Title

Probing Ultrafast Dynamics With Relativistic Electrons

Author

Cesar, David Borden

Publication Date

2019

UNIVERSITY OF CALIFORNIA
Los Angeles

Probing Ultrafast Dynamics With Relativistic Electrons

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

by

David Borden Cesar

2019

ABSTRACT OF THE DISSERTATION

Probing Ultrafast Dynamics
With Relativistic Electrons

by

David Borden Cesar
Doctor of Philosophy in Physics
University of California, Los Angeles, 2019
Professor Pietro Musumeci, Chair

Ultrafast science is one of the frontiers of modern physics. It allows us to explore intense, out-of-equilibrium process by manipulating laser pulses shorter than the thermal time-scale of materials. Here we extend that capability to MeV electron beams by using the high extraction field of a radio-frequency photoinjector to generate dense bunches of electrons. This gives us a unique source which we take advantage of to explore two different systems: the first is a photonic structure called a dielectric laser accelerator (DLA), and the second is an accelerator based version of an ultrafast electron microscope (UEM).

DLA is an advanced accelerator technology which takes advantage of the high damage threshold of dielectrics (as compared to metals) to sustain GV/m accelerating gradients. It works on the same principle as a conventional RF linac, but it is scaled down 100,000 times from microwave frequencies to optical frequencies. This scaling has important consequences for the beam dynamics of the accelerator, and it leads us to consider a richly nonlinear system in which the stable accelerating region is surrounded by areas of chaos. In order to interrogate the dynamics of this system, we have adapted the beam from an RF photoinjector to fit inside the sub-micron aperture of a DLA. We then perform time-resolved spectroscopic measurements to determine the interaction strength of the accelerator. We record accelerating fields of up to 1.8 GV/m and energy gains as high as 315 keV, but we also find that self-phase modulation can cause dephasing if left uncompensated. Our analysis of the dephasing mechanism, and its compensation, leads us to the design of a DLA with all-optical control of the beam dynamics.

MeV UEM is a branch of ultrafast microscopy which is designed to achieve high spatio-temporal resolution in a single-shot. It requires placing billions of electrons inside a small volume of phase space in order to detect contrast from weakly scattering objects. We show that this is possible using relativistic electrons from an RF photoinjector, but because these electrons are stiff we have to replace the conventional microscope optics with strong permanent magnet quadrupole (PMQ) lenses. We have tested these lenses in a two-stage objective-projector setup and measured a total magnification of 900x. When operating this microscope near the design current of 200 mA we observe a strong distortion of the image. Our analysis suggests that the distortion is caused by nonlinear phase-space correlations from the space-charge kick of a Gaussian bunch. We discuss ways to improve the design and obtain 10 nm-10 ps resolution.

The dissertation of David Borden Cesar is approved.

Brian C. Regan

Eric R. Hudson

James Rosenzweig

Pietro Musumeci, Committee Chair

University of California, Los Angeles

2019

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I joined PBPL, almost by accident, eight years ago, after I asked my physics 105A professor if I could work in his lab. At the time I didn't know anything about lasers or accelerator physics, but Pietro promised me that the lab would offer ample opportunity to alternate between theory, computation, and hands-on experimental work. Since that day, the people of this lab have invited me in, taught me how me how to be a physicist, and inspired me to believe in our work. For that I owe thanks to a great many people, and it is my pleasure to be able to list many of them here.

First and foremost, I would like to express my gratitude to the people who spent countless hours with me in UCLA's basement, listening to the clack-clack-clack-clack of the Pegasus Klystron. I would never have been able to complete this research without your guidance, your patience, and your company. Thank you Joel, Kent, Jared, Jorge, and Pietro. During my time here, I have also shared Pegasus with Renkai, Josh, Cheyne, Eylene, Xinglai, Paul, Eric, and especially Emma. I thank each of you for contributing to our beamline and for inspiring me to get the most out of my time in the bunker.

Next I would like to acknowledge all the help I have received while building my experiments. Hoss, Evan, Sean, Giacomo, Edmund and Andy have all provided invaluable work drawing parts in Solidworks, soldering feedthroughs, crimping motors, adding to our Labview controls and more. Andy, in particular, is responsible for a great deal of the two-stage ultrafast electron microscope. I should also like to thank Edgar Peralta, though I have never met him, for leaving us DLA structures to test at Pegasus. Similarly, I should like to thank Radiabeam for making us PMQs to test in the ultrafast microscope. It is only due to these collaborations that I have been able to participate in such a rewarding set of projects.

I also want to thank the many physicists who have graciously shared their thoughts with me over the last several years. Many of these discussions have started in the back corner of the office where I sit, so allow me to acknowledge Gerard, Emma, Joe, Nick, Ivan, and Youna. Thank you all for allowing me to interrupt you, be it for darts or physics. I will always value your feedback on my work. Next, I thank the PBPLers from this era who have kept me company in lab, at conferences, and during group lunch. Thank you Ryan, Nathan, Claudio, Siara, Xinglai, Yusukue, Oliver, Brian, Jamie, Paul, Eric, Gerard, Josh, and Walter. You have all helped contribute to the culture that makes PBPL a great place to work. I will cherish our late night arguments, our intense foosball matches, and of course the barbeques, hotpot, tailgates, and Dodger games.

Going outside of PBPL, I would like to thank my ACHIP colleagues Uwe, Willi, Andrew, Dylan, Alex, Ken, Josh, and Martin, among others for making our collaboration meetings a worthwhile experience. And outside of accelerator physics, I would like to give thanks to my roommates Arjun and Paul for helping me study for the comp and for keeping our apartment quiet and peaceful.

Finally I would like to thank my adviser, Pietro, for shaping me into the scientist I am today. For the last eight years he has held his door wide open and graciously engaged with every idea I could think to bring to him. I am grateful to have had the chance to work for such a dedicated adviser.

VITA

2011	Undergraduate researcher, Particle beam physics laboratory (PBPL), UCLA.
2013	B.S. (Physics) UCLA, Los Angeles, California.
2013–2014	Teaching assistant, Physics and Astronomy, UCLA. Taught 18L, a modern physics laboratory for physics majors
2015	M.S. (Physics), UCLA, Los Angeles, California.
2014–2019	Graduate student researcher, Physics and Astronomy, UCLA.

PUBLICATIONS

D. Cesar and P. Musumeci, “Transverse Beam Dynamics in Plasma Guided Laser Driven Acceleration.pdf,” *Proc. PAC 2013*, 2013

D. B. Cesar, P. Musumeci, and D. Alesini, “Ultrafast gating of a mid-infrared laser pulse by a sub-pC relativistic electron beam,” *Journal of Applied Physics*, vol. 118, no. 23, p. 234506, Dec. 2015.

A. Ody, P. Musumeci, J. Maxson, D. Cesar, R. J. England, and K. P. Wootton, “Flat electron beam sources for DLA accelerators,” *Nucl. Instrum. Methods Phys. Res. A*, vol. 865, pp. 75–83, 2016, Proc. of HBB 2016.

D. Cesar et al., “Demonstration of Single-Shot Picosecond Time-Resolved MeV Electron

Imaging Using a Compact Permanent Magnet Quadrupole Based Lens," *Phys. Rev. Lett.*, vol. 117, no. 2, p. 024801, Jul. 2016.

K. P. Wootton et al., "Dielectric laser acceleration and focusing using short-pulse lasers with an arbitrary laser phase distribution," *AIP Conference Proceedings*, vol. 1812, no. 1, p. 060001, 2017, Proc. of AAC 2016.

J. Maxson, D. Cesar, G. Calmasini, A. Ody, P. Musumeci, and D. Alesini, "Direct Measurement of Sub-10 fs Relativistic Electron Beams with Ultralow Emittance," *Phys. Rev. Lett.*, vol. 118, no. 15, p. 154802, Apr. 2017.

P. Musumeci, D. Cesar, and J. Maxson, "Double-shot MeV electron diffraction and microscopy," textitStruct Dyn, vol. 4, no. 4, May 2017.

D. Cesar, J. Maxson, P. Musumeci, X. Shen, R. J. England, and K. P. Wootton, "Optical design for increased interaction length in a high gradient dielectric laser accelerator," *Nucl. Instrum. Methods Phys. Res. A*, Feb. 2018. Proc. of EAAC 2017.

D. Cesar et al., "High-field nonlinear optical response and phase control in a dielectric laser accelerator," *Commun. Phys.*, vol. 1, no. 1, p. 46, Aug. 2018.

D. Marx et al., "Single-shot reconstruction of core 4D phase space of high-brightness electron beams using metal grids," *Phys. Rev. Accel. Beams*, vol. 21, no. 10, p. 102802, Oct. 2018.

D. Cesar et al., "Enhanced energy gain in a dielectric laser accelerator using a tilted pulse front laser," *Opt. Express*, vol. 26, no. 22, pp. 29216–29224, Oct. 2018.

CHAPTER 1

Introduction

Accelerators have been part of the fabric of basic science research since at least 1932 when Cockroft and Walton built an 800 keV electrostatic accelerator and split lithium atoms [2]. Their unique ability to create large numbers of high-energy particles has given physicists a tool to explore the microscopic constituents of matter and the fields that bind them. Over time, the intense scientific demand for more particles, higher energies, and better quality beams has fueled a history of continuous development in which advances in accelerator technology proceed side-by-side with discoveries in basic physics.

In the 1960's researchers realized that bending electrons make for a natural source of high brightness x-rays, and soon synchrotron sources started spurring the growth of crystallography and absorption spectroscopy [3]. As a result, the scope of accelerator physics expanded to include machines whose primary purpose is creating images of materials and their dynamics. Fast forward to the present day and accelerators have become an invaluable tool for the imaging community. Propelled along by the invention of the x-ray free electron laser (XFEL) and the advent of MeV electron microscopy, accelerator based techniques have become an invaluable tool for ultrafast imaging.

In addition to serving other sciences, accelerator physics has a rich history as a discipline in its own right. Its scientific inquiry is aimed at understanding the evolution of a classical ensemble of particles and trying to find stable, or at least quasi-stable, arrangements for the beam during nonlinear acceleration and radiation process. The basic physical laws governing these systems are well known, and they can, with the aid of modern computers, be simulated in intricate detail. This has been a boon to machine development and optimization, but what room does this engineering leave for the accelerator physicist? Today, as in the past, the role of physics is invaluable in reducing the complex dynamics of an accelerator into bite-sized discussions about optics, instabilities, growth rates, scaling laws, and other physical prototypes which can help to advance our intuition and lead us to the invention of new machines.

In this thesis we study some of the accelerator physics which can be done with an ultrafast electron beam from an RF photoninjector. We operate the injector with relatively low bunch charge so that we can achieve the low emittance needed to probe microscopic structures without sacrificing the high current needed to study ultrafast dynamics. This is ideal for injecting electrons in a dielectric laser accelerator (DLA), and also for performing single-shot ultrafast electron microscopy (UEM). Each of these projects has its own motivation which I will address sequentially. I will start by providing context for DLA, then I will do the same for ultrafast electron microscopy (UEM), and finally I will close the chapter with an introduction to the Pegasus laboratory where our work takes place.

1.1 Dielectric laser acceleration

Dielectric laser acceleration (DLA) describes a particular style of "advanced accelerator" which uses a dielectric structure to reshape the fields of an ultrafast laser into an evanescent slow-wave mode which can be used for high-gradient acceleration [4]. It is an optical-scale version of the conventional RF linac [5] with the substitution of dielectric for metallic boundaries in order to sustain GV/m gradients without breakdown. Like other advanced accelerators, it promises to be smaller, cheaper, and more widely available than its metallic forebearers.

The main attraction of DLA is its ability to use high-power, high-repetition rate, high-efficiency laser systems. Lasers can easily focus their energy into a microscopic volume with enough ferocity to ablate matter, but paradoxically they have difficulty transferring energy to a relativistic beam in free space. The essential difficulty is captured in an infamous no-go theorem which states that since particles move slower than c they can never move synchronously with a photon. Thus, any energy an electron gains from a plane wave will be taken away when it slips to a decelerating phase [6, 7]. Most advanced accelerator schemes can be classified by how they beat the Lawson-Woodward theorem: the inverse free electron laser uses a static B-field, the plasma wake-field schemes immerse the beam in plasma, the inverse Cerenkov scheme immerses the beam in a dielectric, and DLA (analogous to inverse Smith-Purcell radiation) uses evanescent waves. The proliferation of "inverse" schemes reveals a pleasing symmetry: all accelerators can be radiators and vice-versa.

If beating the no-go theorem is the first piece of the advanced accelerator puzzle, then the next is managing the ferociously intense electric fields required for high gradient acceleration. It requires a certain audacity to believe that we can take command of one of the most intense forces available in a laboratory, but this is exactly what we do in DLA. We are able to use diffractive optics to precisely shape the forces inside the structure and harness GV/m fields for acceleration. The resulting electron motion is high oscillatory, strongly coupled, and very nonlinear, but it can be stable. One of the joys of DLA is being able to study the rich dynamics of this system and to appreciate how we can limit the destructive power of the laser by balancing it with itself.

The visceral appeal of these high-intensity laser-electron interactions is what attracts many of us to work on advanced accelerators, but the motivation given for funding them is primarily economic: the large accelerator facilities that drive high energy physics and serve as x-ray light sources require kilometers of underground tunnels and cost billions of dollars. Even a small facility like Pegasus is relatively expensive for a university laboratory, and so only a few such labs exist. This means that the scope of science done with electron beams is heavily constrained by cost: from who has access, to what they can study, and how much beam time they can get. Advanced accelerators are funded on the premise of making a GeV cheap enough to build them in bulk. Research into DLA is expressly attached to the idea of making accelerators so small and robust that they will become a widely available tool for industry, medicine, and discovery science[1].

In addition to the economic argument, advanced accelerators hope to add capabilities that conventional accelerators can't have. DLA, for example, intrinsically bunches electrons

---

[1] Bob Byer, a PI of the ACHIP collaboration, once introduced DLA by saying that it will "democratize the accelerator", meaning that it will do for accelerators what the diode and YAG did for lasers.

on a sub-optical time-scale and so is a promising source of attosecond electron probes [8]. Making such a source widely available could help drive research in the burgeoning field of ultra-fast microscopy.

In practice, no advanced accelerator technology is currently mature enough to deliver on these promises [2]. In theory, DLA should work just as well as its successful cousin, the RF linac, but only recently have advances in nanofabrication, fiber lasers, and electron sources made building a DLA seem feasible. Thus, despite being one of the original advanced accelerators [13, 14, 15], DLA now lags well-behind its counterparts. A 13.5 million dollar grant from the Gordon and Betty Moore foundation created the Accelerator On a Chip (ACHIP) collaboration (of which UCLA is a member) [16] to address that gap. The stated goal of the collaboration is to develop a nanofabricated accelerator driven by a fiber coupled laser source and capable of accelerating electrons from a few KeV to at least 1 MeV.

This is an ambitious goal which aims to take DLA all the way from its first proof-of-principle experiments to an operational accelerator. For now, however, most DLA experiments have a more modest goal: measuring the fields produced inside the DLA. In a conventional RF linac, this would be done by observing the displacement of a bead as it is dragged through the cavity [17, 18], but the aperture of a DLA is much too small to probe in this way. Instead, we have to probe the DLA through a sort of "time resolved energy loss electron spectroscopy" in which we observe the displacement (or energy change) of *electrons* which we have managed to fit into the sample and synchronize with the laser fields. We often call these "acceleration experiments", but in reality we are measuring an increase in the energy spread of the electrons and then comparing it to simulations in order to judge if the structure performed as expected. This type of measurement really has more in common with ultrafast microscopy techniques [19, 20, 21] then with actual accelerator design.

The first experiments to convincingly measure energy exchange in a DLA were published simultaneously in 2013, and these quickly became a template for further experimentation. They both study the interaction of electrons with laser driven diffraction gratings, but one

---

[2] The FEL could be considered an exception since it often used for beam manipulation [9, 10, 11], and even as an accelerator it is stable enough to be used as a regular beamline component [12].

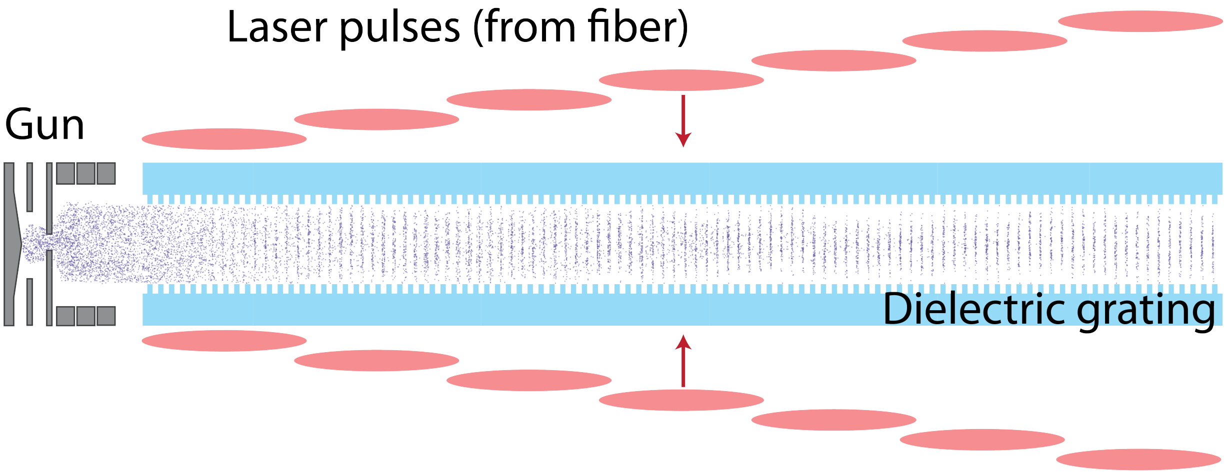

Figure 1.1: Cartoon of a DLA accelerator. A kV tip source feeds a DLA which bunches and focuses the beam. The laser pulses would be delivered by a fiber and timed to accelerated a slice of the beam along the DLA.

of the papers [22] measures highly relativistic (60 MeV) electrons, while the other measures sub-relativistic electrons (28 keV) [23]. As expected, they both find that the electron energy spread increases linearly with the input electric field, but only when the input laser is polarized along the acceleration axis and the phase velocity is perfectly matched to the electron beam. This rules out a number of ponderomotive or electron scattering effects, and suggests that they are indeed observing the DLA mechanism. The papers conclude by calculating an effective acceleration gradient of 250 MV/m in the relativistic case and 25 MV/m in the sub-relativistic case.

