagnostic cells for beam measurements.

and act as a drift space so that the beam can longitudinally compress. These cells are physically equivalent to the active cells and thus provides some variability in the beam line configuration. Before arriving at the Blumlein powered induction cells, the beam has been compressed to less than 70 ns. A finally velocity tilt is provided by the Blumleins to further compress the beam to the space charge limit before entering a neutral drift plasma. The volumetric plasma is created with sufficient density in order to cancel the beam's space charge defocusing force. This allows for a final compression of the beam to 1-2 ns. After traversing the cells, the beam enters a final neutral drift plasma to neutralize the beam before arriving at the target. The plasma neutralizes the beam to allow pulse compression. The result is an ion beam that is both compressed longitudinally to less than 1 ns and accelerated to greater than 1 MeV with a spot size of ~ 1 mm. The 12 active induction cells and 27 focusing solenoids in NDCX-II combine to give a large number of parameters that need to be optimized to deliver the final pulse on target. Both high flux on target and short pulse durations are desirable. The large number of parameters makes the machine very

challenging to tune in the presence of limited diagnostics. Warp simulations are employed as a guide using experimental waveforms to accurately represent the laboratory situation. These simulations are found to have reasonable fidelity with the experiment but are time consuming for laboratory support of operations.

Our intent was to use NDCX-II to explore ML applications. Initially, we sought to both test an auto tuning algorithm (see Ch. 10) to improve facility operations while also sampling the large parameter space and collecting data to be used for ML applications such as surrogate modeling and optimization. We were operating on a very limited budget that was only sufficient to cover operation of NDCX-II to support a proposal to secure more funding. However, we were not awarded more funding and the project was limited to previously collected data that was archived. We use the remaining chapters to document work carried and to show how an ML project in accelerator physics might take shape.

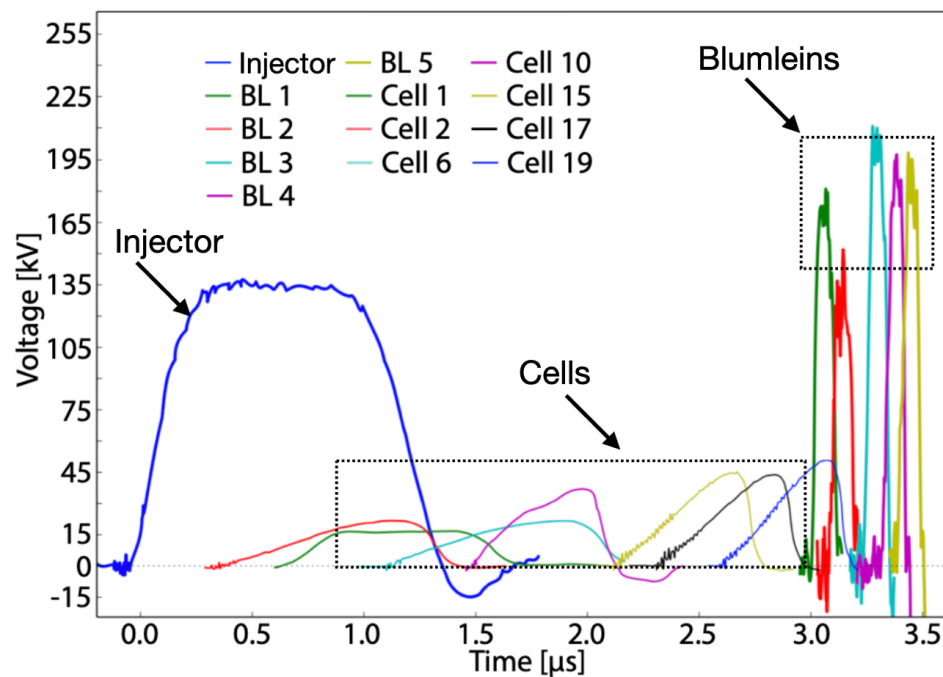


Figure 9.3: Voltage waveforms for the 12 active induction cells and the injector [43]. The first seven cells are custom waveforms labeled by the Cell # on the plot. The remaining five active induction cells are powered by Blumleins and are labeled with BL on the plot.

Chapter 10. Auto Tuning Algorithm for Safe Extremum Seeking in NDCX-II

Once a device has been optimized for performance there is no guarantee that the found optimum settings will remain constant over time—i.e., there is no certainty the optimal applied voltage for the solenoids Fig. 9.2 will remain the same. This “drift” is due to changes in the equipment over time, such as operating temperature and equipment degradation. However, during normal operation, the drift should not be substantial and, in principle, can be found again by sampling the parameter space near the previously found optimum. When sampling the parameter space one must be careful since some settings can be deleterious to the machine. Additionally, the settings are highly interdependent. If the timing is for one of the conductors, this will have an effect for the downstream conductors. With such a narrow optimum, this makes tuning especially difficult and intractable to do by individually adjusting parameters which makes for a perfect case in applying ML techniques. To autotune NDCX-II, we used a local, model-independent extremum seeking (ES) algorithm [51, 52, 53]. The algorithm’s simple implementation and quick convergence made it highly desirable since we could monitor the tuning in real-time. Consider a time-varying and analytically unknown cost function that we seek to maximize $C(\vec{p}, t)$ along with the noise-corrupted measurement $\hat{C}(\vec{p}, t) = C(\vec{p}, t) + \eta(t)$. The varied parameters are given by the vector $\vec{p} = (p_1, p_2, \dots, p_n)$ which, in the case of NDCX-II, will be the application time and the applied bias for each of the 27 cells. The parameters are adjusted by the dynamical equation:

$$\frac{dp_j}{dt} = \sqrt{\alpha\omega_j} \cos[\omega_j t - k\hat{C}(\vec{p}, t)] \quad (10.1)$$

where ω_j is the dither, α is the dithering amplitude, and k is the feedback gain. Local minima can be escaped by increasing α , and k is analogous to the learning rate as seen in gradient based methods. When the dynamics have settled the algorithm will oscillate around p_j with frequency $\sqrt{\alpha/\omega_j}$. Intuitively, the algorithm will converge because by applying a different dither to each parameter, the parameters are orthogonal in Hilbert space since

$$\lim_{\omega_1, \omega_2 \rightarrow \infty} \int_0^t \cos(\omega_1 \tau) \cos(\omega_2 \tau) d\tau = 0, \quad (10.2)$$

when $\omega_1 \neq \omega_2$.

Chapter 11. Data Collection, Retrieval, and Preprocessing

Using Eq. 10.1 the parameters can be varied to find optimum settings while simultaneously sampling the parameter space within a bounded region. The parameters we varied were the solenoid timing, i.e. when the solenoid voltage was applied to give radial focusing, and the applied voltage on the Blumleins. In reality, the applied voltage and time of application can be varied for both the solenoids and active induction cells. However, we decided to keep the custom waveforms and only vary the voltage on the Blumleins (final five active induction cells). Control of NDCX-II is wrapped in a Python interface where the tuning algorithm can be easily implemented to vary the conductor settings. For a beam pulse or “shot”, all voltage and current waveforms are stored in a database for the conductors, ion source, and Fcup. Functionality to query the database has been written in Python making data extraction easy. Individual shot data is collected in a shotnumber data object that contains individual shot classes holding data attributes. After querying the database, a shotnumber or list of shotnumbers can be collected, the shots iterated, and data for individual conductor objects extracted. The process is illustrated graphically in Fig. 11.1.