The dichotomy between relativistic and sub-relativistic experiments continues through to today. The difference originates from the observation that it is easier to slow the phase velocity of light by a fraction of a percent than it is to slow it to $\beta = 1/10^{\text{th}}$. As a result, sub-relativistic structures have turned from fused silica to silicon [24, 25], which makes up for a low damage threshold with a high index of refraction. The newest versions are now capable of producing a few hundred MV/m [24, 26]. Relativistic experiments, on the other hand, have taken a more direct route to higher gradients. Using the same structure [27] as in the original demonstration, the relativistic experiments have increased the gradient by lowering the laser pulse length from 1 ps [22] to 90 fs [28] and finally to the 45 fs laser here at Pegasus [29]. The shorter laser pulse allows for higher electric fields before damage, and thus pushes the effective gradient up to 850 MV/m (and the peak gradient all the way up to 2 GV/m).

As it has become clear that DLA structures can produce the gradients expected of an advanced accelerator technology, experimentation has begun a shift towards measuring and controlling the beam dynamics in a DLA. There are two approaches to this: one is to modify the DLA structure and the other is to modify the drive laser. This latter technique is rather unique to DLA: it is only because we build highly broadband structures, capable of supporting an ultra-fast laser pulse, that the drive laser is able to influence the beam dynamics. We demonstrated this to great effect in [29] by adding a chirp to the drive laser in order to compensate for de-phasing due to the nonlinear Kerr effect.

Exploring beam dynamics requires building and powering longer DLA structures than are used for gradient measurements. To this end, we have introduced a technique (called pulse front tilt or PFT) which allows us to extend the interaction length by more than 50 times [30, 31]. At sub-relativistic energies this also requires modifying the structure to have a tapered phase velocity to match the accelerating electron velocity [32]. But the main difficulty in making ever longer accelerators is that the beam dynamics in a DLA are inherently unstable.

The instability arises from a "saddle-point" type potential that is a general feature of vacuum based acceleration schemes. It can be seen as a consequence of either Earnshaw's theorem [33], or of the Panofsky-Wenzel theorem [34], but in either case it poses a major challenge for DLA. In order to stabilize the beam envelope we will have to add some new force. In a conventional accelerator we would add external magnetic focusing, but for DLA this is several orders of magnitude too weak. Consequently we have no choice but to focus using the field of the drive laser itself [35].

This leads to increasingly complicated structure designs [36, 37] which may become difficult to fabricate. In order to make a more flexible device, it should be possible to control the dynamics with a programmable laser phase [38]. Either way, the resulting dynamics lie in a richly nonlinear part of the accelerator physics cannon which is full of tune spreads, decoherence, coupled-motion-resonances and more. The result is a fascinating system which borders on chaos as it accelerates a carefully confined beam up to high energy.

Another challenge that comes with shrinking down an accelerator is that the dynamic aperture of the device decreases drastically. This makes it hard to fit any electrons into a DLA, let alone enough to load the wakefield and operate efficiently. In order to make up for this, a DLA can run at MHz repetition rates to increase flux, and then re-cycle unused laser energy to increase efficiency [39]. Even this may not produce a useable amount of electron flux, so there is ongoing research into developing more suitable electron sources. The leading candidates are arrays of tip based sources [40] and magnetized beams which can be flattened with skew quadrupoles [41]. What both solutions have in common is that they take advantage of the planar nature of DLA structures to fit more charge through the device. Whether this is suitable for the potential applications of a DLA accelerator remains to be seen.

The final piece of our collaboration's plan to rehabilitate DLA is to integrate this accelerator onto a "chip". Here "chip" is meant to invoke the scale and precision of industrial semiconductor techniques. This will require developing a whole new suite of technology to couple the laser into the structure and then divide and distribute it to the accelerator. So far, much of this work has been preliminary testing and conceptual designs [42, 43, 44, 45], but it is absolutely essential to the final utility of DLA, since without a robust, scalable, production-level device it is unlikely that DLA will become a widely used technology.

Compared to other advanced acceleration techniques, the main attraction of DLA is its potential to become a laboratory-scale device. DLA lags far behind the GeV scale energy gains achieved by of plasma techniques [46, 47, 48] and can't support anywhere near the beam power of THz dielectric wakefield experiments [49]. Nor can it trap and accelerate or manipulate an entire electron beam in the way that IFEL can [10, 11, 9]. But it does have the potential to be a truly small, "shoebox size", laboratory source of relativistic, attosecond, high-brightness electrons.

Such a source would be an obvious candidate for time-resolved electron microscopy (or diffraction) measurements like we discuss in the next section. It could also be combined with a laser-driven undulator [50] to make an euv source for lithography and for chip inspection. There has also been some some exploration of DLA for radio-biology applications [51]. The larger hope, however, is that making relativistic electrons widely available will lead to unforeseen applications.

1.2 Ultrafast transmission electron microscopy

Transmission electron microscopy (TEM) has a long history of helping scientists make discoveries by imaging the structure of materials. It is well-suited to this task because the De Broglie wavelength of electrons ($\lambda = h/p$) is small enough to probe atomic seperations. But it is not particularly well suited to observing material *dynamics*, because dense bunches of electrons experience strong repulsive forces. Modifications to the microscope can reduce the temporal resolution from milliseconds down to femtoseconds, and thus ultrafast electron microscopy (UEM) hopes to enable the study of phenomena such as solidification, crystallization, magnetic switching, phase transitions, and dislocation motion [52, 53, 54, 55].

Before discussing how UEMs work, it is worth digging into what we mean by time-resolved dynamics. Dynamics are literally the evolution of a system under the action of forces. In order to study dynamics in a systematic way, we need to be able to control the forces which drive the dynamics. And, if we are interested in fast processes, then we have to be able to control the forces on a fast time-scale. This form of reasoning leads us to the 'pump-probe' methodology in which pump refers to the force used to manipulate the system and probe refers to the technique used to measure the state of the system. For as much time as we will spend talking about ultrafast probe techniques, the ultrafast pump is equally important. The key technology here is the ultrafast laser, since it can be used to rapidly heat a sample [52], or to drive electromagnetic modes in a sample [56].

The three main styles of pump-probe UEM experiments are caricatured in Fig. 1.2 to emphasize the different ways we can image beam dynamics. On the left is a stroboscopic

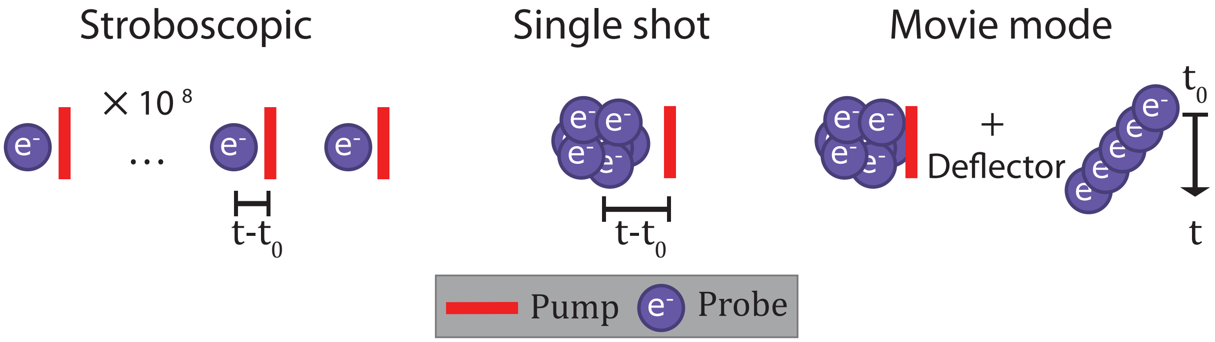

Figure 1.2: Cartoon of laser-pump, electron-probe schemes. From left to the right the techniques are designed to observe increasingly stochastic phenomena.

scheme, in which an image is formed by making bunch trains which have billions of electron-pump pairs with the same delay $t - t_0$. This technique only works with highly repeatable dynamics, but its low charge density helps to preserve the spatial resolution of a conventional microscope. If instead the sample dynamics are stochastic, then a single-exposure imaging technique will be needed. In the middle frame we show a dense electron bunch capable of making a short exposure, although the increased electron density will now start to effect the resolution. If we need to go beyond "still-shots" and follow a single process from start-to-end, then we need a "movie-mode" technique, like we show in the right-most part of the cartoon. Here, a deflecting structure (like a streak camera) maps the $t - t_0$ coordinate of a long electron beam onto a spatial axis.

Research into building such a UEM started around the 1980's when Bostanjoglo and colleagues began modifying commercial microscopes to have pulsed electron sources[3] [57]. They found two ways to do this. The first uses a 'beam-blanking' element: that is, a deflector (either pulsed electro-'static' or an RF cavity), which can periodically dump the beam into a wall. This creates a modulated electron current which is synchronized to an electronic trigger for use in pump-probe measurements [58]. The second way to make a pulsed electron beam is to to use an ultrafast laser to either trigger or gate photo-electrons [59]. Photo-emission is prompt enough [60] that the electrons will be well synchronized with the laser system. Most

---

[3] In principle you could pulse the electron detector, rather than the beam itself, but this would cause unnecessary radiation damage to the sample.

modern UEM's use a photo-emission source rather a beam-blanking cavity, but both can work.

The term "UEM" (which we have been using rather broadly) actually comes from Ahmed Zewail's group at Caltech. In 2005 they realized that by using an 80 MHz laser they could take a stroboscopic image with only a single-electron per pump-probe pair, and thus they could avoid the Coulombic interactions that had plagued earlier devices [54]. In single electron mode their UEM (aside from the electron source and detector) is exactly the same as a commercial microscope, and so it can make high-resolution images via diffraction contrast, phase contrast, and electron energy loss spectroscopy (EELS) [61]. Another advantage of single-electron UEM is that the electron source size can be very small. For example, Roper's group has recently developed a laser-triggered field-emission tip with a normalized transverse emittance of only a few picometers (rms) [62] (which is only about 10 times worse than the limit set by the Pauli-exclusion principle). This is much brighter than conventional photo-emission, (even when ignoring Couloumb repulsion), and it allows them to make high-contrast images even when the sample is very weakly scattering.

At around the same time as the Caltech UEM was being developed, a group at Lawrence Livermore National Laboratory (LLNL) began developing a modern version of single-shot imaging that they called the dynamic transmission electron microscope (DTEM) [63]. Like the other instruments we have discussed so far, the DTEM is also a modified commercial TEM. The main modifications are the installation of a tantalum cathode, access ports for the lasers, and also an additional condenser lens [64, 52]. The condenser lens is needed because, unlike single-electron UEM, the single-shot DTEM has Coulomb repulsion, and so the microscope optics need to be changed to account for the increased beam divergence. The end result is a unique machine which is capable of ns-nm imaging in a single-shot. The beam quality, however, is constrained by Coulomb repulsion, and so the imaging is usually done with amplitude contrast (rather than phase contrast) [65].

This is the point in the story where particle accelerators enter the picture. In order to extend single-shot imaging below the nanosecond time-scale of the DTEM, Li and Musumeci proposed that MeV electrons from an RF photogun be used for microscopy [66]. The MeV electrons don't experience the Coulomb force so strongly because they have an increased "relativistic mass", and so the resolution of their proposed machine is 10 ps-10 nm. But, unlike the previous machines we have discussed, the MeV microscope can't be based on a modified TEM. Instead, all of the components of the beam-line have to be designed from scratch.

The switch to relativistic electrons is not unprecedented. A similar evolution happened in ultra-fast electron diffraction (UED) a few years ago [67]. Like UEM, high quality UED patterns have to be performed in single-electron mode, since obtaining single-shot patterns requires a high current beam (although not as high as for UEM). For this reason, it is advantageous to used relativistic electrons to suppress space charge [68]. The resulting diffraction patterns are not as highly resolved as those from conventional UED, but it is still possible to monitor Bragg peaks as a function of the time delay $t - t_0$ [69]. That is often enough information to measure material dynamics, and so a number of labs have built MeV UED machines over the last few years [70, 71, 72].

Building on the parallels between UED and UEM, a number of these groups have also begun developing the technologies needed for an MeV UEM. All of them start with an RF photogun in order to generate the requisite beam brightness, but they differ in how they replicate the beam optics of a TEM. The basic problem is that high energy electrons are very stiff and so conventional optics have a long focal length. One way around this is to use 2 T solenoid magnets [73, 74], but such lenses are very large and expensive to build. An alternative approach, suggested in the original plan by Li and Musumeci, is to use permanent magnet quadrupoles (PMQs). The PMQs have the advantage of being quite compact, but they are much more difficult to tune than conventional electromagnetic lenses [1].

Another challenge for the single-shot UEM is managing the effects of the Coulomb interaction. This has several layers, but at the simplest level we can think of the Coulomb interaction as the potential of a smooth charge distribution (in the rest frame of the beam). In this case, it acts as a defocusing force which causes the image plane to move as function of the beam current. If the force is purely linear, then it won't distort or blur the image and it can be easily corrected by adjusting the focal length of the objective lens. However, if the force is nonlinear, then it will add aberrations. These aberrations are more difficult to correct than conventional aberrations, because they depend on the current distribution within the microscope (which in turn depends on all of the optics). Therefore it is important to shape the electron beam distribution in a way which limits the non-linear contribution of space-charge. Li and Musumeci [66] have shown that once the force is linearized the spatial resolution of the microscope can be as good as 10 nm, and it will be limited only by the stochastic, or collisional, part of the Coulomb interaction.

So far no MeV UEM has made a study of any material dynamics. In [1] we developed a PMQ objective lens for a UEM, but the magnification was too small to see anything below 1μm. In this thesis, we will show images using 2 PMQ stages and 900x magnification, but again with limited resolution. For now the best results come from the two-solenoid setup in [73]. They get 4 ps-100 nm resolution (at 2600x magnification), and have shown that they can generate amplitude contrast from a variety of materials. But to make a useful MeV UEM will require additional innovations. A microscope needs be able to easily switch between diffraction and imaging mode, adjust the magnification, and get different kinds of contrast. In this thesis we will review some of the requirements for such a device, and we will show off the parts of the technology which we have already developed.

The ultimate goal of the program here at Pegasus is to add capabilities that no other other imaging modality can achieve. Among electron-microscopy techniques, it is the only one which can form real-space images of stochastic phenomena. This will allow us to get direct information about the local (non-periodic) structure of materials, such as dislocations or topological phase transitions. Such information might also be accessible through coherent diffraction imaging with x-rays, but these should be seen as complimentary techniques, because electrons scatter differently than x-rays. That means they can have a higher elastic to in-elastic scattering ratio, and also that they can give you information about different parts of the material structure. The result is that we should be able to see the evolution of structures which were before hidden by the blur of time.

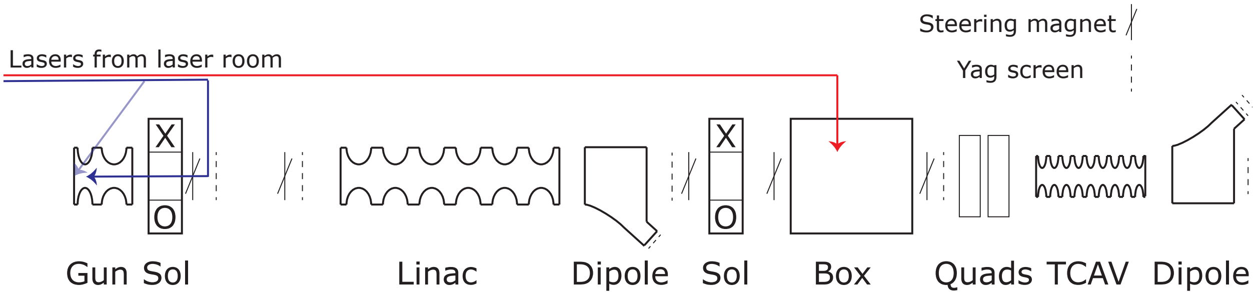

Figure 1.3: Layout of the Pegasus beamline. The electron beam travels from left to right in this figure.