Once extracted, the data is filtered and stored in a comma separated value (csv) file. This format is easily manipulated by the Pandas library [54] in Python and a file can be loaded into a dataframe object which is useful for data analysis and preprocessing for machine learning algorithms. Data was mainly filtered by looking at the total charge delivered on target. In our experiments, typical values ranged between 0.1 nC and 10 nC. Thus, values that were significantly different were ignored. Bad shots typically resulted from a misfiring

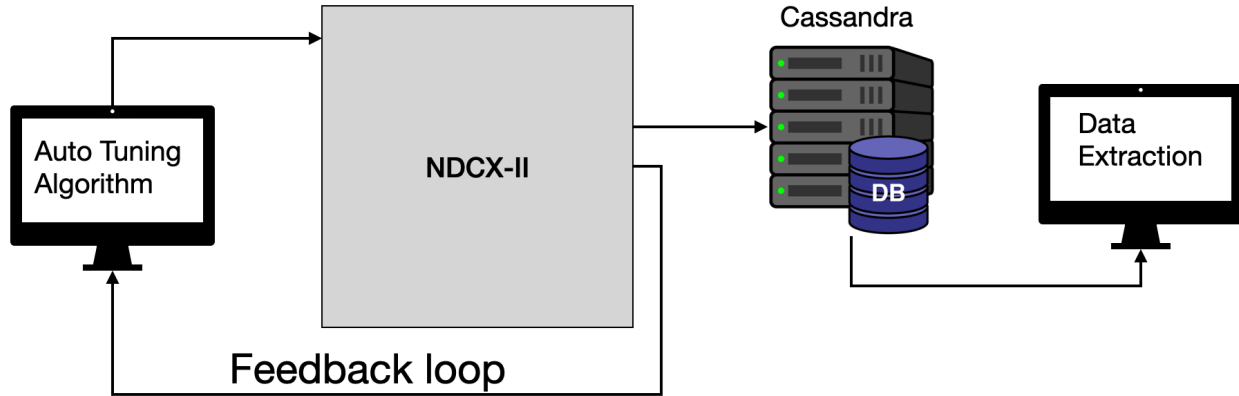


Figure 11.1: Graphical representation of the data collection and retrieval process for NDCX-II. Conductor settings are initialized and the accelerator is commanded through a Python interface at the computer. After each shot, the experimental data is stored into a shotnumber object in the Cassandra database. The database can be queried through Python interface and the data collected through the data classes shotnumber and shot.

in NDCX-II and resulting in no beam. This was verified by looking at the voltage waveform measured by the ion source which indicates a successfully extracted beam from the source. Once this initial screening for bad shots was done, a scatter plot was made for each column to visualize the outputs. This allowed us to quickly spot any outliers that we may have missed, track down the shot data, and determine if we should remove the shot or not.

After pre-processing the data we are left with 382 data points that will comprise the data—although incredibly low, this amount of data will serve the purposes outlined in the introduction. For a full ML model, more shots could/should be used in order to be a viable predictor. In the following chapter, we will build a fully connected NN to predict the peak voltage V_{pk} and time of voltage peak t_{pk} measured at the Fcup using V_{pk} and t_{pk} from the recorded waveforms on the conductors objects (solenoids, compression cells, and Blumleins). Fig. 11.2 shows the distribution for t_{pk} (left) and V_{pk} (right) for the data. A quick view of the data distribution raises a concern, there is appears to be a large number of data within a small width of time. This concern is valid, a quick analysis reveals that approximately 52%

of the training data is between $2.24 \mu\text{s}$ and $2.26 \mu\text{s}$. Within $2.24 \mu\text{s} < t_{\text{pk}} < 2.26 \mu\text{s}$, 40% of V_{pk} is below 6 kV and we can expect early on that any NN trained with this data set to struggle to predict values outside this range. Thus, we should expect clustering within this range of data and, depending on the associated V_{pk} values within this range, we might see clustering in both dimensions.

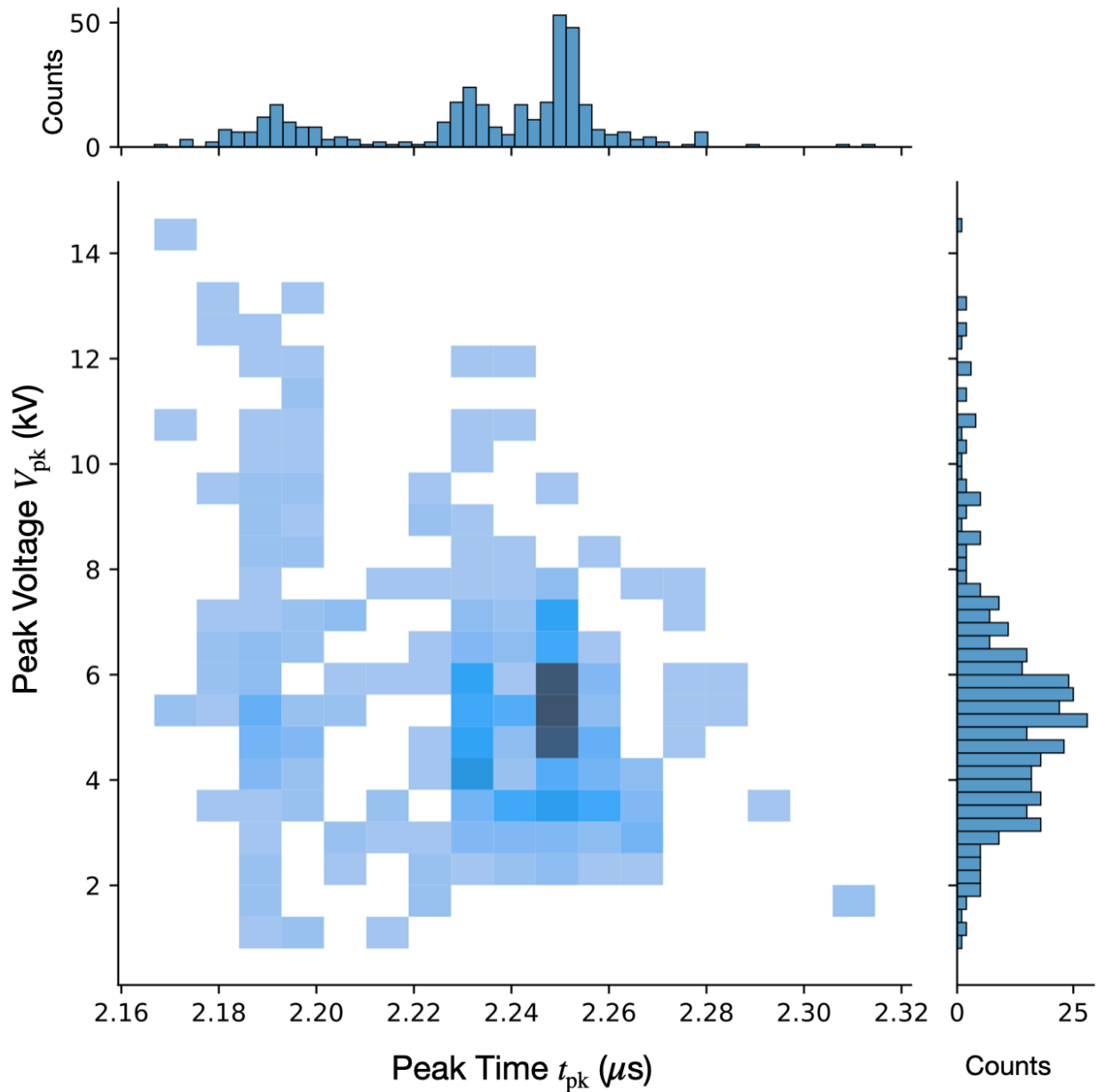


Figure 11.2: Visualization of the data extracted from NDCX-II. The main plot is a 2D histogram comprised of the individual histograms for t_{pk} (top) and V_{pk} (right).

Chapter 12. Training a Neural Network to Predict Peak Current and Arrival Time Using Detector Waveform Data

12.1 Neural Network Architecture

A common architecture when developing a neural network model is a fully connected dense neural network which is shown schematically in Fig. 12.1. The first layer is the input layer which will have n_x parameters that are to be used for training and the final output layer is shown as \hat{y} which will have n_y parameters that are to be predicted. Between the input and

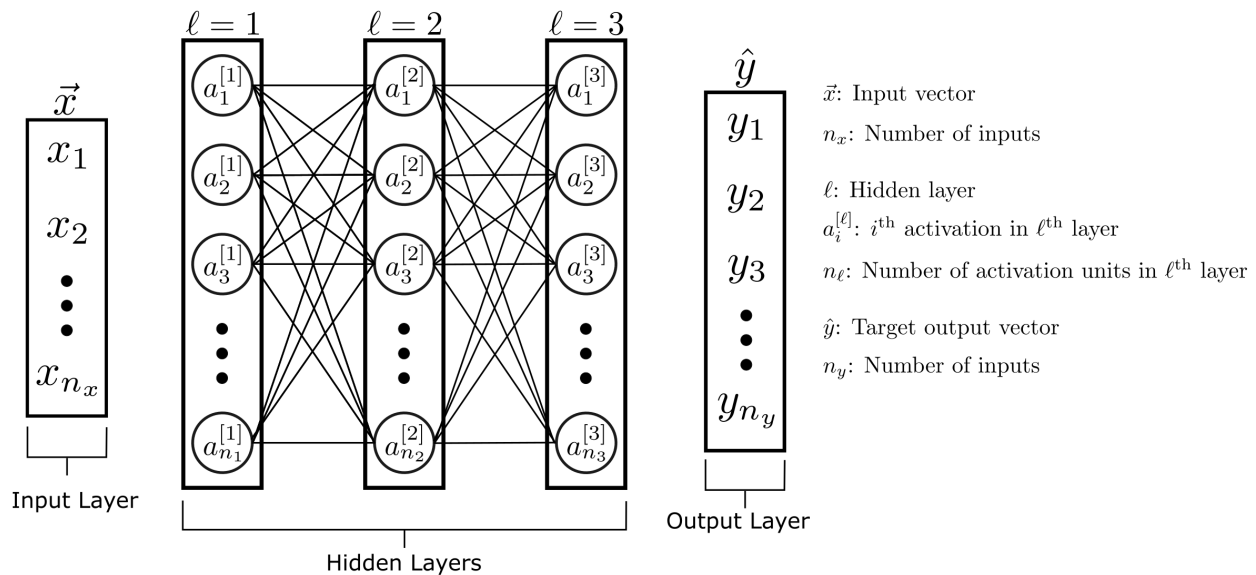


Figure 12.1: Schematic of a dense fully connected neural network architecture. We adopt the notation given by Andrew Ng’s Coursera course [46]. The inputs are given as a single column vector \vec{x} of which there are n_x entries or data samples. A single output layer is given by the vector \hat{y} containing n_y targets. The hidden layers are given by ℓ and each layer is comprised of n_ℓ neurons. Additionally, each layer has an activation function such as rectified linear unit (RELU).

output layers are the hidden layers denoted ℓ . Each hidden layer is comprised of activations or “neurons” $a_i^{[\ell]}$ and has an associated activation function per layer. Our model is more complex but uses this architecture as a building block. Instead of a single input layer feeding into hidden layers for all the conductor data, we instead treat the conductors with individual neural networks. For example, rather than combine peak voltage and time of peak voltage values for all conductors into a single input layer, we instead create three input layers comprised of the peak voltage and time of peak voltage for the respective conductor. These three layers are then used as inputs for respective individual fully connected three layer neural networks that will train on the data for each conductor separately. The final layers for the individual neural networks are then combined to make a single output layer that is used as the input for a fully connected two layer neural network that trains on all the data jointly. A schematic is shown in Fig. 12.2. The three input vectors are \vec{x}_{sol} for solenoids, \vec{x}_{cc} for compression cells, and \vec{x}_{bl} for Blumleins. Each individual NN then trains on the individual conductor peak voltage and time of peak voltage. This architecture allows for a high degree of variability in training since the hyper parameters for the individual NNs can be tuned for each conductor. Note that, although the a_i labels are the same in Fig. 12.2, they are different for each individual NN. The final layers of each individual NN ($\ell = 3$) are concatenated to form a single input layer \vec{x} which is then fed into full connected two layer NN before giving the prediction of the peak voltage and time of peak voltage of the Fcup.

To better highlight the difference, consider the input data coming from the seven compression cells alone. Lst. 12.1 is pseudocode to help illustrate the process. After collecting the data from a csv file, we create two dense hidden layers that both use the scaled exponential linear unit (SELU) activation function with a L2 regularizer. Hidden layer 1 has 100 activations while the second has 80. The last layer predicts two outputs using a linear

activation function. After creating the model, additional hyper parameters such as the batch size, number of epochs to train, the optimization function to be used, etc. are created and then the `cc_data` is used to fit the model. Now consider training on both the compression cells and the Blumleins. When specifying the inputs, we would have to calculate the shapes and then construct the layers accordingly. This does not seem like an issue at first sight. However, what if we wanted to specify a different activation function for the Blumleins? We could not do this by combining all the data into single input at the start. To circumvent this, we instead create the layers independently and then use TensorFlow [55] to concatenate the final layers from the individual networks into one at the end. Lst. 12.2 abbreviates how this might be done by adding the Blumlein data into the procedure from Lst. 12.1. The process is the same as the simple network, although now it is repeated for both conductors. We therefore, create two NNs for both conductors and the final output network is then fed into a NN that trains on all the both the compression cell and Blumlein data. Thus, we go from individual training to coupled training which allows more flexibility in architecture. This is helpful for projects with differing source distributions but that are being combined to predict an output.