1.3 Introduction to Pegasus

Pegasus[4] is a small electron beamline which is currently designed to study ultrafast accelerator physics in a university setting. It sits in a 10 m x 3 m radiation bunker and is accompanied by a small laser room with a 3-mJ 45 fs Ti:Sapphire laser. The accelerator itself consists of an RF photogun and a short 11-cell linac. That makes Pegasus quite small by accelerator standards, but what it lacks in beam energy it makes up for with high brightness and advanced diagnostics [81, 82].

An outline of the Pegasus accelerator at the time of this thesis is shown in Fig. 1.3. Electrons originate on the left, inside an RF photogun [83], where they are accelerated to 99% of the speed of light. After the gun, the main components are: a solenoid for collimating the beam (which is divergent coming out of the gun); a linac [84] which can either boost the energy (to a maximum of 7 MeV) or put a negative chirp on the beam to initiate velocity bunching; a dipole spectrometer for measuring the beam energy; a second solenoid for focusing into the experimental box; a quadrupole doublet for refocusing the beam downstream;

---

[4] Pegasus no longer stands for anything, but the name does have historical precedent. The history starts in 1991 when the bunker was built for the Saturnus accelerator. By 1998, Saturnus had completed its measurements of SASE FEL gain [75, 76, 77] and thereafter the project was reinvented as Pegasus, which stood for “Photoelectron generated amplified spontaneous radiation source”, and was intended to house a 17 MeV plane wave transformer photoinjector for testing beam-radiation interactions [78, 79, 80]. After a fallow period, the beamline was recommissioned by Pietro Musumeci in 2007 [81]. The name Pegasus stuck, but the focus of the lab shifted away from radiation generation to the creation, measurement, and applications of low emittance electron beams.

a transverse deflective cavity (TCAV) for streaking the longitudinal beam distribution [85]; and a final spectrometer which is often used in combination with the TCAV for measuring longitudinal phase space. Interspersed are steering coils to center the beam and screens which can be inserted to check the beam profile.

The high-gradient photoinjector is the heart and soul of the Pegasus laboratory. It is what allows us to produce relativistic beams with both high current density and low emittance. It can achieve these contrary goals because it rapidly accelerates electrons to relativistic energies, where space-charge forces are less effective at tearing the beam apart. This feature was originally exploited to make a high-current source for free electron lasers [86], but the genius of Pegasus (and similar beamlines [72, 68, 74, 87, 88]) is to realize that when operated at low bunch charge ($< 1\,\mathrm{pC}$) it can have improved emittance ($\epsilon_n < 25\,\mathrm{nm}$) and thus become a useful source for probing ultrafast scattering.

The way it works, is that the RF gun is a highly resonant copper cavity which can build-up accelerating gradients of $100\,\mathrm{MV/m}$ before breakdown (or more if the gun is cryogenically cooled [89]). The field inside the gun looks like $E_z = E_0 \cos(kz) \sin(\omega t + \phi)$ with $\omega = ck$ such that there is a traveling wave component moving along the gun at nearly the speed of light. However, when we photoemit electrons they have to start from rest and so they slip relative to the accelerating wave. If they aren't accelerated fast enough, then they risk being slammed back into cathode. Whether or not this will happen is controlled by $\alpha = qE_0/(km_ec^2)$, the energy gain per period. If the beam becomes relativistic within a half period, then it will (nearly) stop slipping and we can say that the electrons have been trapped by the accelerating wave. Then we can use a classic perturbation analysis [90] to find the launch phase, ϕ_0, for which the electrons end up at the peak accelerating phase[5]: $(\pi/2 - \phi_0)\sin(\phi_0) = 1/\alpha$.

For the work in this thesis we use a 1.6 cell gun [83] with a peak field of up to $92\,\mathrm{MV/m}$ ($\alpha \approx -3$), a launch phase of about 24°, and thus an extraction field of $qE_0 \sin(\theta_0) = 38\,\mathrm{MV/m}$. This produces a total energy of about $4.5\,\mathrm{MeV}$. If we want higher beam en-

---

[5] Note that Kim [90] defines α as 1/2 the value we use.

ergy then we have to split some of our 10 MW of RF power to the linac. Typically we do the split so that the peak field in the gun drops to roughly 65 MV/m ($\alpha \approx -2$) and a launch phase of around 10° for an extraction field of only 11 MV/m. This still is still enough to eject the electrons 3.5 MeV, and it gives us enough power in the linac to double the energy to 7 MeV.

The source of "ultrafast" physics at Pegasus is a 45 fs, 3 mJ Ti:Sapphire laser. It is a commercial system (a Coherent "Legend-Elite") which consists of an diode-pumped oscillator that can be synchronized to the RF system, a stretcher, a regenerative amplifier, and a compressor. After the compressor we pick-off a small amount of the laser (e.g. 0.5 mJ) and up-convert it to UV (266 nm) in order to photo-emit electrons from the copper surface of the RF gun. The photo-emisson from copper is prompt [60], and so we can source electrons as short as the 45 fs laser pulse. Sometimes this actually too short, and so we can use dispersive optics to stretch the UV up to 10 ps.

The remaining laser energy is available for pump-probe experiments or for driving laser-electron interactions. It can be used directly (e.g. for DLA), or it can be down-converted to infrared light in order to couple with solid-state materials [91], or even rectified to a single-cycle THz pulse such that one wavelength is the duration as the electron beam [11]. In any case, it comes from the same laser used for photo-emission thus it is automatically synchronized with the electron beam.

1.3.1 Modes of operation

Applications for the Pegasus beam inevitably involve making a tradeoff between total charge, beam current, and emittance. This happens because if we try to increase the charge density too much, then space-charge will eventually tear the beam apart. To get the most out of the electron beam we have to shape the UV laser so that photo-emitted beam starts out with the optimal combination of parameters. There are two extremes to can consider: the "pancake" regime, in which the photo-emitted beam is much wider than it is flat; and the "cigar" regime, in which the beam is much longer than it is wide.

One way to approach this analysis is to calculate the maximum charge density which can be extracted from the cathode before the self-field of the beam is so strongly repellent that it suppresses emission (i.e. when the electric field behind the beam point back into the cathode). In the steady-state this is the called the Child-Langmuir problem, and $J \propto V^{3/2}/d^2$ for a diode with voltage V and gap d. A similar scaling applies to the cigar beam, except the length of the diode is replaced by the length of the electron beam, and the voltage is replaced by the cathode field $V/d = E_c = E_0 \sin(\theta_0)$. For the pancake beam, we instead calculate the electric field for an infinite sheet of charge, which (with its image charge) makes an electric field σ/ϵ_0. Formalizing these ideas leads to the scaling laws for the charge/current [92, 93]:

$$\begin{aligned} \frac{Q}{\pi\sigma_r^2} &\lessapprox \epsilon_0 E_c, && A >> 1 \\ I &\lessapprox I_0 \frac{\sqrt{2}}{9}\left(\frac{q_e E_c \sigma_r}{m_e c^2}\right)^{3/2}, && A << 1 \end{aligned} \tag{1.1}$$

where $I_0 \approx 17\,\mathrm{kA}$ is the Alfven current, σ_r is the radial size of the beam, and A is the aspect ratio of the *electron beam* near the cathode. This is different than the aspect ratio of the laser, since the electron velocity is not initially c. This leads to a longitudinal compaction from $c\sigma_t$ to $\frac{1}{2}\frac{q_e E_c}{m_e}\sigma_t^2$, so that in terms of the laser dimensions we find $A = \sigma_r m/\sigma_t^2 E_c$.

To fully understand the importance of these scaling laws, we have to remember that the source emittance is proportional to the source size σ_r. In the cigar regime this leads to a favorable scaling for transverse beam brightness, since $I/\epsilon^2 \propto 1/\sqrt{\sigma_r}$ can be made very large. By contrast, the pancake regime has constant transverse beam brightness, but it is able to do so with arbitrarily short pulses. In both cases we benefit from having a large electric field at the cathode.

Throughout this thesis we switch between the two modes of operation. For DLA we need a high current bunch, and so we operate near the pancake regime. For UEM we need as much transverse brightness ("coherence") as possible, and so we operate near the cigar regime. To manage the switch we use the UV pulse length σ_t to control the aspect ratio of the bunch. Then, to take advantage of the cigar regime scaling, we also need to make σ_r as small possible. For this we can enter the gun through an oblique incidence port which allows us to focus the UV with a shorter focal length (see Fig. 1.3). The result is that we have made Pegasus into a very flexible machine which can adapt its phase space density to fit a variety of applications in ultrafast science.

1.4 Summary

Both DLA and UEM are broad projects which enjoy attention from many researchers outside DLA. In the introduction we have attempted to provide this context, but for for the rest of the thesis we will focus on the work done here at Pegasus. For that reason, we present here a pseudo-chronological summary of our experiments and list the major results we have achieved.

Our research into DLA comes from two experimental programs. The first [29] aimed to measure GV/m gradients by using a high intensity laser pulse (without pulse front tilt). This was meant to be a simple extension of the work done at the now-closed NLCTA facility [22, 28], but we soon found that the DLA interaction saturated at high intensities. When we realized that the saturation was due to the optical Kerr effect we were able to compensate for it by adding anomalous dispersion to the laser pulse. This realization changed the way we think about DLA, because it showed that we could control the beam dynamics by tuning the laser (rather than building the beam dynamics into the structure). The second DLA experiment [94, 30] built on the first by modifying the drive laser to have a tilted pulse front, and thus to allowed us to extend the DLA interaction from 20 μm to nearly 1 mm. In addition to setting a new record for DLA energy gain, this geometry gives us better access to the phase of the drive laser. This led us to propose a new experiment [38] which uses a spatial light modulator to control the beam dynamics in a 2 cm DLA.

The research into an MeV ultrafast microscope is based on a proposal by Li and Musumeci [66] that got funded, in partnership with Radiabeam Technologies, by a department of energy STTR (small business technology transfer) grant. The first experiment we did for the UEM was to test a PMQ triplet as a 30 x objective lens [1]. The magnification of this experiment was fairly limited, but it was enough to demonstrate that PMQs are an affordable short-focal-length lens for MeV microscopy. Later we tried the UEM again, with a much more complicated experimental design that had two triplets, a sample stage, and an aperture stop. Unfortunately we did not include a design to image the back focal plane, which led to difficulty generating contrast and ultimately a loss of resolution. Nonetheless, we were able to use the experimental campaign to study nonlinear-dynamics caused by space-charge in the microscope column, and we were able to identify weak points in the microscope design and start preparing an updated proposal for future experiments.

For this thesis we divide each research subject into two chapters, one theoretical (and computational), and one experimental. Part of the purpose of the theoretical chapters is to prepare the reader to understand the experiments, but the theoretical sections also go beyond this to analyze and discuss what future experiments might look like. The experimental chapters are arranged such that they first describe our techniques and then describe our measurements. We try to use each measurement to highlight a specific physical principle and then discuss its implication for machine design.

CHAPTER 2

DLA theory and modeling

A dielectric laser accelerator is a tool for transferring energy from a laser to an electron beam. It does this by slowing down the phase velocity of light to match the velocity of the electron beam. We can think of this a two step process: first the laser interacts with the dielectric to produce the synchronous accelerating mode, and second the beam accelerates under the force of the prescribed field[1].

The first step, where the laser and dielectric combine to produce an accelerating mode, can be rather complex. Some DLA structures work by producing a photonic band gap, some use a “slow-down” layer, and others, like the ones we use, can be thought of as diffraction gratings [4]. But, no matter what mechanism the structure uses, it will produce a free-space accelerating mode which looks like $E_z \propto E_0 \cos(k_z z - \omega t)$. If the mode's phase velocity matches the electron velocity ($\omega = \beta c k$), then it will have an accelerating gradient proportional to the applied field (E_0).

In many ways this process is familiar to accelerator physicists, because a DLA looks a lot like a miniaturized version of a conventional RF linac. Instead of klystrons we have solid-state lasers, instead of metal cavities we have dielectrics, and instead of ps-mm scale dimensions we have fs-μm dimensions. But otherwise the shape of the accelerating mode is the same and so we can expect to find analogies for all of the conventional beam dynamics [95]. This can be comforting: it means that we should be able to find an arrangement which can stably accelerate an electron bunch without destroying the beam brightness. However, it also belies the difficulty of the process, because controlling beam dynamics in a DLA works nothing

---

[1] We typically ignore the reverse process in which the beam interacts with the dielectric to produce radiation. This process would become important if we had a high current beam or if we had a very weak laser.

like controlling them in a conventional accelerator.

One of the major differences between our DLA structures and conventional RF structures is that our DLA structures are non-resonant. Resonances can help build up the field strength of a weak power source, but they also lead to long pulse durations and lower damage thresholds. It is not entirely understood why long pulses cause more damage, but it is certainly related to the fact that for a fixed electric field longer pulses have more energy density (fluence). In fused silica, where bulk damage has been most thoroughly studied, we know that pulses shorter than 10 ps beat out thermal transport and suppress damage by melting. It is thought that for these short pulses damage is caused by an avalanche triggered by free electrons from multi-photon photo-emission (see § B.4 and reference [96]), for which shorter pulses are again advantageous since a fixed electric field implies that there are less photoelectrons available to trigger an avalanche[2].

In order for an ultrafast laser to remain ultrafast inside the DLA, the dielectric structure has to treat all frequencies the same (i.e. have a broadband "impedance"). That is, if we think of the input laser as a sum over plane waves with slightly different frequencies then the field in the DLA is also a sum over the same set of plane waves with the same amplitudes. This prevents the laser from stretching, but it also means that the field in a DLA is sensitive to the phase-relationship between the plane-waves in the laser. If the phase of the laser of is evolving, then the electron beam will see a changing phase. As we develop a model of the DLA interaction we will see that this has profound consequences for the beam dynamics.

In this chapter, we will first present a simple explanation of the fields inside a DLA. Then we will consider what happens to "stiff" electrons in these fields and we will find out how much acceleration we can expect from a DLA. This will lead us to consider the importance of dephasing and analyze the Kerr phase shift. After presenting simulations of these basic phenomena, we will turn to understanding beam dynamics and designing a stable accelerator. Finally, we will conclude with a brief consideration of energy and momentum conservation in the DLA.

---

[2] Interestingly we often cite the damage limit as a single threshold fluence even though it has been measured to be time dependent [97].

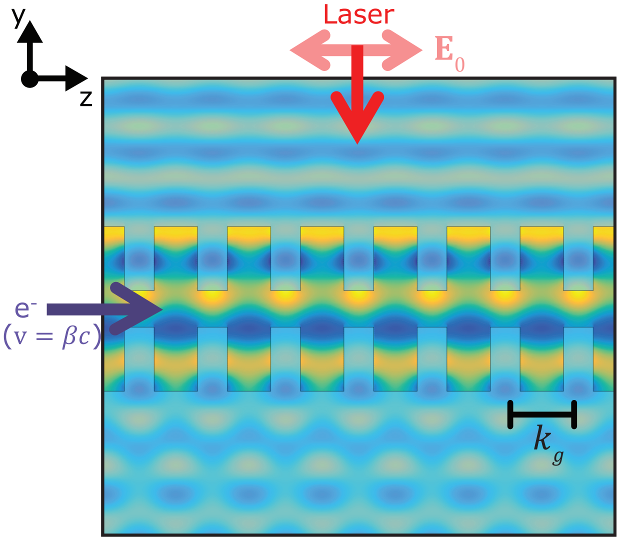

Figure 2.1: Schematic of grating-style DLA showing how a plane wave laser is diffracted into an evanescent accelerating mode.

2.1 Eigenmodes of a grating-style DLA

To date all experimentally tested DLAs[3] have been based on a simple design similar to the one proposed by Plettner [99]. The idea is to use a periodic dielectric (a "grating") to convert a plane wave into an evanescent accelerating mode. This makes DLA synonymous with "inverse Smith-Purcell radiation", an idea which long predates Plettner (but originally used metallic structures) [13, 15]. To distinguish between different types of periodic dielectric structures we often talk about "gratings" [22, 23] versus "pillars" [100, 101]. The two structures work essentially the same way, but the pillars are easier to make since the are etched into a single monolithic assembly, while the gratings are usually made by bonding two separate pieces. The gratings are preferable for experiments with relativistic electrons because they have a larger aspect ratio, and because their accelerating channel is enclosed by dielectric walls. The walls help improve the signal-to-noise ratio by filtering out unaccelerated electrons.

In this thesis we exclusively use dual grating DLAs like the one illustrated in Fig. 2.1 (though all results apply equally well to pillar structures). In this cartoon the laser propa-

---

[3] We should make an exception for [98] which accelerates electrons near the edge of a planar interface. But this was merely a curiosity and was not intended to be a serious accelerator.

gates in the y direction before being diffracted by the DLA into the z direction, where it is matched to the electron beam. The bottom grating helps convert the evanescent accelerating mode back into a propagating mode that continues traveling in the y direction. Because the laser is traveling perpendicular to the electron beam the interaction can only happen for a finite length; however if there are enough electrons interacting then the efficiency can still be high [27, 39], and moreover it is very useful to start with a laser polarized in the z direction in order to concentrate energy in the (transverse magnetic) accelerating mode.