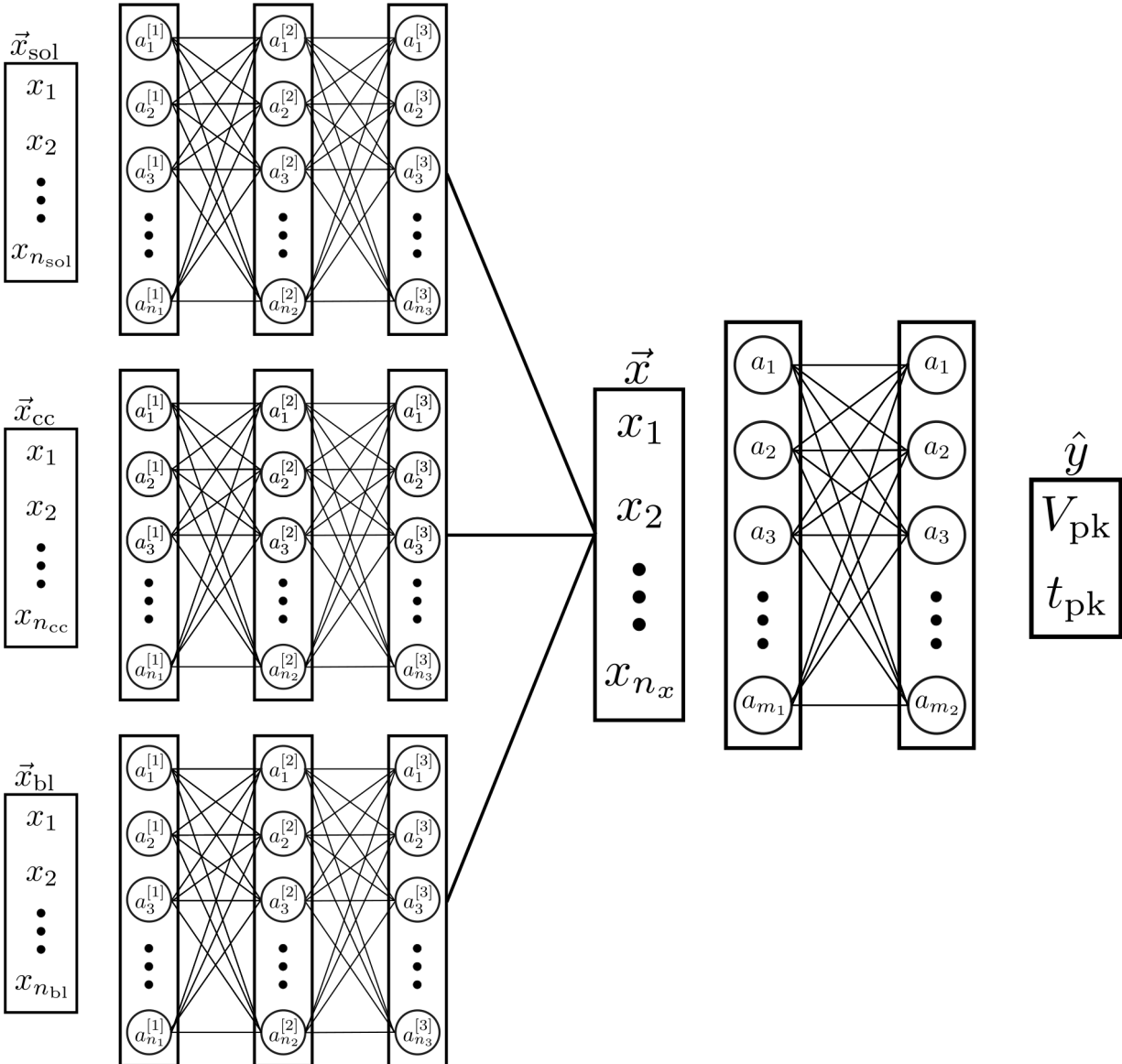


Figure 12.2: NN architecture using in training the extracted NDCX-II data. The input vectors are from the solenoids (\vec{x}_{sol}), compression cells (\vec{x}_{cc}), and Blumleins (\vec{x}_{bl}). Each input is fed into a three-layer neural network that trains on the conductor data alone which allows for more variability in training. The results are then fed into a two layer neural network which trains on all the data together before giving the output \hat{y} which is the peak voltage V_{pk} and time of the peak voltage t_{pk} .

```

1  import tensorflow as tf
2  import pandas
3  import numpy as np

```

```

4
5 data = pd.readcsv("path_to_data")
6 cc_data = np.zeros(N_data, 7) # Data matrix for 7 cells
7 # Loop through cc_data and assign data to column
8 for i in range(7):
9     # Assume cell data is under column cc_#
10    cc_data[:,i] = data[f"cc_{i+1}"]
11
12 # Specify input shape for model. Note cc_data.shape = (num samples, 7)
13 input_cc = tf.keras.Input(shape=(7,))
14
15 # Create layers
16 dense_l1 = tf.keras.layers.Dense(
17     100,
18     activation = "selu",
19     kernel_regularizer = tf.keras.regularizers.L2(l2=l2_weight),
20     )(input_cc)
21 dense_l2 = tf.keras.layers.Dense(
22     80,
23     activation = "selu",
24     kernel_regularizer = tf.keras.regularizers.L2(l2=l2_weight),
25     )(dense_l1)
26 dense_l3 = tf.keras.layers.Dense(2, activation="linear")(dense_l2)
27
28 # Create model
29 model = tf.keras.Model(
30     inputs = [input_cc],
31     outputs = [dense_l3],

```

```

32     name = "Simple Model"
33 )
34
35 # With created model we can then fit to the reshaped cc_data.

```

Listing 12.1: Pseudocode for constructing a simple NN from a single data source which is taken to be seven compression cells.

```

1  import ...
2  data = ...
3
4  # Specify input shapes for both now
5  input_cc = tf.keras.Input(shape=(7,))
6  input_bl = tf.keras.Input(shape=(5,))
7
8  # Create layers for cells same as before and add cc_ prefix to
   variable names
9  ...
10 cc_dense_l3 = tf.keras.layers.Dense(2, activation="linear")(dense_l2)
11
12 # Now create layers for Blumleins
13 bl_dense_l1 = tf.keras.layers.Dense(
14     200,
15     activation = "relu",
16     kernel_regularizer = tf.keras.regularizers.L2(l2=l2_weight),
17     )(input_bl)
18 bl_dense_l2 = tf.keras.layers.Dense(
19     150,
20     activation = "relu",

```

```

21     kernel_regularizer = tf.keras.regularizers.L2(l2=l2_weight),
22     )(bl_dense_l1)
23 bl_dense_l3 = tf.keras.layers.Dense(2, activation="linear")(dense_l2)
24
25 # Concatenate layers and then build complete model
26 dense_input = tf.concat([cc_dense_l3, bl_dense_l3], 1)
27
28 # Create layers to train on outputs from both conductor networks
29 dense_l1 = tf.keras.layers.Dense(...)(dense_input)
30 dense_l2 = ...(dense_l1)
31 dense_l3 = tf.keras.layers.Dense(2, activation="linear")(dense_l2)
32
33
34 # Create model
35 model = tf.keras.Model(
36     inputs = [input_cc, input_bl],
37     outputs = [dense_l3],
38     name = "Complex Model"
39 )
40
41 # With created model we can then fit to the reshaped cc_data.

```

Listing 12.2: Pseudocode for constructing a simple NN from a single data source which is taken to be seven compression cells.

12.2 Training and Validation

We split our data into a training and validation set using a 90:10 split (90% of the data is used for training while 10% used for validation). Typically, a 70:30 split is used but we chose a high percentage due to the low amount of data. This results in 382 samples to train on. We then normalized the data using the `StandardScaler` method provided by the Keras application programming interface (API) [56]. This transformation is done by subtracting the mean value of the data from each data point and then normalizing by the standard deviation of the whole dataset. That is, for some collection of data x , the transformed data \tilde{x} is given by

$$\tilde{x} = \frac{x_i - \bar{x}}{\sigma_x}, \quad (12.1)$$

where x_i is the i^{th} datapoint and σ_x is the standard deviation of x . Eq. 12.1 is applied to each t_{pk} and V_{pk} for each conductor individually, as well as the Fcup. The transformation is done over the training data and applied to the validation data. Care should be taken to not create a transformation that has used the validation data as this would result in data “leakage” and information from the validation set will “spill” into the training data. After the training is done, we apply the inverse transformation get the results in the laboratory units. Fig. 12.3 shows the distribution of t_{pk} and V_{pk} for the Fcup before the transformation (Panels A and B respectively) and the transformed distributions (Panels C and D respectively).