We can start an analysis of a grating style DLA, like that in Fig. 2.1, by applying the grating equation familiar from introductory physics: $N\lambda_0 = d_g\left(\sin(\theta_i) \pm \sin(\theta_f)\right)$. This suggests that an incoming plane wave will be split into a set of modes indexed by the diffraction order N. For the cases of interest the diffraction angle θ will be imaginary. As with total internal reflection, an imaginary angle implies an evanescent wave. To see this, we should recall that θ is related to a plane-wave wavevector as $\theta_{i,f} = \arctan(k_z^{(i,f)}/k_y^{(i,f)})$ where $k_0^2 = k_z^2 + k_y^2$ and $k_z^{(i,f)}$ are the z-components of input and output wavevectors. Then we can rewrite the grating equation as a sort of momentum conservation:

$$Nk_g = k_z^{(i)} + k_z^{(f)} \tag{2.1}$$

where $k_g = 2\pi/d_g$ is the so-called grating momentum. In order to accelerate electrons we want to match the phase velocity of the mode to the electron velocity, so we get the phase-matching relation:

$$\omega = \beta c k_z^{(f)} = \beta c(Nk_g + k_0 \sin(\theta_i)) \tag{2.2}$$

We can immediately see that the phase-matched modes are evanescent, because in free space $\boldsymbol{k} \cdot \boldsymbol{k} = \omega^2/c^2$. Taking k_z from the grating equation and setting k_x to zero for a planar structure[4] leaves us with $k_y = \sqrt{k_0^2 - k_z^2}$. Since $k_0^2 - k_z^2$ is less than zero we can see that $\exp\left(ik_z z + ik_y y\right)$ will be evanescent in y with characteristic decay length $\Gamma = ik_y$.

Now you might protest that the grating equation derived in elementary physics classes should not apply here, since it is based on an argument about the interference of plane

---

[4] Setting $k_x = 0$ also implies that the all of the input fields are invariant x variation. This is only approximately true in experiments.

waves which propagate away from the grating. But, in fact, it does apply equally well to the evanescent modes. In Appendix A we make this argument rigorous by exploiting the symmetry properties of the grating and of Maxwell's equations. We show that the grating equation can be seen as a consequence of Bloch's theorem for periodic structures. We also discuss the importance of other symmetries, most notably reflection symmetry. The reflection symmetry (e.g. about the x-axis) allows us to split the fields into transverse magnetic (TM) and transverse electric fields (TE) which only have (E_z, E_y, B_x) or (B_z, B_y, E_x) components respectively.

Since our laser starts polarized with (E_z, B_x) we will only produce a TM mode. Then, by applying Maxwell's equations to the waveform we calculated from the grating equation, we can write down the fields of the m^{th} order mode in the DLA as the real part of:

$$\begin{aligned}
\boldsymbol{E_m} \cdot \hat{z} &= iE_0 \left(c_{(1,m)} e^{-\Gamma_m y} + c_{(2,m)} e^{\Gamma_m y} \right) e^{i(k_m z - \omega t)} \\
\boldsymbol{E_m} \cdot \hat{y} &= E_0 \gamma_m \left(c_{(1,m)} e^{-\Gamma_m y} - c_{(2,m)} e^{\Gamma_m y} \right) e^{i(k_m z - \omega t)} \\
\boldsymbol{B_m} \cdot \hat{x} &= -\frac{E_0}{c} \beta_m \gamma_m \left(c_{(1,m)} e^{-\Gamma_m y} - c_{(2,m)} e^{\Gamma_m y} \right) e^{i(k_m z - \omega t)}
\end{aligned} \tag{2.3}$$

where we have abbreviated $k_z^{(f)} = k_m$. Since k_m and ω determine the mode velocity we can define $\beta_m = \omega / c k_m = k_0 / k_m$. This lets us express the evanescent decay length as $\Gamma_m = i k_y = \sqrt{(k_m^2 - \omega^2/c^2)} = k_0/(\beta_m \gamma_m)$. DLAs need to have a narrow gap in order to produce a high gradient (typically no larger than λ), so $\Gamma_m y$ is often a small number (although for subrelativsitic particles it can be greater than 1).

Finally, we should draw attention to the (complex) constants c_1 and c_2, since their origin isn't obvious from our earlier discussion of the grating. They correspond to modes decaying from the top and bottom gratings and come from the sign ambiguity in $\exp(\pm i k_y)$. For perfectly symmetric or anti-symmetric illumination we get $c_1 = \pm c_2$ and the exponentials combine to form a "cosh" or a "sinh". In order to accelerate electrons we would like the E_z to be a cosh-mode so that the acceleration is approximately uniform across y, while the transverse forces will be shaped like a sinh and thus be approximately linear.

We will often use the on-axis mode amplitude $E_m = E_0 |c_{(1,m)} + c_{(2,m)}|$ to represent the gradient in the DLA. The factor $|c_{(1,m)} + c_{(2,m)}|$ tells how to convert the incident field to gradient, and so we will often call it the "diffraction efficiency", but we caution that it is not the an energy efficiency, and so it can actually be greater than 1. It is related to number of other factor which have been used by the DLA community. For example, Peralta [102] defines an "acceleration factor" as $f_A = E_m/E_{\text{max}}$ so that he can optimize the gradient before damage (assuming damage occurs when λE_{max} is above the band gap of fused silica). To connect this to experiments, he also defines a "field enhancement factor" as $\eta_{\text{fe}} = E_{\text{max}}/E_0$, so that he would say the gradient is $E_m = E_0 \eta_{\text{fe}} f_A$ and thus, in his terms, our diffraction efficiency would be $|c_{(1,m)} + c_{(2,m)}| = \eta_{\text{fe}} f_A$ [5].

2.2 Electron acceleration in a DLA

Now that we have shown a DLA can support an accelerating mode (Eq. 2.3), we are ready explore simple acceleration processes. This analysis will focus on the acceleration of a single electron in a short DLA. It is preparation for helping us interpret the result of experiments we performed at Pegasus.

Since we wish to talk about the acceleration of an electron we now need to define, separate from the laser, an electron velocity β and energy $\gamma m_e c^2$. We assume the electron is suitably 'beam-like' that $\beta_z \approx \beta$ and thus that energy and velocity are directly related: $\gamma^2 = 1/(1 - \beta_z^2)$. Since confusion could occur between the electron velocity and the phase velocity of the laser, we will explicitly subscript the mode velocity β_m (as well as the associated quantity γ_m).

It will also help to define a quantity ψ to represent the phase of accelerating mode. For a plane wave this is simply:

$$\psi_{\text{plane wave}} = \psi_0 + k_m z - \omega t \tag{2.4}$$

where k_m comes from Eq. 2.3 and the diffraction equation. In this expression z and t should be seen as implicit functions of the path length along the accelerator (s), such that for a

---

[5] Note that if E is to be defined *inside* the fused silica then there should be a Fresnel factor floating around [103]. Peralta calculates the Fresnel factor incorrectly and so interpreting his formulas requires caution.

linear reference trajectory $z = s$ and $ct = \int_0^z (1/\beta) ds$. Additionally, if the laser is not a pure plane wave, then ψ can have additional s dependence, as in $\psi = \psi_{\text{plane wave}} + \psi_{\mathcal{E}}(s)$. Throughout this chapter we will develop the idea that controlling ψ is a very powerful tool for accelerator design.

Kinematic approximation To calculate the motion of relativistic electrons in a short DLA it is typically safe to assume the electrons are extremely rigid and thus that they don't deviate from the trajectory they were on before the DLA. In 1D that amounts to assuming that $z = z_0 + \beta_0 ct$, or equivalently that $\psi = \psi_0 + z(k_m - \omega/\beta_0 c)$, even while the electron energy changes along the DLA according to $\partial_z \gamma = q_e E_z / m_e c^2$. This is sometimes called the kinematic approximation or the use of 'ballistic' trajectories (although in our case ballistic means neglecting gravity).

In this case, we can write the energy gain along an unperturbed trajectory C as:[6]:

$$\Delta E = q_e E_m \int_C ds \cos\left(\psi(s)\right) = q_e E_m \int_{-L/2}^{L/2} dz \cos\left(\psi_0 + k_0 z \left(\frac{1}{\beta_m} - \frac{1}{\beta}\right)\right)$$

where we have used the definition $k_m = k_0/\beta_m$ and we have used an on-axis mode amplitude of $E_m = E_0 |c_{(1,m)} + c_{(2,m)}|$ and a DLA of length L. If the phase velocity of the mode is matched to the beam then $\psi(s) = \psi_0$ and the integral is independent of z. If the velocities are mismatched then the phase grows linearly with z and the energy gain will oscillate. The velocity mismatch is typically small so we will expand β around the small quantity $\eta = (\gamma - \gamma_m)/\gamma_m$[7]. The energy gain can then be simplified to:

$$\Delta E = q_e E_m L \sin(\psi_0) \text{sinc}\left(\frac{k_0 L}{2} \frac{\eta_0}{(\gamma_m^2 \beta_m^3)}\right) \tag{2.5}$$

---

[6] Compared to Eq. 2.3 we have specialized to a cos instead of a complex exponential. There is no loss of generality since we have kept ψ_0 to represent the initial phase of the electron relative to the wave.

[7] You might think it makes more sense to use a variable we can call $\eta_p = (\gamma\beta - \gamma_m\beta_m)/\gamma_m\beta_m$ which is the normalized change in momentum. This isn't quite as nice as using the change in energy directly since $\partial_z \gamma m_e c^2 = \partial_t p$. That is, when we use z as the independent variable (in place of t) energy is a canonical momentum (and p is the new Hamiltonian). However, there is no problem writing the momentum version. Depending on whether you use η or η_p as a small parameter gives slightly different approximations: $\left(\frac{1}{\beta_m} - \frac{1}{\beta_0}\right) \approx \eta / \left(\beta_m^3 \gamma_m^2\right) \approx \eta_p / \left(\beta_m \gamma_m^2\right)$.

The fact that the energy gain is linear in the field amplitude is one of the defining characteristics of DLA. So is the sinusoidal dependence on the phase of the electrons. By contrast, the sinc function is often unseen in experiments since the argument is deliberately kept close to zero in order to maximize the energy gain.

Tuning the phase velocity In order to prevent the sinc function in the previous section from limiting the acceleration, we need to produce a DLA whose phase velocity is well-matched to the electron beam. The grating equation (Eq. 2.1) tells us how to do this: we can change the incidence angle of the laser, θ_i, in order to tune the phase velocity. This is apparent in the phase matching condition (Eq. 2.2), which is important enough that we reprint it here:

$$\omega = \beta c(mk_g + k_0 \sin(\theta_i)) \tag{2.6}$$

One way to build intuition for this relationship is to consider a plane wave propagating at an angle. It has a phase variation $k_0 \cos(\theta_i)y + k_0 \sin(\theta_i)z + \omega t$ as it approaches the structure. At the boundary it has to match the accelerating mode and so, up to the k_g provided by the structure, the $k_0 \sin(\theta_i)$ has to be inherited by the DLA (as the Bloch phase in appendix A).

We usually choose $m = 1$ and $k_g = k_0$ such that $\theta_i = 1/\beta$ (in the small angle approximation). Alternatively we could write $\beta_m = 1/\theta_i$ for the mode velocity. Then we can write the energy gain from the previous section in terms of an error in the incidence angle, $\Delta\theta_i$ instead of the energy deviation η. We find: $\Delta E \propto \mathrm{sinc}(k_0 L \Delta\theta_i/2)$.

Effective length Once we have tuned the phase velocity, the energy gain will no longer be limited by dephasing, but by the region which can be illuminated by the drive laser pulse. Since the drive laser propagates transversely to the electron beam it is typical for the laser-electron interaction to last as only long as the (ultrafast) pulse length, as in Fig. 2.2(a). To extend this interaction we can tilt the laser pulse by about 45° so that the interaction can be extended as far as the transverse length of the laser allows, as in Fig. 2.2(b). In either case it is the drive laser envelope which limits the interaction, not the DLA.

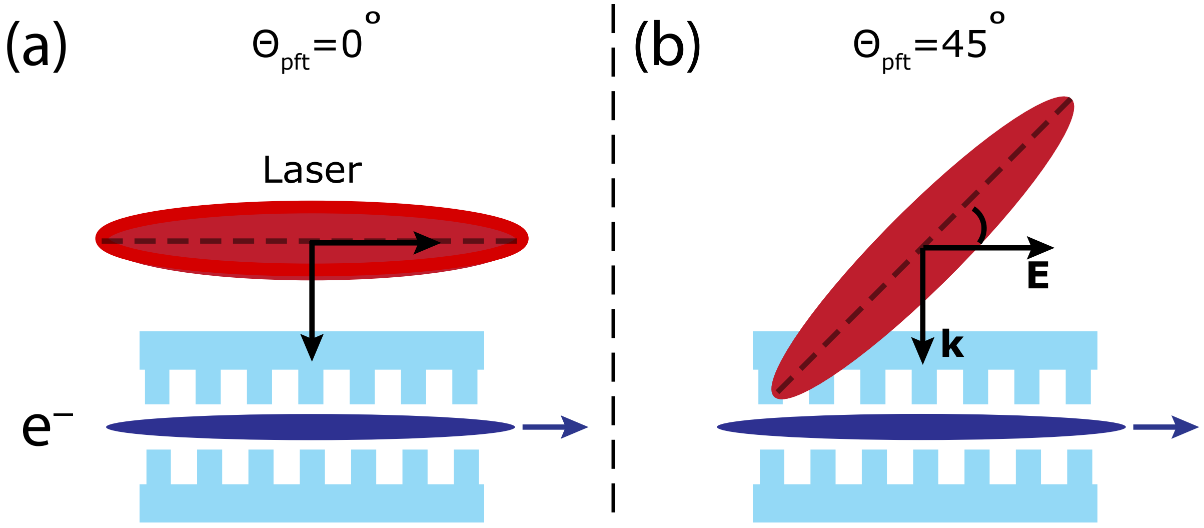

Figure 2.2: Cartoon showing a DLA illuminated by a flat pulse and a tilted pulse. In (a) the temporal duration of the flat pulse limits the interaction length, while in (b) the spatial extent of the tilted pulse provides a much longer interaction length.

By laser envelope we mean to describe the profile, $\mathcal{E}$, of a laser field which looks like $E_z = E_0\mathcal{E}(\boldsymbol{r},t)\exp(i(k_0 z - \omega_0 t))$. $\mathcal{E}$ is a slowly varying complex-valued function in the sense of the paraxial wave equation. Siegman's "Lasers" [104] and Diels and Rudolph's "Ultrafast laser phenomena" [105] give a good account of what exactly this means (though Siegman labels the envelope u instead of $\mathcal{E}$), so we will only develop it briefly.

The primary function of $\mathcal{E}$ is to bound the laser to a finite spot-size (and pulse duration), but since $\mathcal{E}$ is complex-valued it also effects the phase. To emphasize this we can write $\mathcal{E} = |\mathcal{E}|e^{i\psi_{\mathcal{E}}}$ such that the phase of accelerating wave is now $\psi = k_0 z - \omega_0 t + \psi_{\mathcal{E}}$. There is some ambiguity in choosing the carrier frequency (k_0 and ω_0), but any choice we make can be compensated by an appropriate $\psi_{\mathcal{E}}(\boldsymbol{r},t)$. We typically choose the carrier frequency to eliminate any linear components from $\psi_{\mathcal{E}}(\boldsymbol{r},t)$ at the $\boldsymbol{r},t$ where the intensity is maximized.

The accelerating mode for a laser with envelope $\mathcal{E}$ is simply the TM mode for the carrier wave (ω_0, k_0) multiplied by the envelope. In this case the energy gain becomes (for $\eta = 0$):

$$\Delta E = q_e E_m \sin(\phi_0)\left|\int_C \mathcal{E}(\mathbf{r},t)ds\right|, \tag{2.7}$$

and we can interpret the integral as an effective length $L_{\text{eff}} \equiv \left|\int_C \mathcal{E}(\mathbf{r},t)dz\right|$ for the DLA.

Fig. 2.2 makes it obvious that the effective length of the pulse-front-tilted beam is longer than the flat-pulse. But L_{eff} isn't just the spot-size of the laser, because $\mathcal{E}$ can have a phase component which lowers the energy gain. For example, a Gaussian pulse with a radius of curvature (or its longitudinal analog, chirp) will have a quadratically varying phase. As the phase grows the integral will start to oscillate rapidly and average to zero. As a result, the tails of the of the pulse won't contribute to the energy gain and so aren't counted in the effective length.

Frequency domain We can gain some intuition for the effective length by breaking the laser envelope into plane-waves. Formally we do this by Fourier transforming $\mathcal{E}$ to get its $(\omega, \boldsymbol{k})$ components: $\mathcal{E}(\boldsymbol{k}, \omega) = \mathcal{F}(\mathcal{E}(\boldsymbol{r}, t))$. Now we notice that even though $\mathcal{E}$ is propagating in the y direction, the fact that it has finite z extent necessitates that it have k_z components. This means that we can think of the laser as being assembled from from a set of rays, each of which are incident on the DLA with a a slightly different angle. If we apply the grating equation to each ray then we can construct the composite accelerating mode.