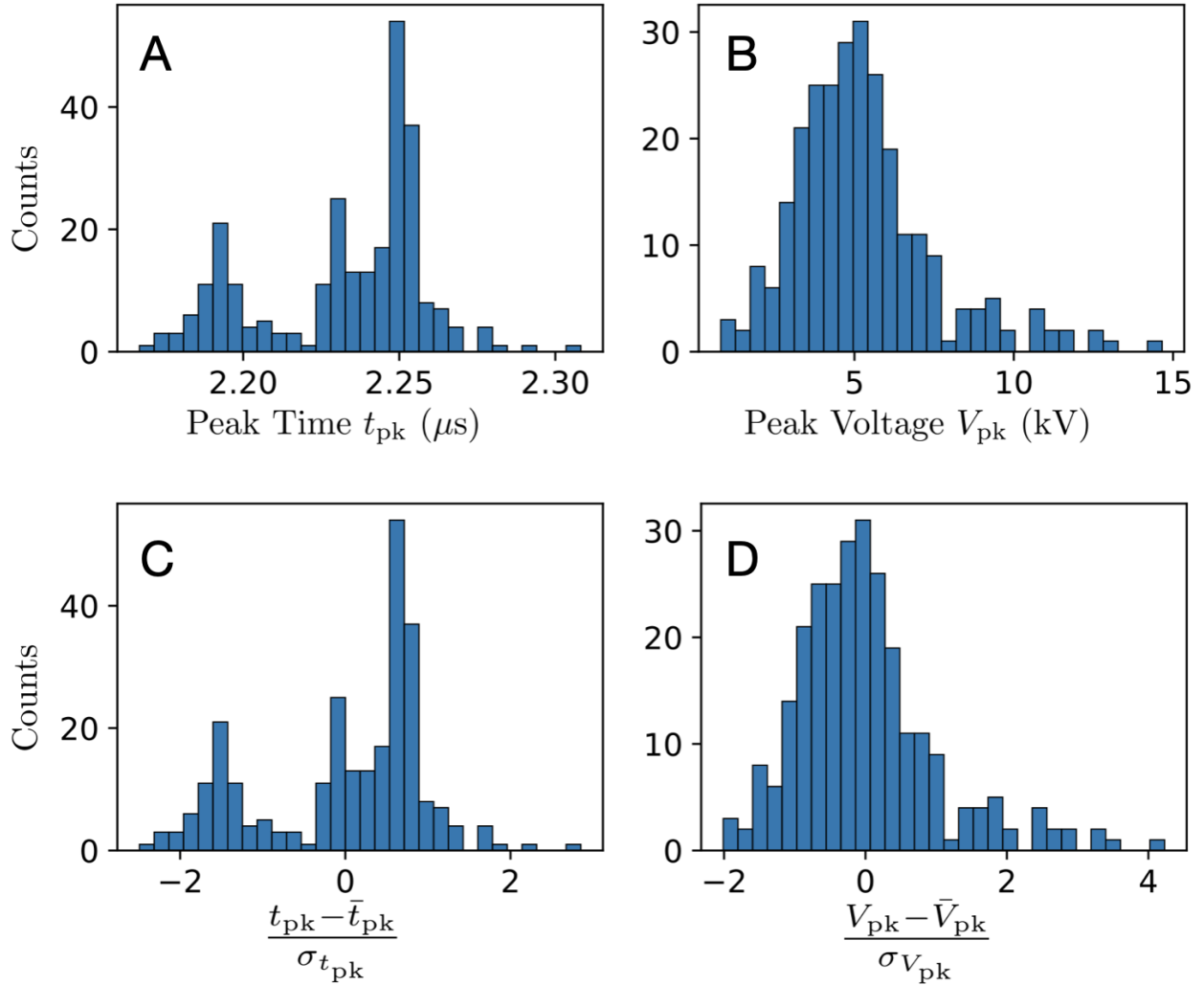


Figure 12.3: Training data t_{pk} (Panel A) and V_{pk} (Panel B) before applying the `StandardScaler` and the transformed data for t_{pk} (Panel C) and V_{pk} (Panel D) respectively.

Our NN architecture has many hyperparameters due to the complexity illustrated in Fig. 12.2. Following this schematic, we kept the number of neurons the same for each layer in the individual conductor NNs and used $n_1 = 70$, $n_2 = 50$, and $n_3 = 40$. Due to the training data being strongly clustered, we chose the weight of the L2 regularization to be small enough to be practically 0. This is due to the large clustering of data. With this clustering, valid data points will appear as outliers and a high regularization weight will create an insensitive NN. Since we wish to predict these values, choose to keep the NN

Architecture	
n_1, n_2, n_3	80, 60, 40
m_1, m_2	30, 20
Activation Functions	SELU for hidden layers, Linear for output
Regularizer, Weight	L2, 1×10^{-10}
Training	
Batch Size	150
Epochs	4500
Learning Rate	0.5×10^{-6}

Table 12.1: Summary of the various hyperparameters used which corresponds to the NN architecture shown in Fig. 12.2.

sensitive to outliers. For the activation function, we used the scaled exponential linear unit (SELU) function. After testing the exponential linear unit (ELU) function, rectified linear unit (RELU) function, and tanh function, we found that SELU better performed and was able to better predict values outside the problematic time bias present in the data. The individual conductor NNs were then concatenated to form an input vector that will feed into the final two layer NN which trains on all the data. We chose $m_1 = 30$ and $m_2 = 20$, and kept the same settings for the regularizer and activation functions. A batch size of 150 and trained over 4500 epochs which took about 2.5 minutes. We chose a learning rate of 0.5×10^{-6} . A summary of the hyper parameter settings is given in Table 12.1. These hyper parameters were tuned by hand and do not represent a global optimum. Lastly, we chose ADAM as our optimizing function [57] and chose the root-mean-square-error (RMSE) as our loss function to minimize. The training took approximately 3.5 minutes to complete with these settings and Lst 12.3 gives an outline of the algorithm in pseudo code at the end of the section.

To tune the hyper parameters, we used the average relative error over the predicted