In the previous section, we constructed the composite accelerating mode by assuming it was proportional to $\mathcal{E}(\boldsymbol{r}, t)$. We can now see that this is equivalent to assuming the diffraction efficiency is the same for each ray. That way we can pull E_m out of the Fourier integral and transform $\mathcal{E}$ alone. In the introduction to this chapter we argued that this assumption should hold because the DLA needs a large bandwidth to support ultrafast pulses, but we will leave a proof to the simulation section. For now we will continue calculating the energy gain in the Fourier domain. We can re-write the energy gain integral (Eq. 2.35) as:

$$\Delta E = q_e E_m \sin(\theta_0) \left| \int_C ds \mathcal{F}\left(\mathcal{E}\left(k_z, \omega\right)\right) \right| \tag{2.8}$$

which we can rewrite for a kinematic trajectory $t = z/(\beta c)$ as:

$$\Delta E = q_e E_m \sin(\theta_0) \left| \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} dk_z \mathcal{E}\left(k_z, \omega\right) e^{i\left(k_z z - \omega\left(\frac{z}{\beta c}\right) + \psi_0\right)} \right| \tag{2.9}$$

where we have tacitly assumed the physical DLA is long enough that we can extend the limits of integration to infinity. Switching the order of integration to do z first turns the exponential

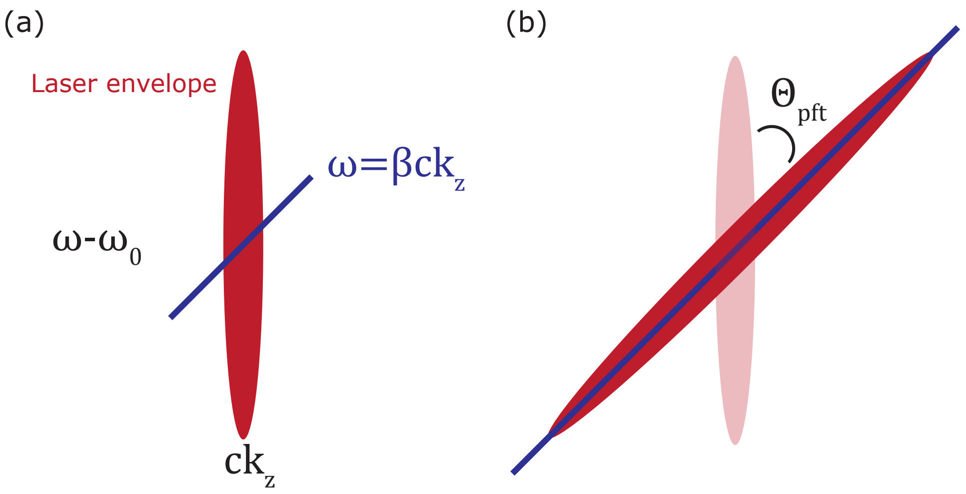

Figure 2.3: Acceleration in the frequency domain: (a) Flat pulse front and (b) Tilted pulse front. The purple line shows the components of the laser envelope which can interact with an electron moving at velocity β. Small changes in the electron velocity will correspond to a vertical displacement of the line (due to the vertical axis being offset by ω_0).

in a delta function: $\delta(k_z - \omega/\beta c)$. Then we can use the delta function to eliminate the k_z integral. That leaves:

$$\Delta E = q_e E_m \sin(\theta_0) \left| \int_{-\infty}^{\infty} d\omega \left(\mathcal{E} \left(k_z = \omega/\beta c, \omega \right) \right) \right| \tag{2.10}$$

Thus in frequency space the acceleration of a particle is given by an integral of the Fourier transformed laser envelope along the phase-matching condition.

The frequency space picture of acceleration is shown in Fig. 2.3 for both flat and tilted pulse fronts. The red ellipses show a contour of the frequency-space envelope $\mathcal{E}(k_z, \omega)$ and the purple lines show the path of integration corresponding to an electron moving at velocity β. The ellipses have a large aspect ratio because for an ultrashort laser $\sigma_t < \sigma_z$ such that there is larger bandwidth in ω than in k_z [8]. To take full advantage of the laser pulse we have to correlate angle (k_z) with frequency (ω) such that the major axis of the envelope lies along the electron trajectory. This angular dispersion becomes a tilted pulse front in the time-domain [106].

---

[8] Recall that the laser is propagating in y so that $\sigma_t \approx c\sigma_y$ not $c\sigma_z$.

With a longer interaction comes increased sensitivity to dephasing. In Fig. 2.3 we can see how this works in the frequency domain by realizing that a small change in the particle velocity β looks like a vertical shift of the purple line. Clearly this has a much more significant impact on the tilted pulse front than on the flat pulse front. This corresponds to the sinc-like dephasing we saw in the time-domain expression of Eq. 2.5. By contrast the impact on a flat pulse front is negligible, in fact the aspect ratio of our drive laser is so extreme that we can approximate the integral from Eq. 2.10 as occurring at fixed ω (while k_z still changes), so that the energy gain is:

$$\Delta E_{\text{no PFT}} \propto q_e E_m \sin(\phi_0) \, |\mathcal{E}(\omega = \omega_0)| \tag{2.11}$$

2.3 Kerr dephasing

The primary source of dephasing in our experiments is a nonlinear phenomena called the optical Kerr effect. It occurs when a high intensity laser stretches the dielectric into anharmonicity and generates a nonlinear polarization density. It can be described phenomenologically by an index of refraction $n = n_0 + n_2 I$. At 800 nm in fused silica this expression can be fit to measurements to find $n_0 = 1.45$ and $n_2 = 2.45 \cdot 10^{-16}\text{cm}^2/\text{W}$ [107]. For our purposes we can treat the Kerr effect as instantaneous and local [108] so that the intensity dependence causes different slices of the lasers to experience a different index of refraction.

An immediate consequence of the Kerr effect is that the peak of a laser develops a phase shift relative to its low-intensity tails. After propagating a distance d through a nonlinear medium, that phase shift is approximately:

$$\Delta\psi_{nl} = n_2 I k_0 d \tag{2.12}$$

In terms of the envelope $\mathcal{E}(\boldsymbol{r}, t) = |\mathcal{E}(\boldsymbol{r}, t)| e^{i\psi_{\mathcal{E}}(\boldsymbol{r}, t)}$ the phase becomes $\psi = k_0 z - \omega_0 t + \psi_{\mathcal{E}}(\boldsymbol{r}, t) + \Delta\psi_{nl} |\mathcal{E}(\boldsymbol{r}, t)|^2$. A phase matched DLA will have $k_0 z - \omega_0 t(z) = 0$, but, unless the envelope phase and the Kerr phase compensate for each other, then there will be dephasing, and the effective length of the accelerator will suffer.

During typical operation at Pegasus the phase shift is dominated by the nonlinear phase change $\Delta\psi_{nl}$. The DLA gratings we use in this thesis are only about 1 µm tall, but they are etched into the bottom of a $d = 500\,\mu\text{m}$ thick fused silica wafer. At an intensity of 3TW/cm^2, well before the damage threshold, that adds up to $\Delta\psi_{nl} = \pi$ and an electron crossing through the laser will alternate between de-accelerating and accelerating phases. This effect only gets worse at higher intensities and so it causes the electron energy gain to saturate as a function of the drive laser power.

At even higher intensities the Kerr effect creates a strongly curved index which refracts that pulse envelope refracts and causes it to change shape within the $d = 500\,\mu\text{m}$ thick fused silica wafer. Near the damage limit (11TW/cm^2) we can use simulations (see § 2.4.1 and appendix B) to show that the laser starts to self-focus then self-steepen, and eventually filament. But even as the pulse collapses, the main effect on DLA acceleration comes from the rapidly changing phase. Thus we are often satisfied with ignoring these changes to the envelope and just using $\Delta\psi_{nl}$ from Eq. 2.12.

Compensation At a given intensity it is possible to balance the Kerr phase $\Delta\psi_{nl}|\mathcal{E}(\boldsymbol{r},t)|^2$ with an appropriate envelope phase $\psi_{\mathcal{E}}(\boldsymbol{r},t)$ in order to make ψ constant along the accelerator. To see how this works let us consider an example relevant to the Pegasus experiments: we will take the case of a Gaussian drive laser in the flat pulse configuration (figures 2.2(a) and 2.3(a)) whose dispersion can be adjusted by means of grating compressor [109]. We will assume that the laser has $c\tau << w_z$ so that the electrons are essentially stationary in z (relative to the envelope) during the time the laser passes by. In this case we only need a 1D envelope and so we write:

$$\mathcal{E}(t) = \exp\left(-\frac{t^2}{\tau_G^2 - 2i\phi_2}\right), \tag{2.13}$$

where $\tau_0 = \tau_G\sqrt{2\ln(2)}$ is the fully compressed pulse width[9] and ϕ_2 is the group velocity dispersion. Note that in this section we use t as shorthand for $t - y/v_g$, where v_g is the group velocity of the laser. Thus we are referencing a position relative to the laser envelope and

---

[9] Convention dictates that τ represents the intensity FWHM. τ_G is instead the 1/e point of the electric field (equivalent to the waist w, used for transverse laser dimensions).

not an absolute time. We can get away with this because for fixed y (e.g. the center of the DLA) there is no appreciable difference.

Even before we add the Kerr phase, the effect of dispersion on the envelope requires some consideration. As we change ϕ_2 the pulse gets longer, which increases the effective length of the accelerator, but it also grows a phase, which decreases the effective length. Moreover, the mode amplitude has to decrease as the $\sqrt{\tau}$ in order conserve energy. The net effect of GVD on the energy gain is not at all obvious until we consider the pulse in the Fourier domain. In this case $\mathcal{E}(\omega) = \exp\left(-2\ln(2)(\omega-\omega_0)^2/\Delta\omega^2\right)\exp(i\phi_2(\omega-\omega_0)^2)$ with a bandwidth $\tau_0\Delta\omega = 4\ln 2$. Now we can see the ϕ_2 is just adding a spectral phase. That simplifies things, because from Eq. 2.11 we know that the energy gain only depends on $|\mathcal{E}(\omega)|$. Thus the energy gain is independent of ϕ_2 in the absence of the Kerr effect.

Unfortunately the Kerr phase is a time-domain phenomena and its effect on $|\mathcal{E}(\omega)|$ is not obvious. To make progress we can split the time-domain envelope into an amplitude and phase, as in $\mathcal{E}(t) = |\mathcal{E}(t)|e^{i\psi_{\mathcal{E}}}$ and then assume $|\mathcal{E}(t)|$ is constant while the phase grows a Kerr contribution. This yields (assuming $k_0 z - \omega_0 t(z) = 0$):

$$\begin{aligned} |\mathcal{E}(t)| &= \exp\left(-\frac{t^2}{\tau_G^2 + 4\phi_2^2/\tau_G^2}\right) \\ \psi &= -\frac{2\phi_2 t^2}{\tau_G^4 + 4\phi_2^2} + \Delta\psi|\mathcal{E}(t)|^2 \end{aligned} \tag{2.14}$$

The first term in ψ is called chirp[10] and it is analogous to a radius of curvature in the spatial-domain[11]. The second term is the Kerr phase, and it is shaped like a Gaussian. To leading order (in t), the Gaussian is also quadratic and so we can choose ϕ_2 to roughly balance the two effects and make ψ appear constant.

The effect of ϕ_2 and $\Delta\psi$ on the pulse is shown in the pulse envelopes of Fig. 2.4. Changing just the dispersion (leftmost column) causes the pulse to broaden and the chirp to grow. Changing just the Kerr phase (top row) causes a large Gaussian phase. The two effects are

---

[10] To see why its called chirp look at the "instantaneous frequency" $\omega_i(t) = \partial_t(\omega_0 t + \psi_{\mathcal{E}}) = \omega_0 - t\frac{16\ln 2^2\phi_2}{\tau_0^4 + 16\ln 2^2\phi_2^2}$. The frequency is linearly varying along the pulse. If it were sound it would make a rising or falling "tone" (depending on ϕ_2) like a bird "chirping".

[11] We have left out the temporal analog to the Guoy phase shift. It is not so important because it adds a phase which depends not on $t - y/v_g$ but on y alone.

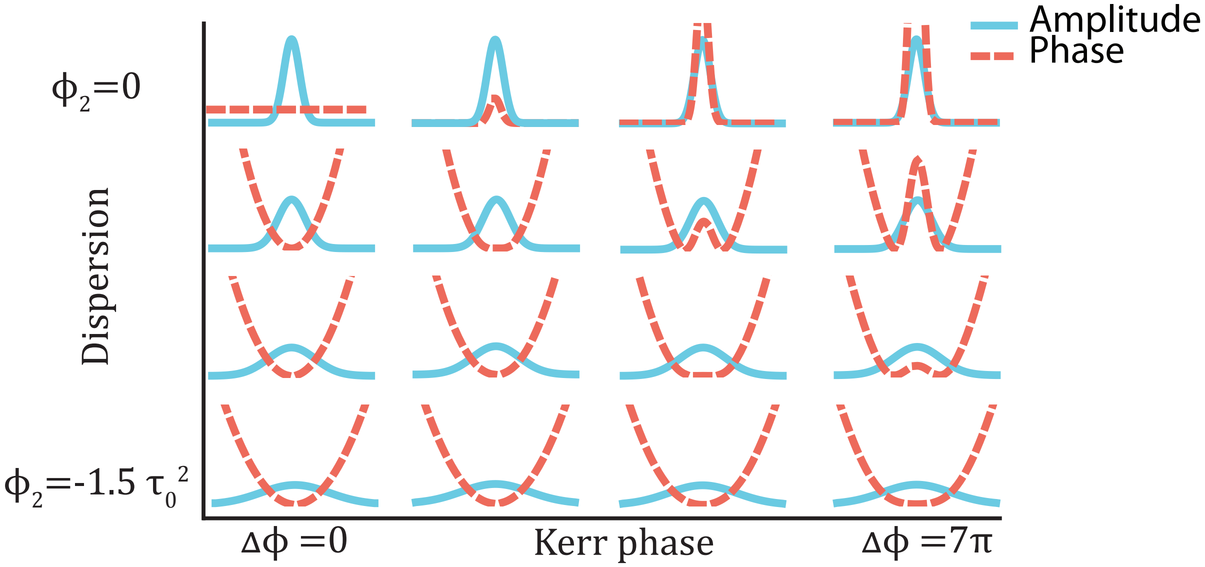

Figure 2.4: Cartoon showing how dispersion and the Kerr phase effect the envelope $\mathcal{E}$. Along the diagonal are a set of envelopes whose dephasing has been (mostly) compensated.

balanced on the main diagonal and so the phase is approximately flattened. This allows for much higher energy gain in the DLA.

In practice this means we can use ϕ_2 as a “knob” to tune the $\psi(s)$. We were already using θ_i to control the 1st order component (§2.2), and now we can also control the 2nd order component. For short DLAs we use this ability to flatten the phase and make ψ constant, but for longer DLAs we can actually choose a gently curved phase to matching the increasing velocity of the electron beam.

2.4 DLA simulations

We use simulations to connect our empirical knowledge of the laser envelope to the electron spectra produced by the DLA. This approach is illustrated in Fig. 2.5 as a chain of simulation tools. Each simulation is specialized to the relevant physics in order to aid interpretation and make the computational load feasible.

Starting from the left of Fig. 2.5, we initialize our simulations with measurements of the laser spatio-temporal profile. Assuming that there is no spatio-temporal coupling (or, in the

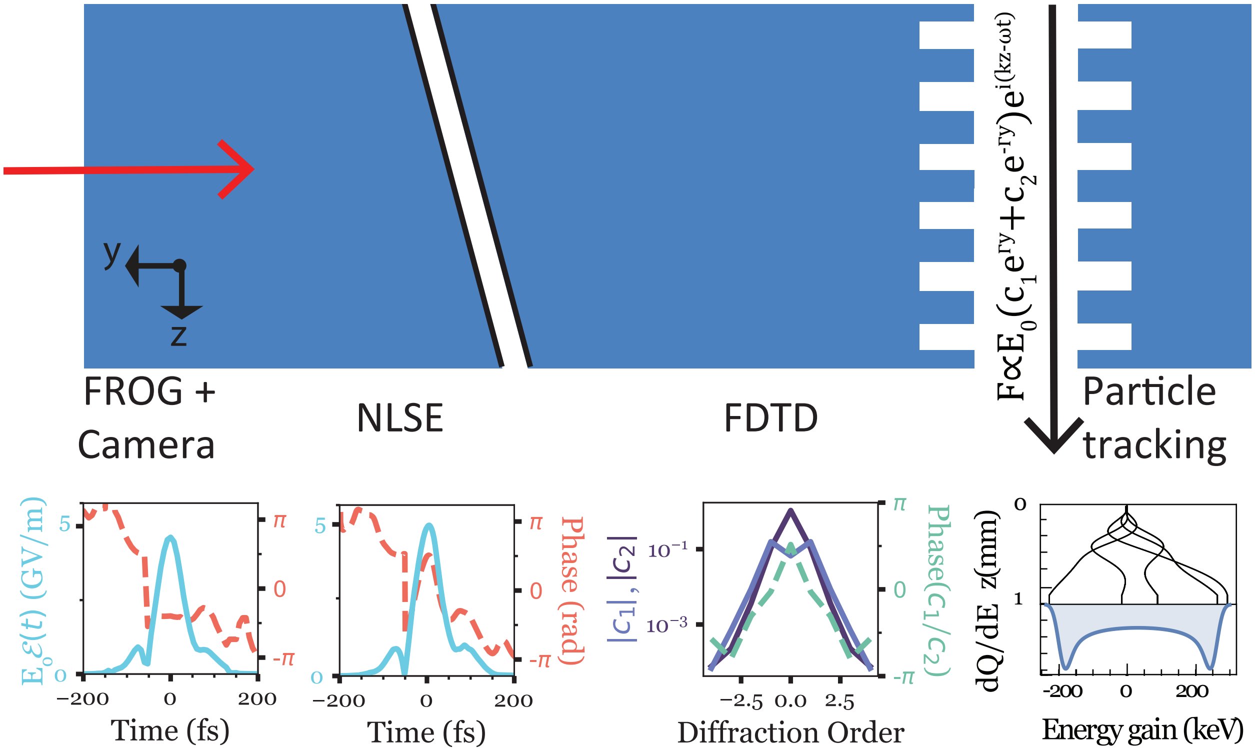

Figure 2.5: A chain of simulations connects the measured laser profile to the electron spectra. From left to right the laser progresses from measured data through nonlinear propagation in bulk fused silica to full electro-magnetic modeling at the DLA grating and finally particle tracking. The sub-figures are representative of the output from each stage. From left to right: the measured amplitude and phase from FROG; the amplitude and phase after numerical propagation through a 1D NLSE; the efficiency of a symmetric dual grating DLA; and finally particle tracking in the analytic field given the known pulse envelope and the grating efficiency.

case of a tilted pulse front, that the coupling is known exactly), we can easily propagate the fields in free space to the top surface of our DLA to find the pulse envelope's intensity and phase. From there we simulate propagation through $\approx 500\,\mu\text{m}$ of bulk, nonlinear, dielectric, which causes self-phase modulation and a small increase in the pulse intensity due to delf-focusing. Then we simulate the sub-wavelength diffraction gratings using a full electromagnetic code in order to extract the diffraction efficiency. Finally we track electrons through the accelerating fields to determine the electron spectra.