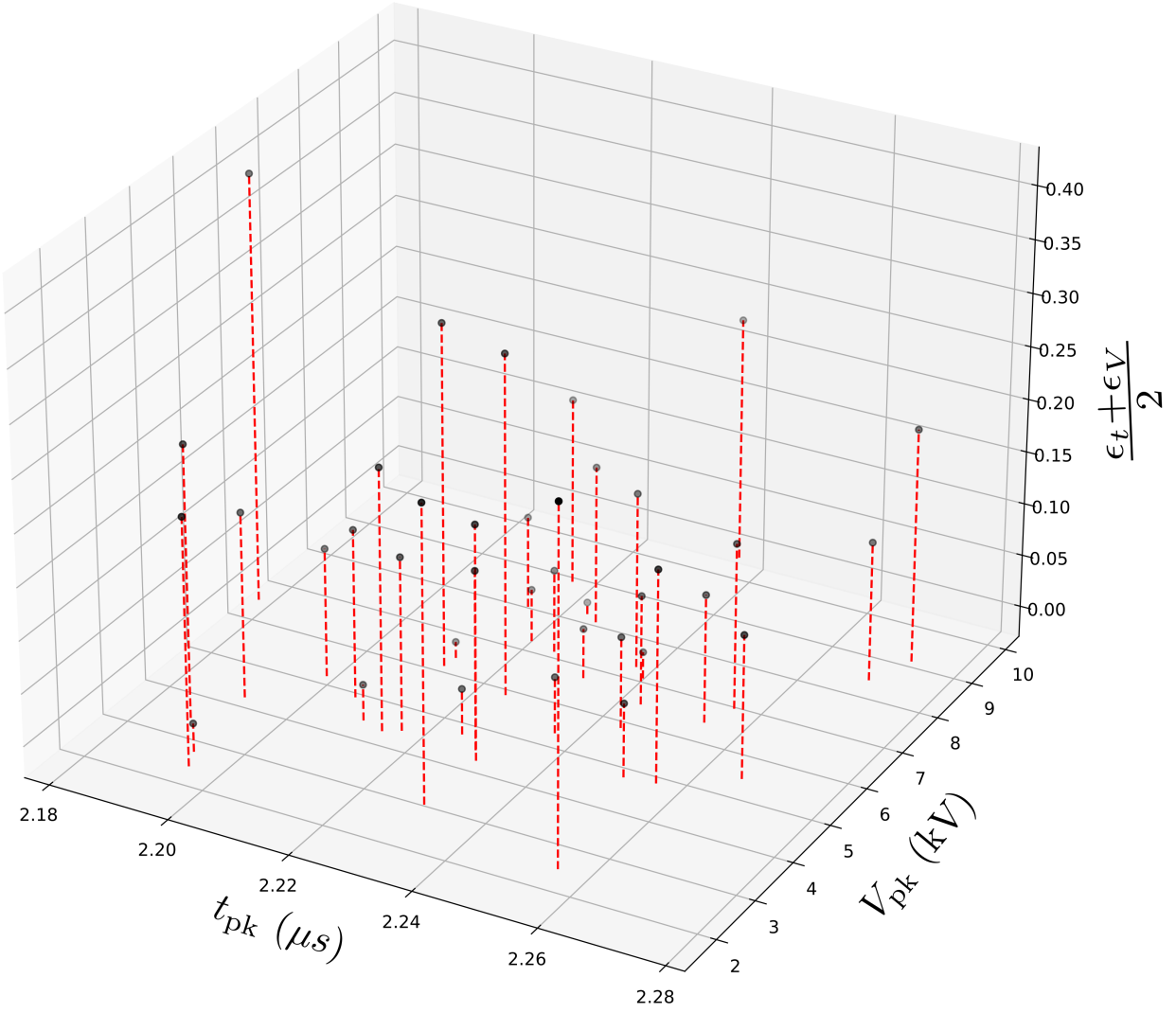


Figure 12.4: Calculated $\epsilon_{\hat{y}}$ for each (t_{pk}, V_{pk}) in the validation data.

values. The relative error ϵ_{ξ} of a parameter ξ is calculated by computing

$$\epsilon_{\xi} = \left| \frac{\xi_{\text{pred}} - \xi_{\text{actual}}}{\xi_{\text{actual}}} \right|, \quad (12.2)$$

where ξ_{pred} is the predicted value and ξ_{actual} is the actual value from the data. We seek to predict the t_{pk} and V_{pk} from the Fcup and so, we take as a metric the average of the

respective relative errors $\epsilon_{\hat{y}}$ calculated using

$$\epsilon_{\hat{y}} = \frac{\epsilon_t + \epsilon_V}{2}, \quad (12.3)$$

where ϵ_t and ϵ_V are calculated using Eq. 12.2 for t_{pk} and V_{pk} respectively. We used Eq. 12.3 to calculate the error for each of the validation points after training and plotted the results in Fig. 12.4. On the lower x and y axes are t_{pk} and V_{pk} for each of the validation points used. Extending in the vertical z direction is the calculated $\epsilon_{\hat{y}}$ for each point. Note that, even with the regularizer weight effectively set to 0, the outliers are difficult to get an accurate prediction. As expected, values near the cluster have relatively low error and data away from this region has the highest $\epsilon_{\hat{y}}$ values. We binned $\epsilon_{\hat{y}}$ using a bin width of approximately 0.05 to better see the distribution in Panel A of Fig. 12.5. From the distribution we see that only a handful of points have $\epsilon_{\hat{y}} > 0.25$ and roughly 40% of the validation points have $\epsilon_{\hat{y}} < 0.15$. Although not perfect, this is a good start for a model. Metrics other than $\epsilon_{\hat{y}}$ could be developed to help guide the tuning and, of course, additional data would alleviate the clustering problem and improve performance. Panel B shows the computed loss over the training and validation data for the 4500 epochs used. Although not the characteristic elbow (usually, the elbow marks a point where the loss is approximately constant for increasing steps), the marked point shows where the loss curves cross which suggests we may be slightly over training at this point. Decreasing the epochs to approximately 2800 does not result in a significant speedup in time and, since training a NN is an inherent stochastic process, this point is subject to change when retraining the model with the same data and the same settings. Thus, it is more beneficial to look for improvements elsewhere such as the learning rate, various settings for the ADAM optimizer, different loss functions, etc.

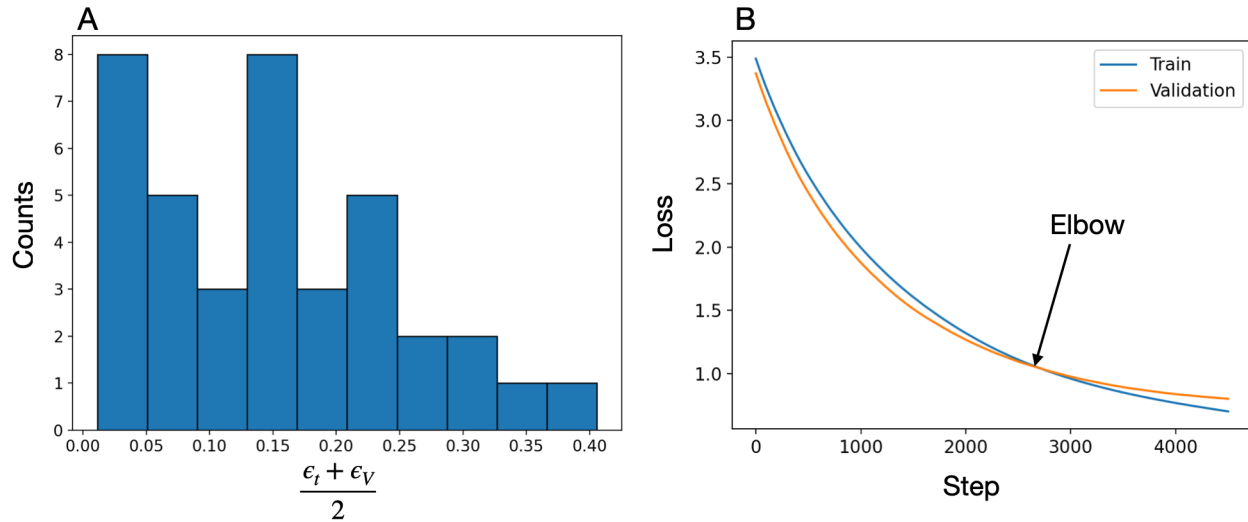


Figure 12.5: Distribution of the $\epsilon_{\hat{y}} = (\epsilon_t + \epsilon_V)/2$ over the validation set after training the NN is shown in Panel A. The bin widths are approximately 0.05. Panel B gives the RMSE loss score over the training data (blue) and validation data (orange) at each epoch. Although the validation loss does not quite plateau to zero change over a step, there is a crossing point between the training and validation curve that can be taken as the elbow.

```

1  import ...
2  data = ...
3
4  # Specify input shapes for conductors
5  in_sol = tf.keras.Input(shape=(56,)) # 2x28 parameters for t and V
6  in_bl = tf.keras.Input(shape=(10,)) # 2x5 parameters
7  in_cell = tf.keras.Input(shape=(14,)) # 2x7 parameters
8
9  # Regularizer weight
10 wL2 = 1e-10
11
12 # Set number of neurons. Will keep same for all conductors
13 an1 = 80
14 an2 = 60
15 an3 = 40

```

```

16
17 # Create layers
18 dense_S1 = layers.Dense(
19     an1,
20     activation="selu",
21     kernel_regularizer=tf.keras.regularizers.L2(l2=wL2),
22 )
23 dense_S2 = layers.Dense(
24     an2, activation="selu", kernel_regularizer=tf.keras.regularizers.
25     L2(l2=wL2)
26 )
27 dense_S3 = layers.Dense(
28     an3, activation="selu", kernel_regularizer=tf.keras.regularizers.
29     L2(l2=wL2)
30 )
31 # Layer creation is identical for Blumleins and compression cells
32 dense_B1 = layers.Dense(
33     an1,
34     activation="selu",
35     kernel_regularizer=tf.keras.regularizers.L2(l2=wL2),
36 )
37 dense_B2 = layers.Dense(
38     an2, activation="selu", kernel_regularizer=tf.keras.regularizers.
39     L2(l2=wL2)
40 )
41 dense_B3 = ...