Most importantly, we will show that these simulations agree with more basic calculations. For example, the electromagnetic finite difference time domain (FDTD) simulation [110] is largely unnecessary because the DLA grating is broadband: that is, it diffracts all the plane-waves in our laser with the same efficiency. Consequently, the effect of the DLA can be taken into account by specifying two complex numbers (one for each grating).

2.4.1 NLSE

In the slowly varying envelope envelope approximation a laser is modeled by a rapidly oscillating phase contained inside the complex envelope function. As the laser propagates the envelope evolves according to differential equation which has the same form as a Schrödinger equation. In this system the Kerr effect acts a nonlinear source term, so the we call the propagator a generalized nonlinear Schrödinger equation (NLSE) [111]. In this section we consider a pulse propagating in $\hat{y}$ with envelope $\mathcal{E}(\mathbf{r}, t)$. When writing down the NLSE we absorb E_0 into $\mathcal{E}$ to simplify the notation. With that in mind, the pulse propagation is governed by:

$$\frac{\partial \mathcal{E}}{\partial y} = \left[\hat{D}_f + \hat{D}_s + \hat{N}\right] \mathcal{E} \tag{2.15}$$

where $\hat{D}_f, \hat{D}_s, \hat{N}$ are the diffraction, dispersion, and nonlinear operators, respectively:

$$\begin{aligned}
\hat{D}_f =& \frac{i\lambda}{4\pi n_0}\left(1 - \frac{i\lambda}{2\pi c}\partial_t\right)\left(\partial_x^2 + \partial_z^2\right) && (2.16)\\
\hat{D}_s =& -\frac{i}{2}k^{(2)}\partial_t^2 - \frac{1}{2}\xi && (2.17)\\
\hat{N} =& \frac{i2\pi n_2}{\lambda}\left|\mathcal{E}^2\right| - \frac{1}{\mathcal{E}}\left(\frac{n_2}{c}\right)\partial_t\left(\mathcal{E}\left|\mathcal{E}^2\right|\right) \\
& - \frac{i2\pi n_2}{\lambda}T_r\partial_t\left|\mathcal{E}^2\right| - \frac{\beta_6}{2}\left|\mathcal{E}^{10}\right| && (2.18)
\end{aligned}$$

A detailed discussion of this equation is provided in appendix B. In brief, we have set up the equation to be solved by a finite difference method. To efficiently propagate the beam we use the well-known split-step algorithm to alternate between Fourier domains [111, 112] for applying the the linear and non-linear kicks respectively. Our method is built on the 'Generalised Adaptive Fast-Fourier Evolver' (GAFFE) [113] which allows us to grow (or shrink) the mesh used to describe the pulse envelope in order to preserve memory.

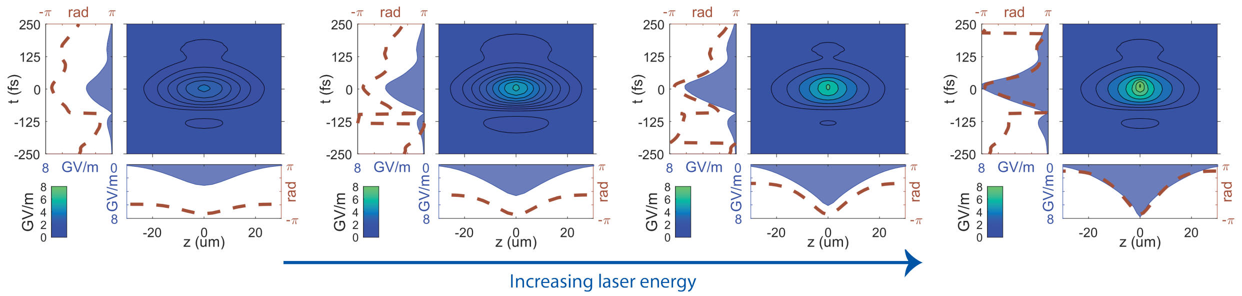

Figure 2.6: Pulse envelope as a function of laser intensity for a flat pulse. Each plot shows a (z-t(y)) contour of the laser envelope after propagation through 500 µm *of fused silica. The Kerr effect causes the intensity dependent phase as well as self-focusing and self-steepening of the laser pulse.*

We illustrate the effects of nonlinear propagation through 500 µm of fused silica in figures 2.6 and 2.7 by a series of 2D beam profiles each having increasing beam energy. Without nonlinear effects the pulse envelope would grow, but not change shape. In actuality, however, the high energy beam profiles have a strong phase change and noticeable self-focusing.

For Fig. 2.6 in particular we are modeling the laser used in the flat-pulse experiments and so we show a slice of the beam along the z and $t(y)$ directions. We initialize the simulations according to the measured beam profiles (from FROG and from a camera[12]), yielding a profile very much like the leftmost (lowest energy) image in the figure. At higher energies we can see, in the line-outs, that the phase begins to follow the intensity profile (i.e. the square of the electric field). At even higher energies we can see self-focusing (in the sharpening of the bottom lineout) and even some self-steepening (similar to self-focusing, but in the time-domain). Although it is not obvious from this figure, the peak intensity is actually increaseing faster than the pulse energy. The increase, however, is usually small for the relevant parameters[13] and the main effect is the self-phase modulation.

When extending the analysis to model a tilted pulse front we need to move to a rotated

---

[12] The camera only gives the beam intensity, not a phase. We choose a radius of curvature which makes the beam focus at the DLA plane, which is consistent with experimental procedure and wavefront measurements.

[13] For example, in the rightmost plot the peak electric field would be lowered by 30% in the absence of non-linear effects. Also, note that the effect of self-focusing was ignored in damage measurements for DLA [96].

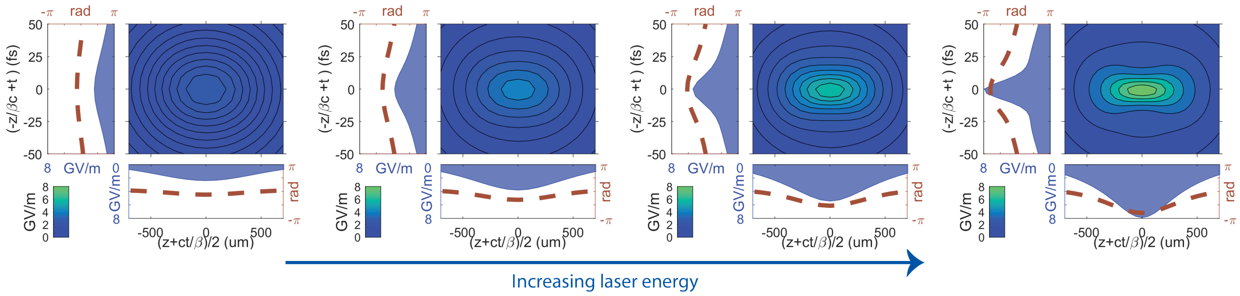

Figure 2.7: Pulse envelope as a function of laser intensity for a tilted pulse front. Each plot shows an equivalent to a (z-t) contour, but rotated to be along the eigen-axes of the (coupled) pulse envelope. Compared to Fig. 2.6 the Kerr effect effects the narrow t direction much more strongly since it now experiences diffraction in addition to dispersion.

coordinate system to represent the beam along its eigen-axes. In the representation of Fig. 2.7 we show a slice of the beam in $(t - z/\beta c) \times (z + ct/\beta)$ which is roughly equivalent to the $(t) \times (z)$ view shown above; except that in the flat case the electrons propagate in time, while the PFT case they propagate along the spatial direction. We also simplify the simulation by initializing a perfectly Gaussian beam (fit to the FROG and spatial camera measurements). Most of the pulse evolution is analgous to the flat-pulse, but there is one notable difference: the tilted pulse shows very strong self-focusing in the time-like dimension (the left-inset of the right-most plot). This sharp focus happens because the laser is very narrow in the "time-like" dimension and so the "Kerr-lens" has a short focal length. Normally the kerr-lens is defocusing in time (unless the material has anomalous dispersion), but with our rotated envelope the time-like dimension experiences diffraction. The diffraction operator has the opposite sign of the dispersion operator (and much greater magnitude), so that the net propagation causes the pulse to focus. This is a novel effect, which to our knowledge was first reported by us in [94]. Unfortunately, however, we made no direct measurements of this profile.

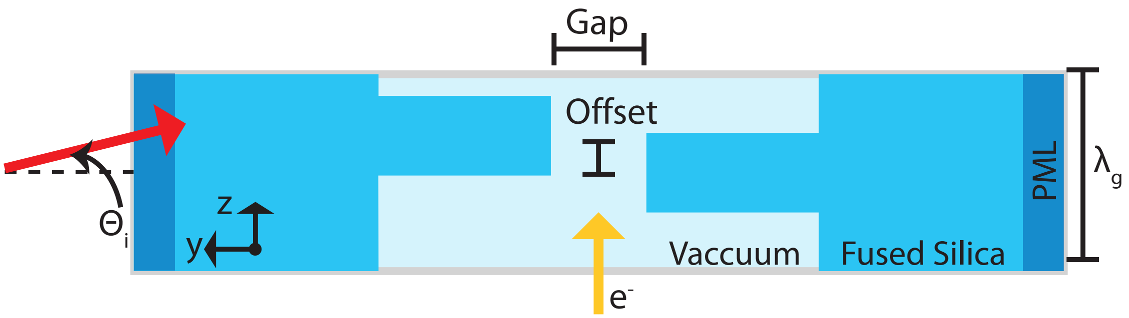

Figure 2.8: Unit cell for FDTD simulations showing the direction of laser propagation

2.4.2 Peralta Structure

All of the structures used in this experiment are from the same batch of structures fabricated by Edgar Peralta [102] based on a design by Thomas Plettner [27]. We use two types structures, those with a gap of 800 nm and those with a gap of 400 nm. Both structures have a tooth length of 725 nm and a tooth width of 325 nm. Although SEM images show rounded corners near the substrate (at the bottom of the etch), we will model the structures as having perfectly rectangular teeth.

Our FDTD simulations use the unit-cell pictured in Fig. 2.8. A commercial code is used to calculate the propagation of an ultrafast laser pulse and then we Fourier transform the fields in the gap to understand the frequency response of the DLA. We want to study the grating efficiency as a function of ω, $\mathbf{k_B}$, gap size and tooth offset, so we set up the simulation with Bloch periodic boundary conditions. This forces $\mathbf{E}$ to be periodic up to a phase $e^{ik_\mathrm{B}\mathbf{z}}$ determined by the incident angle $k_\mathrm{B} = (\omega/c)\sin(\theta_i)$[14]. A PML (phase matching layer) is tuned to eliminate reflections from the boundaries. In principle the results depend on what direction the electrons and laser come from, but given one orientation we can find the others using the symmetry relations from section § A.4.

To quantify the grating efficiency we decompose the field into modes indexed by their phase velocity, β_m. We can do this by Fourier transforming the FDTD fields from (t, z) to

---

[14] All of the frequency components need to have the same Bloch periodicity, so k_B is calculated from the angle for the central wavelength. This means the laser in the simulation has a slight pulse front tilt.

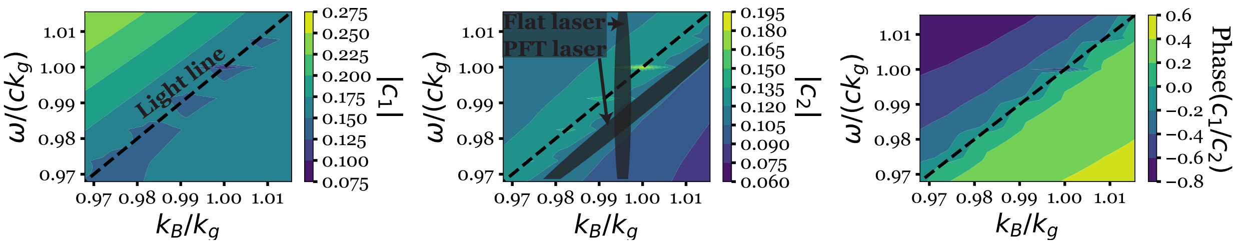

Figure 2.9: Diffraction coefficients as a function of frequency and Bloch wave-number. Notice that the results are relatively constant along the electron trajectories (dashed lines), while there is some variation perpendicular to the light line.

(ω, k_z). Since the structure is exactly Bloch periodic in space we know that the fields can only have discrete spatial frequencies $k_z = Nk_g + k_\mathrm{B} \cdot \mathbf{z}$ for integer diffraction orders N. To get the response at other frequencies we have to run batches of simulations with different k_B. Then we fit the amplitude from each simulation to a sum of exponentials decaying from each side of the gap: $E_z/E_0 = c_1(\omega, k_z)e^{\Gamma y} + c_2(\omega, k_z)e^{-\Gamma y}$ with $\Gamma = \sqrt{k_z^2 - \left(\frac{\omega}{c}\right)^2}$ (as in the eigenmode Eq. 2.3). This allows us to characterize the entire DLA by just two (complex) numbers $c_1(\omega, k_z)$ and $c_2(\omega, k_z)$.

In Fig. 2.9 we show the frequency response of a 400 nm gap DLA structure with no tooth offset. For reference we shade in an outline of the FWHM intensity contour of two typical laser pulses (with and without PFT), but note that by changing the incidence angle we could slide these contours along the k axis. Since we don't care about the overall phase of the fields, we show only three numbers: $|c_1|$,$|c_2|$, and the angle between c_1 and c_2. All 3 parameters are roughly constant along the light-line, while they show a small, but significant, change as a function of phase velocity. Thus it is fair to treat all components having the same phase-velocity as having a single efficiency. Note that above the light-line the modes begin propagating, and so $|c_1|$ refers to the downward propagating component and $|c_2|$ to the reflection.

Next, in Fig. 2.10, we show how the DLA response depends on the tooth offset. This time we only show the diffraction coefficients for a single phase velocity (resonant with 6 MeV electrons), but we do so for 3 structures: An 800 nm gap Peralta structure, a 400 nm

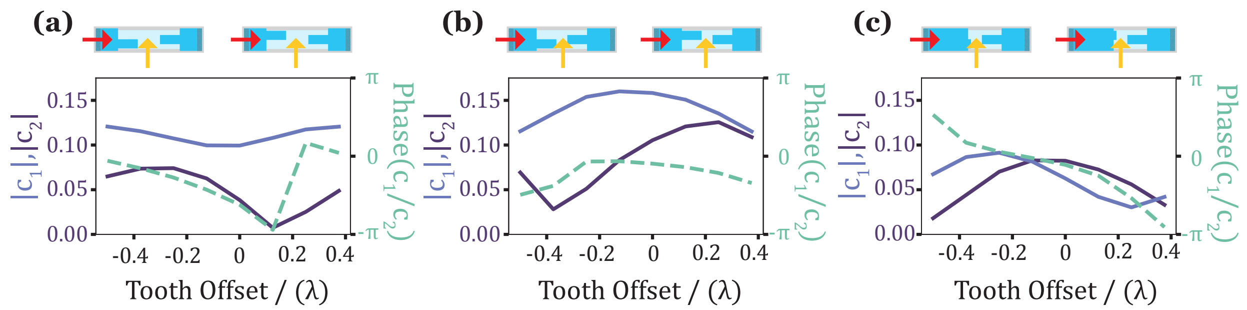

Figure 2.10: Diffraction coefficients as a function of tooth offset for three structures. (a) an 800 nm gap Peralta structure, (b) a 400 nm gap Peralta structure, and (c) an asymmetric structure with symmetric fields. Note that the cartoon structures (which are drawn to scale) define the geometry and the sign of the tooth displacement.

gap Peralta structure, and a purposefully asymmetric structure with a 400 nm gap (all with single-sided illumination). Since c_1 comes from the same side as the laser it is generally the stronger than c_2, but we can symmeterize the fields by shrinking the tooth on the c_1 side to make the diffraction less efficient. This is how the asymmetric structure is able to have symmetric fields with single-sided illumination.