```

```

41 dense_CC1 = ...
42 ...
43
44 # Combine layers into a single input layer
45 dense_NN = tf.concat([dense_S3, dense_B3, dense_CC3], 1)
46
47 # Create the NN that will train on all conductors
48 dense_NN1 = layers.Dense(
49     30, activation="selu", kernel_regularizer=tf.keras.regularizers.L2
50     (l2=wL2)
51     )(dense_NN)
52 dense_NN2 = layers.Dense(
53     20, activation="selu", kernel_regularizer=tf.keras.regularizers.L2
54     (l2=wL2)
55     )(dense_NN1)
56 dense_NN3 = layers.Dense(2, activation="linear")(dense_NN2)
57
58 # Create model
59 model_NDCX_data = keras.Model(
60     inputs=[in_sol, in_bl, in_cell],
61     outputs=[dense_NN3],
62     name="NDCX_II_data",
63 )
64 model_NDCX_data.summary()
65
66 # With created model, can now train
67 n_epochs = 4500
68 b_size = 150

```

```

67
68     opt = keras.optimizers.Adam(learning_rate=0.5e-6)
69
70     model_NDCX_data.compile(
71         loss="mse", optimizer=opt, metrics=[tf.keras.metrics.
RootMeanSquaredError()]
72     )
73
74     # Specify the input data and targets for the training set. Recall the
data is normalized based on a transformation found over the training
data and applied to the training and validation data
75     in_train = [normed_x_sol_train, normed_x_bl_train, normed_x_cell_train
]
76     in_train_target = [normed_peak_voltage_train, normed_peak_time_train]
77
78     # Specify the input data and targets for the validation set
79     in_val = [normed_x_sol_val, normed_x_bl_val, normed_x_cell_val]
80     in_val_target = [normed_peak_voltage_val, normed_peak_time_val]
81
82     # Fit the model to the input data
83     history = model_NDCX_data.fit(
84         x=in_train,
85         y=in_train_target,
86         batch_size=b_size,
87         epochs=n_epochs,
88         validation_data=(in_val, in_val_target),
89     )
90

```

```

91     # The loss curves can now be visualized through
92     fig, ax = plt.subplots()
93     ax.plot(history.history["loss"], label="Train")
94     ax.plot(history.history["val_loss"], label="Validation")
95     ax.set_xlabel("step")
96     ax.set_ylabel("loss")

```

Listing 12.3: Pseudocode for the NN architecture used to train the NDCX-II data.

12.3 Summary and Concluding Remarks

The goal of this chapter was to document the work that was done and to show how an ML project in accelerator physics might take form. We have implemented an auto tuning algorithm to track drift and quickly find optimal working conditions for an accelerator with a large set of tuning parameters and complex optimization landscape (highly correlated set of parameters). We also built a NN to predict key parameters at the target location from the input parameters. For a fully working ML interpolation one would use a larger data set, more input parameters and predict more key parameters (such as the width of the pulse or the capacity of the beam to heat a target). The resulting NN could then replace the long Warp runs and be used for scanning of the parameter space and in situ optimization.

This model, although trained on a relatively small data set, demonstrates how complex an ML can be. Fortunately, one of the more difficult aspects of an ML project was relatively easy in this project—data collection and preprocessing. This part in the project pipeline can be an endeavor on its own but, due to efforts from Dr. Persaud, the data storing and ultimate extraction is trivial. Our network architecture is not conventional but allows for a

high degree of variability when optimizing and tuning hyper parameters. Hyper parameter optimization is challenging and here we only optimized a select few. There are more settings that can be toggled with the optimizer, loss functions, and activation functions. Should the project receive more funding these hyper parameters would need to be investigated.

The current model can also be improved by using simulation data to supplement the training data. A full 3D simulation of NDCX-II exists using Warp. Preliminary results showed that data created in Warp was useful in training the NN. However, implementing Warp runs to create more training data was beyond the scope of this project, but seems like an interesting approach worth exploring in future studies. It would be interesting to incorporate synthetic data from Warp to help train the NN and see results should the project receive funding in the future. With an easy to use data encapsulation, a working full simulation in Warp, and work NN model to build off of, NDCX-II is even more appealing for testing and developing ML algorithms and processes that future accelerator and accelerator physicist can benefit from.

BIBLIOGRAPHY

- [1] A. W. Chao, K. H. Mess, M. Tigner, and F. Zimmermann, *Handbook of Accelerator Physics and Engineering*. WORLD SCIENTIFIC, 2nd ed., 2013.
- [2] A. Sessler and E. Wilson, *Engines of Discovery*. WORLD SCIENTIFIC, 2014.
- [3] R. W. Hamm and M. E. Hamm, “The beam business: Accelerators in industry,” *Phys. Today*, vol. 64, pp. 46–51, June 2011.
- [4] T. P. Wangler, *RF Linear Accelerators Second, Completely Revised and Enlarged Edition*. Wiley-VCH, 2008.
- [5] S. Lund, “Phy 905 lectures 09 longitudinal physics: Beam acceleration.”
- [6] S. Lund, “Longitudinal physics: Beam accelerator.” https://people.nsl.msu.edu/~lund/msu/phy905_2020/lec_lund/, 2020.
- [7] M. Reiser, *Theory and Design of Charged Particle Beams Second, Updated and Expanded Edition*. Wiley-VCH, 2008.
- [8] “United states particle accelerator school.” <https://uspas.fnal.gov/>.
- [9] “Cern document server.” <https://cds.cern.ch/?ln=en>.
- [10] S. Lund and J. Barnard, “Transverse particle dynamics.” https://people.nsl.msu.edu/~lund/uspas/bpisc_2020, 2020.
- [11] J. Barnard and S. Lund, “Introduction.” https://people.nsl.msu.edu/~lund/uspas/bpisc_2020/lec_set_01/01.introduction.pdf, 2020.
- [12] H. Wiedemann, *Particle Accelerator Physics*. Springer International Publishing, 2015.
- [13] S. M. Lund and B. Bukh, “Stability properties of the transverse envelope equations describing intense ion beam transport,” *Phys. Rev. ST Accel. Beams*, vol. 7, p. 024801, Feb. 2004.
- [14] J. Barnard and S. Lund, “Introduction.” https://people.nsl.msu.edu/~lund/uspas/bpisc_2020/lec_set_01/01.introduction.pdf, 2020.
- [15] W. Magnus, N. Y. U. I. of Mathematical Sciences, and U. S. A. F. O. of Scientific Research, *Hill’s Equation: General theory*. No. pt. 1 in United States. Air Force Office of Scientific Research Technical Note No, New York University, Institute of Mathematical Sciences, Division of Electromagnetic Research, 1957.