Electron-spectra measurements are not very sensitive and so we can't expect to measure the variations we have just simulated. But, when designing a longer structure to trap and accelerate electrons, we do care about the field structure in the accelerator. These simulations illustrate two important features of the field structure: (1) the DLA is very broadband, especially for input plane waves having the same phase velocity; and (2) the fields can be symmetrized by adjusting the tooth dimensions and offsets.

2.4.3 Measuring tooth offset

It is clear that the tooth offset is an important parameter for determining the fields in the accelerator grating. Unfortunately, the Peralta structures are made by bonding two wafers together and it is difficult to do so with sub-micron accuracy. Moreover, we find it likely that Peralta bound the structures with a twist, as shown in Fig. 2.11(e), such that one wafer has many offsets. Consequently we must measure each structure to determine the tooth offset.

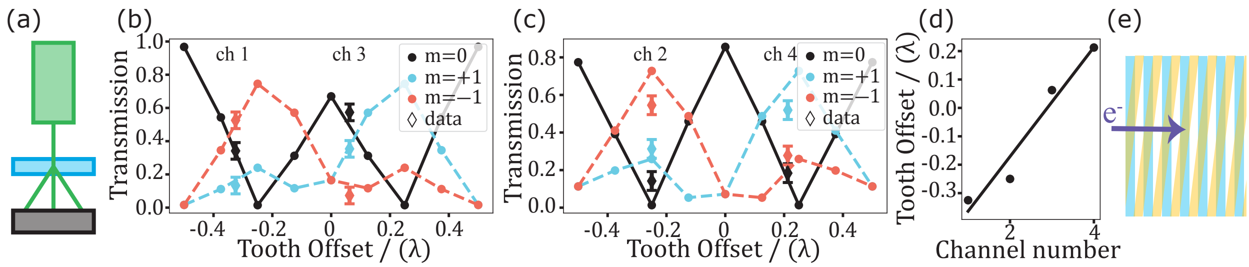

Figure 2.11: Measurement of tooth offset: (a) Cartoon showing diffraction measurement; (b) Simulated diffraction from a peralta structure with an 800 nm gap and measurements of 2 800 nm acceleration channels; (c) Measurements on a 400 nm gap; (d) The best-fit offset for each channel; (e) a cartoon of a rotated grating.

To measure tooth offset we measure the far-field diffraction pattern formed by using a green laser diode (at 532 nm), as illustrated in Fig. 2.11(a). We can compare the intensity of light in the m=-1,0 and 1 orders to simulations of the diffraction intensity in order to identify the tooth offset. In Fig. 2.11(b) and Fig. 2.11(c) we show simulations of the 800 nm and 400 nm gapped structures, respectively, and then we overlay measurements (squares) of four acceleration channels at the most likely tooth offset[15]. We can then plot the tooth-offset as a function of distance from the alignment channel (the channels are each placed 2 mm apart from each other and are 0.5 mm wide–measurements take place on the larger, 1 mm part of each structure). The measured linear variation is consistent with a tilted wafer.

2.5 Beam dynamics

So far we have only been considering the acceleration of stiff electrons which propagate on unperturbed “kinematic” trajectories, but in a long accelerator this isn’t true, and we need to consider the full beam dynamics. These are significantly more complicated, so where possible we try to simplify them. In practice this means that we use the kinematic approach for all

---

[15] All of our published data, with one exception, comes from channels 1 and 2. The exception is the gradient-saturation plot which contains data from both channel 2 and a corollary to channel 2 on a different chip which I have not measured. That chip, which is the one I broke with a bouncing screwdriver, has too much material on the bottom of it (ablated aluminum from its mount) to probe with the green laser.

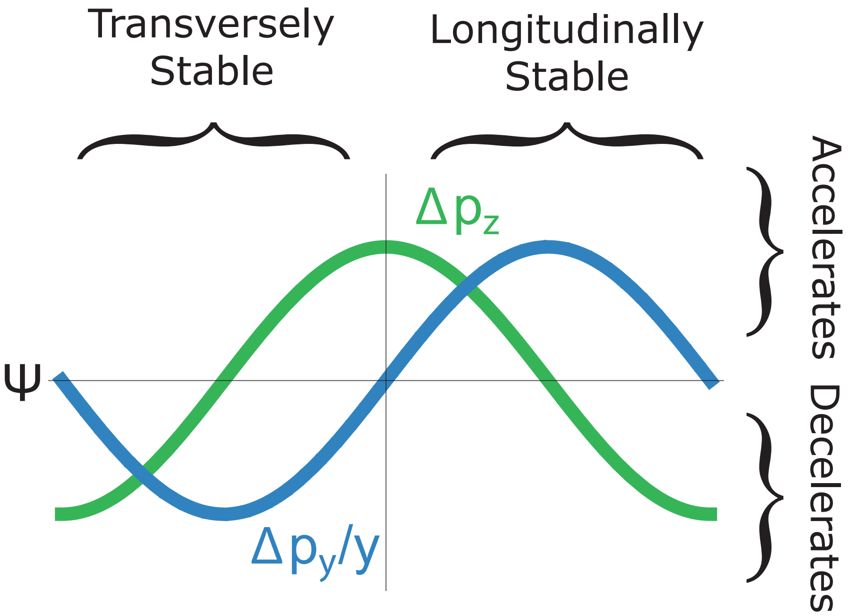

Figure 2.12: Force diagram for a free-space TM mode. The longitudinal force always lags the transverse force by $\pi/2$.

but the longest tilted pulse front interactions. Nonetheless, we will present the dynamics in some detail in order to show that it is possible to design a stable MeV scale DLA accelerator.

A physically intuitive way to start our study of the TM mode is to boost the fields from Eq. 2.3 to a frame moving at the phase velocity $\beta_m < 1$. In the boosted frame (indicated by tildes) the TM mode become entirely electric. It is then obvious that it can be derived from the scalar potential:

$$\Phi = E_0 \frac{\beta_m \gamma_m}{k_0} e^{i\frac{k_0 \tilde{z}}{\beta_m \gamma_m}} \left(c_{(1,m)} e^{-\frac{\tilde{y}}{\beta_m \gamma_m}} + c_{(2,m)} e^{\frac{\tilde{y}}{\beta_m \gamma_m}} \right) \tag{2.19}$$

This potential has no minima, only saddle points, and so the motion is inherently unstable. Naranjo et.al. [33] identify this as an analogy of Earnshaw's theorem from electrostatics and it holds for any *free space* mode since $\nabla^2 \Phi = -\nabla \cdot \boldsymbol{E} = 0$. It is only useful, however, for a resonant interaction, since we need to be able to "freeze" both the fields and the electrons in order to have a completely static arrangement.

Back in the laboratory frame, we can see the consequences of this theorem by plotting the forces on an electron as a function of the phase $\psi = k_m z - \omega t$. In Fig. 2.12 we have labeled regions of this plot to indicate what kind of motion each quadrant corresponds to. Transverse stability requires a restoring force (blue line below zero), while longitudinal stability requires particles which are ahead to be accelerated more slowly (the green line needs a negative slope). Clearly there is no region which is both transversely and longitudinally stable, but there are regions which are accelerating and either transversely or longitudinally stable. These are the useful regions for DLA operation and later we will show that by switching between them we can achieve net stability. For now we re-iterate that this type of force map is a general feature of TM-modes and thus effects all DLA devices.

2.5.1 Longitudinal beam dynamics

Ignoring transverse effects (valid for $\Gamma_m y << 1$), we can write the longitudinal dynamics as [114]:

$$H_{\|}(\gamma, \psi; k_m z) = \gamma(1 - \beta\beta_m) - \alpha_m \cos(\psi) \tag{2.20}$$

where γ is the canonical momentum, ψ is the canonical position, and the independent variable is $k_m z$. The main scale parameter is the small quantity $\alpha_m \equiv q_e E_0 |c_{(1,m)} + c_{(2,m)}| / (k_m m_e c^2)$ which tells us the energy gain per optical period.

The scale parameter α is actually the same quantity we used in the introduction to describe trapping in the RF gun (§ 1.3.1). We can see how this works by looking at the phase portraits in Fig. 2.13. The small α dynamics have a very small seperatrix and so only resonant electrons are trapped by the wave. But as $|\alpha|$ approaches 1 the seperatrix reaches down to $\gamma = 1$ and so it can trap particles from rest. In the RF gun we have $\alpha \approx 2$ so that we can accelerate the meV photoemitted electrons up to MeV energies.

For the DLA, $\alpha \approx 10^{-3}$ and so the bucket height is quite small and unsuitable for acceleration. In order to gain appreciable energy, it is necessary to taper the phase of the accelerator to match the changing velocity of the beam. This amounts to changing variables to a reference frame moving with $\gamma_m(z)$. And since the beam has to fit inside the small seperatrix, we can safely expand the Hamiltonian about small $\eta = (\gamma - \gamma_m)/\gamma$ (and use η as a canonical momentum instead of γ). The result is a set of dynamics which are analogous

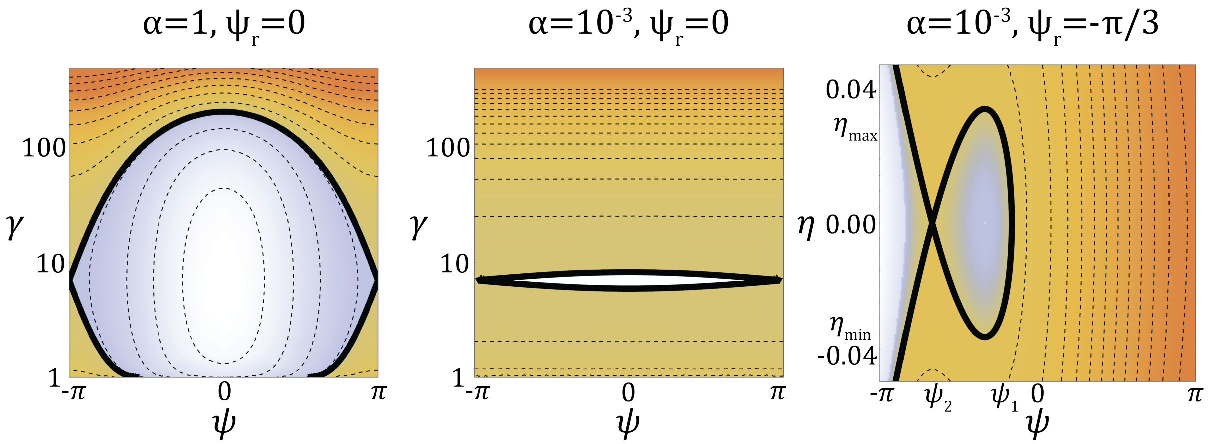

Figure 2.13: Longitudinal phase portraits for a sinusoidal potential with $\gamma_m = 7$. Particles move along the contours of constant H. The parameter α controls the seperatrix height and the phase ψ_r the location of the stable fixed point.

to a forced pendulum with α playing the role of gravity and γ_m that of mass:

$$H_{\|}(\eta, \psi; k_m z) = \frac{\eta^2}{2\gamma_m^2 \beta_m^3} - \frac{\alpha_m}{\gamma_m} \left(\cos(\psi) + \psi \sin(\psi_r)\right) \tag{2.21}$$

In this Hamiltonian the stable fixed point ($\psi = \psi_r$, $\eta = 0$) occurs at the resonant phase and the frequency of small oscillations is: $k_s = k_0 \sqrt{\alpha_m \cos(\psi_r)/\gamma_m^3 \beta_m^3}$ (for historical reasons k_s is called the synchrotron frequency). When $\psi_r = 0$ this is the same as the previous Hamiltonian (in the small η limit), but otherwise this Hamiltonian describes seperatrix which is increasing in central energy like $\partial_z \gamma_m = -\alpha_m k_m \sin(\psi_r)$. For this reason we call this the "moving-bucket" picture of electron acceleration.

The separatrix boundaries can be calculated as (c.f. [115]):

$$\begin{gathered} \cos(\psi_1) + \psi_1 \sin(\psi_r) = \cos(\psi_2)_+ \psi_2 \sin(\psi_r) \\ \psi_2 = \pi \, \mathrm{sign}(\psi_r) - \psi_r \end{gathered} \tag{2.22}$$

$$\eta_{\max} = \sqrt{2\alpha_m \gamma_m \left(\cos(\psi_r) - \cos(\psi_2) + (\psi_r - \psi_2) \sin(\psi_r)\right)} \tag{2.23}$$

Where ψ_1,ψ_2 are the locations where the bucket touches $\eta = 0$, and $\eta_{\max}$ is the height of the bucket at $\psi = \psi_r$. The seperatrix is also called the "bucket" because particles which start inside tend to stay inside, even if we (adiabatically) change the scale parameters [116].

2.5.2 Transverse beam dynamics

The transverse beam dynamics are considerably simpler and yet more vexing. We again ignore transverse-to-longitudinal coupling and expand the forces in the limit $\Gamma_m y << 1$. In this case the transverse force becomes:

$$F_y = qE_0\gamma_m \left((c_1 - c_2) - (c_1 + c_2)\Gamma_m y\right)(1 - \beta\beta_m)e^{i\psi} \tag{2.24}$$

which has both a dipole kick and a quadrupole kick. Normally we try to design $c_1 = c_2$ so that the dipole kick vanishes, but if not we at least try to put a potential minima inside the gap. If c_1 and c_2 have the same phase then we can eliminate the dipole by changing variables to $y = \tilde{y} + |c_1 + c_2|/|c_1 - c_2|$. This shifted position must lie inside the DLA gap or else we have no hope for transverse stability. From here on we will assume $c_1 = c_2$ to simplify the analysis (i.e. we are studying the "cosh" mode, for which the transverse forces are a "sinh" that we are approximating as linear).

In this limit we are left with a time-dependent harmonic oscillator. It is natural to consider the oscillator strength κ defined by $y'' = -\kappa_m y$ such that:

$$\kappa_m = -\frac{\alpha_m \cos(\psi) k_m^2}{\gamma\beta^2}(1 - \beta\beta_m) \tag{2.25}$$

If the mode is resonant ($\beta = \beta_m = \beta_r$) then then it decays like $1/\gamma_r^3$ and is defocusing for a stable resonant phase (to have k_s be a real number we need $\alpha_m \cos(\psi) > 0$). If the mode is not resonant, then we have a form of Hill's equation. Hill's equation has a long history in accelerator physics [34] and we will not go into the details here. In a later section, however, we will use a separation of scales technique to show that the non-resonant motion leads to a phase-independent focusing force which allows us to stabilize the motion.

2.6 Stable accelerator design

In a conventional linac the beam is positioned at a longitudinally stable (and accelerating) phase and the transverse stability is provided by external electromagnets. This isn't possible with a DLA. Not only is the aperture small, but the resonant defocusing scales like E_0/λ_0

and so is 10,000 times larger in a DLA than in an s-band RF linac. In order to magnetically confine electrons to the DLA channel we would need a 10 MT/m quadrupole magnet, more than 1000 times stronger than the strongest quadrupoles we have built (see § 4.3). The only other force we have we have available to solve the problem is the fields of the DLA itself.

The basic idea is to alternate between the longitudinally and transversely stable regions of Fig. 2.12. Because the electron is no longer stationary relative to the wave we have violated the conditions of Earnshaw's theorem and so we are allowed to find a kind of dynamic stability. This trick, of stabilizing a system by adding a time-dependent force, has a long history in physics. The clearest analogies are to the quadrupole ion trap [117], and to the radio-frequency quadrupole (RFQ)—a type of ion linac [5]. The basic principle is also familiar to beam physicists who study quadrupole focusing in FODO (focus-drift-focus-drift) lattices [34], only in a DLA FODO refers to the y and z axes rather than the x and y axes.

The first scheme devised to stabilize motion in a DLA was studied as part of the GALAXIE compact-XFEL project [33]. Like a conventional linac, it operates at a longitudinally stable phase with strong resonant defocusing. It then adds a small phase oscillation which creates a set of non-resonant harmonics in the structure. These harmonics provide "ponderomotive" focusing which can stabilize the transverse motion. Recently, a new scheme, called "Alternating phase focusing" (APF) has made a more exact analogy to the FODO lattice (including transport matrices and a Courant-Synder analysis) [37]. It jumps back and forth between two phases which are symmetrically distributed about the peak acceleration. The phase jumps require a specialized structure, but they do have the advantage of symmetrizing the dynamics and increasing the dynamic aperture of the DLA. We will compare these schemes later, but first we will present a design using the ponderomotive scheme in [38] (a video of beam dynamics in this accelerator design is available from the presentation in [118]).