- [16] E. Courant and H. Snyder, “Theory of the alternating-gradient synchrotron,” *Annals of Physics*, vol. 3, no. 1, pp. 1–48, 1958.
- [17] 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 *2nd International Conference on High-Energy Accelerators*, pp. 274–287, 1959.
- [18] S. M. Lund, T. Kikuchi, and R. C. Davidson, “Generation of initial kinetic distributions for simulation of long-pulse charged particle beams with high space-charge intensity,” *Phys. Rev. ST Accel. Beams*, vol. 12, p. 114801, Nov. 2009.
- [19] P. M. Lapostolle, “Possible emittance increase through filamentation due to space charge in continuous beams.,” *IEEE Trans. Nucl. Sci.*, vol. 18, pp. 1101–1104, 1971.
- [20] F. J. Sacherer, “Rms envelope equations with space charge.,” *IEEE Trans. Nucl. Sci.*, vol. 18, pp. 1105–1107, 1971.
- [21] D. P. Grote, A. Friedman, J. Vay, and I. Haber, “The WARP Code: Modeling High Intensity Ion Beams,” *AIP Conference Proceedings*, vol. 749, pp. 55–58, 03 2005.
- [22] “Warp wiki home.” <https://warp.lbl.gov/home>.
- [23] S. Hanna, *RF Linear Accelerators for Medical and Industrial Applications*. Artech, 2012.
- [24] R. W. Hamm and M. E. Hamm, *Industrial Accelerators and Their Applications*. WORLD SCIENTIFIC.
- [25] S. Sinha, Y. Hou, D. Ni, Q. Ji, A. Persaud, P. Seidl, T. Schenkel, A. Lal, and K. K. Afridi, “A 27.12-MHz 10-kv power amplifier for compact particle accelerators utilizing an optimized matching network,” in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 5452–5457, ieeexplore.ieee.org, Oct. 2020.
- [26] K. B. Vinayakumar, S. Ardanuc, Q. Ji, A. Persaud, P. Seidl, T. Schenkel, and A. Lal, “Demonstration of waferscale voltage amplifier and electrostatic quadrupole focusing array for compact linear accelerators,” *Journal of Applied*, vol. 125, p. 194901, May 2019.
- [27] A. Persaud, Q. Ji, E. Feinberg, P. A. Seidl, W. L. Waldron, T. Schenkel, A. Lal, K. B. Vinayakumar, S. Ardanuc, and D. A. Hammer, “A compact linear accelerator based on a scalable microelectromechanical-system RF-structure,” *Rev. Sci. Instrum.*, vol. 88, p. 063304, June 2017.
- [28] “The future of compact high voltage power is here.” Retrieved May 1, 2021, from <https://www.airitytech.com/>.

- [29] Q. Ji, K. K. Afridi, T. Bauer, G. Giesbrecht, Y. Hou, A. Lal, D. Ni, A. Persaud, Z. Qin, P. Seidl, S. Sinha, and T. Schenkel, “Beam power scale-up in micro-electromechanical systems based multi-beam ion accelerators,” *Rev. Sci. Instrum.*, vol. 92, p. 103301, Oct. 2021.
- [30] S. Lund, “Phy 905 lectures 04.sup equations of motion and applied fields,” September 2018.
- [31] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*. Bristol: Hilger, 1988, 1988.
- [32] A. Wolski, “Maxwell’s equations for magnets,” 2019.
- [33] D. Neuffer, “Longitudinal motion in high current ion beams - a Self-Consistent phase space distribution with an envelope equation,” *IEEE Trans. Nucl. Sci.*, vol. 26, pp. 3031–3033, June 1979.
- [34] S. Linge and H. P. Langtangen, *Programming for Computations-Python: A Gentle Introduction to Numerical Simulations with Python 3.6*. Springer Nature, 2020.
- [35] S. M. Lund, S. H. Chilton, and E. P. Lee, “Efficient computation of matched solutions of the Kapchinskij-Vladimirskij envelope equations for periodic focusing lattices,” *Phys. Rev. ST Accel. Beams*, vol. 9, p. 064201, June 2006.
- [36] F. Gao and L. Han, “Implementing the Nelder-Mead simplex algorithm with adaptive parameters,” vol. 51, no. 1, pp. 259–277.
- [37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [38] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. USA: Cambridge University Press, 3 ed., 2007.
- [39] R. Courant, K. Friedrichs, and H. Lewy, “On the partial difference equations of mathematical physics,” *IBM J. Res. Dev.*, vol. 11, pp. 215–234, Mar. 1967.
- [40] S. Humphries, *Charged particle beams*. Mineola, New York: Dover, 2013.

- [41] W. L. Waldron, W. J. Abraham, D. Arbelaez, A. Friedman, J. E. Galvin, E. P. Gilson, W. G. Greenway, D. P. Grote, J.-Y. Jung, J. W. Kwan, M. Leitner, S. M. Lidia, T. M. Lipton, L. L. Reginato, M. J. Regis, P. K. Roy, W. M. Sharp, M. W. Stettler, J. H. Takakuwa, J. Volmering, and V. K. Vytla, “The NDCX-II engineering design,” *Nucl. Instrum. Methods Phys. Res. A*, vol. 733, pp. 226–232, Jan. 2014.
- [42] A. Edelen, C. Mayes, D. Bowring, D. Ratner, A. Adelman, R. Ischebeck, J. Snuverink, I. Agapov, R. Kammering, J. Edelen, I. Bazarov, G. Valentino, and J. Wenninger, “Opportunities in machine learning for particle accelerators,” *TBD*.
- [43] P. A. Seidl, J. J. Barnard, E. Feinberg, A. Friedman, E. P. Gilson, D. P. Grote, Q. Ji, I. D. Kaganovich, B. Ludewigt, A. Persaud, C. Sierra, M. Silverman, A. D. Stepanov, A. Sulyman, F. Treffert, W. L. Waldron, M. Zimmer, and T. Schenkel, “Irradiation of materials with short, intense ion pulses at NDCX-II,” *Laser Part. Beams*, vol. 35, pp. 373–378, June 2017.
- [44] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, “A high-bias, low-variance introduction to machine learning for physicists,” *Physics Reports*, vol. 810, 2019.
- [45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [46] A. Ng, Y. Mourri, and K. Katanforoosh, “Neural networks and deep learning,” in *Deep Learning Specialization*, Coursera, 2023. DeepLearning.AI.
- [47] J. Portilla, “Python for machine learning and data science masterclass,” Udemy, 2023. Pierian Training.
- [48] A. Geron, *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. Sebastopol, CA: O’Reilly Media, 2017.
- [49] F. Chollet, *Deep Learning with Python*. Manning, 2017.
- [50] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.
- [51] A. Scheinker and M. Krstic, “Extremum seeking with bounded update rates,” *Syst. Control Lett.*, 2014.
- [52] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and A. Lutman, “Demonstration of Model-Independent control of the longitudinal phase space of electron beams in the Linac-Coherent light source with femtosecond resolution,” *Phys. Rev. Lett.*, vol. 121, p. 044801, July 2018.

- [53] A. Scheinker, D. Bohler, S. Tomin, R. Kammering, I. Zagorodnov, H. Schlarb, M. Scholz, B. Beutner, and W. Decking, “Model-independent tuning for maximizing free electron laser pulse energy,” *Phys. Rev. Accel. Beams*, vol. 22, p. 082802, Aug. 2019.
- [54] T. pandas development team, “pandas-dev/pandas: Pandas,” Feb. 2024.
- [55] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- [56] F. Chollet *et al.*, “Keras,” 2015.
- [57] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.