A major advantage to the ponderomotive approach is that it does not require phase-jumps. That means allows us to imprint the slowly-varying phase onto the laser with a spatial light modulator (SLM) and then image the phase profile onto the DLA (Fig. 2.14). The SLM gives us enormous flexibility, because it uses an array of voltage-controlled liquid

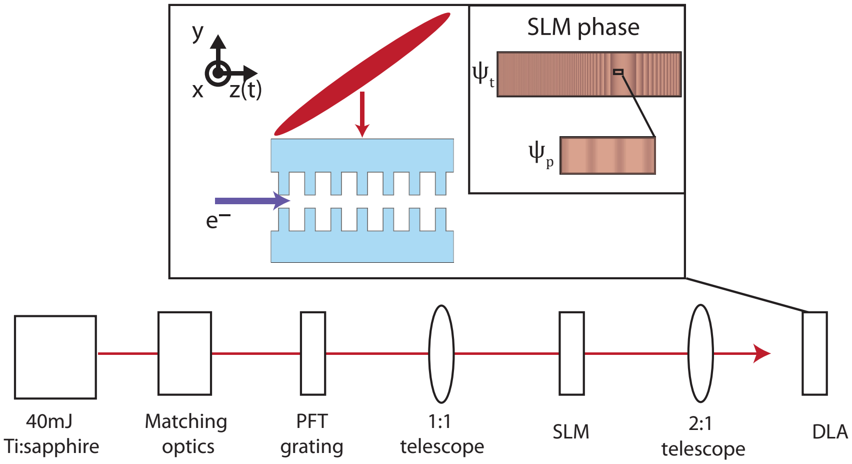

Figure 2.14: Cartoon of the proposed DLA experiment (not to scale) using a tailored phase profile imprinted onto a pulse front tilted laser beam to control the beam dynamics over a 2 cm long structure. At bottom is a block diagram of the optical design. At top is a side view of the DLA interaction, including the z dependent phase profile employed.

crystals to make a surface with a programmable index of refraction. This allows us to directly control ψ along the laser's z envelope. In the tilted pulse front configuration this overlaps almost exactly with the electron coordinate so that we can use the SLM to control the phase an electron experiences. The main constraint is that ψ vary slowly enough that we can collect the light which diffracts off the SLM and image it onto the DLA.

The accelerator we are designing is based on a realistic scenario for the Pegasus beamline. It uses a 2 cm structure driven by a 40 mJ Ti:Sapphire laser ($\tau = 70\,\mathrm{fs}$, $\lambda = 780\,\mathrm{nm}$). It first bunches a 3.5 MeV beam and then accelerates it up to 5 MeV at an average gradient of 150 MV/m. We explain the design of the accelerator by applying approximate stability criteria derived from a simplified analysis of the beam dynamics.

2.6.1 Generating non-resonant harmonics

A simple way to create the non-resonant harmonics needed for focusing is to use a sinusoidal phase modulation $a\cos(\psi_p)$ with $\psi_p \approx (\delta_k z)$. This gets added to the grating momentum

$ik_g z$[16], so that the fields inside the structure will be proportional to:

$$e^{i(k_g z + a\cos(\delta_k z))} = \sum_{m=-\infty}^{\infty} i^m J_m(a) e^{i((k_g + m\delta_k)z)} \tag{2.26}$$

The sum on the right (a Jacobi-Anger expansion) is a series of sidebands with Bessel function amplitudes $J_m(a)$ and mode spacing δ_k. Only one term in the sum can move at the resonant velocity, and for this term we explicitly write $m = r$, as in $J_r(a)$. Additionally, we will allow a and δ_k to be functions of z provided that they change very little during a beat period $(2\pi/\delta_k)$. In this case we will define $\psi_p = \int_0^z \delta_k(s)ds$ so that the instantaneous wavenumber can be broken into modes $k_g + m\delta_k(z)$. The slowly varying condition is the main assumption of our analysis.

If we apply the TM mode from Eq. 2.3 to each of these plane waves we end up with the Lorentz force:

$$\begin{aligned} F_z &= i \sum_m q_e E_1 J_m(a) \cosh\left(\Gamma_m y\right) e^{i\left(\psi + (m-r)\left(\psi_p + \frac{\pi}{2}\right)\right)} \\ F_y &= \sum_m q_e E_1 J_m(a) \frac{\Gamma_m}{k_m} \sinh\left(\Gamma_m y\right)\left(1 - \beta\beta_m\right) e^{i\left(\psi + (m-r)\left(\psi_p + \frac{\pi}{2}\right)\right)} \end{aligned} \tag{2.27}$$

where we have assumed a constant diffraction efficiency of E_1/E_0 and also that the mode is symmetric such that $c_{(1,m)} = c_{(2,m)} = E_1/(2E_0)J_m(a)e^{i\left((m-r)\left(\psi_p + \frac{\pi}{2}\right)\right)}$. To complete this expression let us clarify a number of terms: firstly, we have the mode wavenumber $k_m = \partial_z(\psi + (m-r)\psi_p)$ with its associated phase velocity $\beta_m = \omega/(ck_m)$ and Lorentz factor $\gamma_m = \sqrt{1/(1-\beta_m^2)}$; then we have the evanescent factor $\Gamma_m = \frac{k_m}{\gamma_m}$; and finally we have the phase of a particle relative to the resonant mode $\psi = k_g z - \omega t + \psi_t$, where ψ_t is a slowly changing phase we add to maintain synchronicity as the beam accelerates. It is also useful to define the normalized amplitude of each harmonic as: $\alpha_m = q_e E_0 J_m(a)/(m_e c^2 k_m)$. Note that α_m may be negative since q_e and J_m both have a sign; for example, in our design $\alpha_r > 0$ and $\alpha_0 < 0$.

---

[16] We only use the $N = 1$ diffraction order from Eq. 2.1. The other terms are too far away from resonance to effect the dynamics.

2.6.2 Ponderomotive stability

To analyze motion due to the non-resonant waves we can use the multiple-scales technique. To show how this works we will present a simplified analysis here, and then use the more detailed results from Naranjo et al. [33] when we study the transverse stability of the accelerator.

We will consider the system (following Landau [119]):

$$x'' = -\partial_x \Phi(x) + f(x)\cos(\delta_k z) \tag{2.28}$$

We are interested in variations of a linear oscillator $\partial_x \Phi(x) = f(x) = \kappa x$ with $\delta_k >> \sqrt{\kappa}$ so that the oscillatory term is very fast. The multiple-scales technique is to split x into the sum of an average (a.k.a. secular, slow) motion and a fast (a.k.a. quiver) motion, as in $x = \overline{x} + x_{\text{quiver}}$. If the oscillations due to δ_k are very fast then the quiver motion will have a small amplitude, but large derivatives. In that case we can expand the equations of motion for small x_{quiver} as in:

$$\overline{x}'' + x''_{\text{quiver}} = -\partial_{\overline{x}}\Phi(\overline{x}) - \partial^2_{\overline{x}}\Phi(\overline{x})x_{\text{quiver}} + f(\overline{x})\cos(\delta_k z) + \partial_{\overline{x}} f(\overline{x})x_{\text{quiver}}\cos(\delta_k z) \tag{2.29}$$

We make the quiver motion account for the large fast force $x''_{\text{quiver}} = f(\overline{x})\cos(\delta_k z)$ so that $x_{\text{quiver}} = -(f(\overline{x})/\delta_k^2)\cos(\delta_k z)$. The remaining terms drive $\overline{x}$, but since $\overline{x}$ is slowly varying we are justified in averaging over the quiver period. That leaves us with:

$$\overline{x}'' = -\partial_x \Phi(x)|_{x=\overline{x}} - \frac{1}{2\delta_k^2} f(\overline{x})\partial_{\overline{x}} f(\overline{x}) \tag{2.30}$$

The second term represents a secular force caused by the quiver motion. This happens even though the quiver motion is symmetric, because the quiver amplitude $|f(\overline{x})|$ is slightly different at the two turning points. By definition the quiver force points backwards at the turning points, and so the particle will experience a net, second order, motion away from the maxima of $|f|$. Another way to see this is realize that the second term can be written as a ponderomotive potential $\Phi_p = f(\overline{x})^2/4\delta_k^2$. If f has any extrema then this will be focusing, and if it is large enough then it can overcome an unstable static potential.

Longitudinal stability In our design the resonant harmonic dominates the time-averaged longitudinal dynamics. Thus we can use the Hamiltonian in Eq. 2.21 with $m = r$. In this case $\alpha_r > 0$, so we will have stable acceleration from $\psi_r = 0$ to $\psi_r = -\pi/2$.

The main competition for stability is the non-resonant harmonics. If their phase velocity is close enough to resonance then the particles will be pulled towards two separate "fixed points" and the motion can become chaotic. According to the Chirikov criterion[33, 120] the onset of this instability occurs when a particle's orbit takes it into two separatrices.

In order to check the Chirikov criteria we need to consider both the location of the seperatrices (Eq. 2.23) and the magnitude of the quiver motion. Taking the real part of Eq. 2.27 we can write the oscillatory force as $\partial_z \gamma = -\alpha_m k_m \sin(\psi + (m-r)\left(\delta_k z + \frac{\pi}{2}\right))$ (where we assume δ_k, a, and ψ_t are constants over the quiver period). On average ψ is also constant and so the energy will oscillate. For $(m - r) = 1$ the result would be:

$$\gamma_{\text{quiver}} = -\alpha_m \frac{k_m}{\delta_k} \left(\sin(\psi + (\delta_k z)) - \sin(\psi)\right) \tag{2.31}$$

As an estimate of the Chirikov criteria we can then check whether a particle on the seperatrix of the resonant bucket, oscillating with an additional quiver amplitude γ_{quiver}, will cross into the next closest seperatrix. We should also have γ_{quiver} and its associated ponderomotive potential be small enough that the dynamics are dominated by the resonant bucket. In practice, however, this latter requirement is not so strict since the non-resonant motion can actually help to stabilize the longitudinal motion.

Transverse stability After linearizing $\sinh(\Gamma_m y)$ we will find that that the transverse force is a sum over time-dependent harmonic oscillators with strengths given by Eq. 2.25. The resonant term (κ_r) will, of course, be defocusing, which is the crux of our stability problem. To overcome this resonant transverse defocusing we use the ponderomotive motion caused by the non-resonant harmonics. This yields a term (the $(B + D)$ term of equation

(3) in Ref. [33]) which is phase independent, with oscillator strength:

$$\kappa_p = \sum_{m \neq r} \left(\frac{\alpha_m}{\gamma_r \beta_r^2} \right)^2 \left(\frac{k_m^4}{2(m-r)^2 \delta_k^2} \right) \times \left((1 - \beta_r \beta_m)^2 + \frac{(1 - \beta_r \beta_m)}{\gamma_r^2} \right) \tag{2.32}$$

The total oscillator strength is $\kappa = \kappa_p + \kappa_r$, which can be focusing for all phases if $|\kappa_p| > |\kappa_r|$. It is often useful to use the expansion $(1 - \beta_r \beta_m) \approx \left(\frac{1}{\gamma_r^2} + \beta_r^2 \frac{(m-r)\delta_k}{k} \right)$. Importantly, this expansion shows that at high energy $\kappa_p \propto \gamma_r^{-2}$ while $\kappa_r \propto \gamma_r^{-3}$ so that this scheme can be stable in the ultra-relativistic limit.

Once the motion is transversely stable there remains a question of what the angular acceptance will be. The outermost oscillator will start (at rest) on the boundary and swing to have a maximum angle of $y'_{\text{max}} = \sqrt{-\kappa} y_{\text{max}}$. Only, thus far, we haven't considered the fast oscillatory motion. Just like in the longitudinal case (Eq. 2.31), we can calculate the amplitude of small oscillations by assuming ψ and y are approximately constant. We find that under the influence of harmonics $m \neq r$ the rapid motion is described by:

$$y_{\text{quiver}} = y \frac{\kappa_m}{(m-r)^2 \delta_k^2} \tag{2.33}$$

By construction this number should be small, but when calculating our aperture y'_{max} we should account for the possibility of such motion by reducing y_{max} commensurately.

2.6.3 Lattice Design

Armed with our knowledge of the approximate beam dynamics we will discuss the design of a 2 cm accelerator based on parameters available at the Pegasus facility. Since we start with an unbunched beam our accelerator will need to consist of several sections to capture, bunch, and accelerate the beam. The prototype for such a linear accelerator, the RFQ, consists of 4 subsections: a radial matcher, a shaper, a buncher, and an accelerator. In order to limit our accelerator to a manageable 2 cm we combine stages 1-3 into a rapid buncher. This causes us to have a reduced dynamic aperture and also heats the beam. Nonetheless, we consider this design suitable for a proof-of-principle experiment.

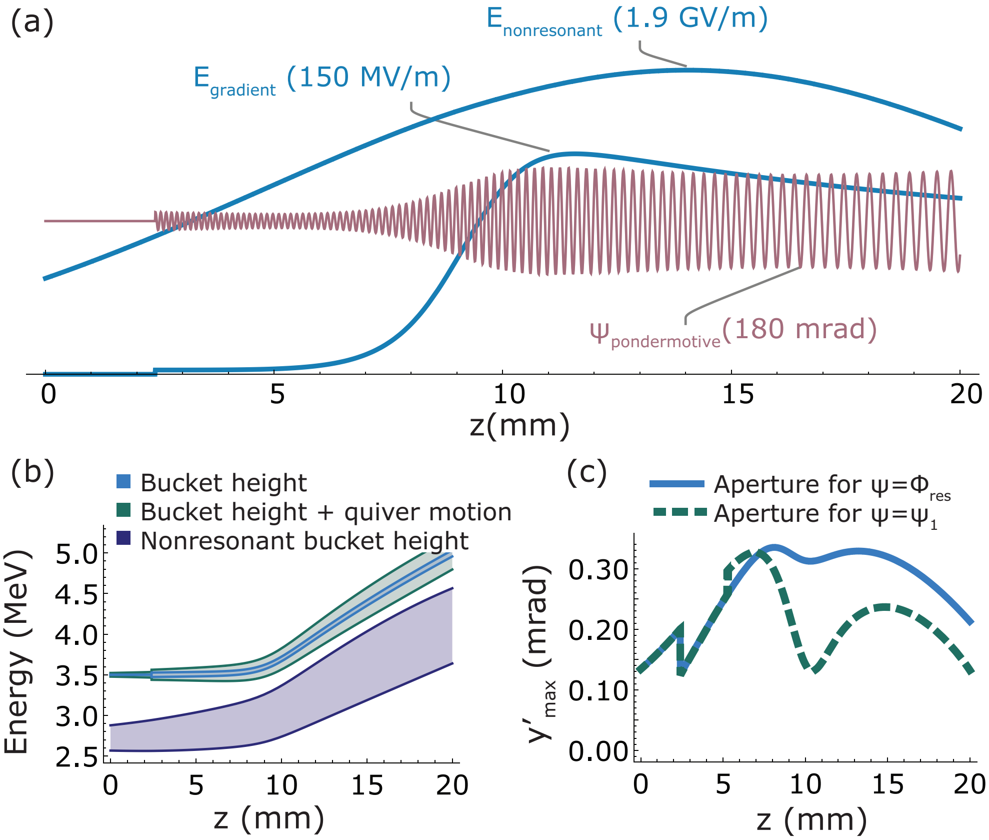

Figure 2.15: DLA lattice: (a) Curves showing the amplitude of the gradient, nonresonant harmonic, and the ponderomotive portion of the phase mask. (b) An approximate longitudinal stability criterion: particles trapped in the resonant potential can't overlap with the nonresonant potential. (c) An approximate transverse stability criterion: we plot the maximum amplitude of Betatron oscillations (negative values would be unstable).

Based on the phase mask discussed in section 2.6.1 we have three free parameters available to us to maintain stability and control the progression between stages of the accelerator. They are the mode spacing δ_k, the mode amplitude a, and the phase ψ_t which is used to control the resonant phase of the accelerator, ψ_r. The mode spacing can be changed without affecting the behavior of the resonant mode, and so we will use it to control the ponderomotive focusing. The mode amplitude lets us control the ratio of the resonant to non-resonant harmonics, and so we will use it control the (resonant) seperatrix height. Finally we will use the resonant phase to control the progression of the accelerator from capture to acceleration. Once we have progressed through the accelerator and choosen values for these parameters we will be ready to integrate the equations of motion.

Before choosing values for these free parameters we need to define the laser available to us. Based on a slightly optimistic reading of previous measurements [29] we will model an accelerator which can diffract 2 GV/m into the first Bloch mode ($E_1 = 2\,\mathrm{GV/m}$). But after losses in the optical transport of Fig. 2.14, we will only have 8 mJ of Ti:Sapphire (wavelength $\lambda = 800\,\mathrm{nm}$) with which to illuminate a structure of 2 cm length in z, and so we choose to focus the laser to a spot size of $w_x \times w_z = 40\,\mu\mathrm{m}\times 1.3\,\mathrm{cm}$. This laser spot under-fills the structure (see Fig. 2.15(a)), but the leading edge of the Gaussian is sufficient to power the buncher and then the peak of the Gaussian can be used to accelerate.

Next we will choose which of the sidebands from Eq. 2.26 will be the resonant harmonic (the one we've been calling r). We will need to use most of the power for ponderomotive focusing, so we should make $J_r(a)$ small. To accomplish this we choose $r = -1$ and use values of $a << 1$ so that most of the energy is in J_0. This choice is convenient because we can think of the motion as being due to just two terms: the $m = 0$ and $m = -1$ harmonics; with the power splitting controlled by $a(z)$. We choose $r = -1$ instead of $r = 1$ so that the non-resonant harmonic is slower than the resonant harmonic and κ_p is focusing for all γ.

Now we are in a position to actually design the accelerator components by choosing a, δ_k and ψ_r. Our choices will result in the accelerator shown in Fig. 2.15(a). We plot $\psi_p \approx a\cos(\delta_k z)$ to show the ponderomotive phase. The focusing is strongest when it has low amplitude (so all the power is in the $m = 0$ non-resonant harmonic) and long period δ_k. We also directly plot the envelopes for the nonresonant field $E_1 J_0(a)$ and for the accelerating gradient $E_1 J_{-1}(a)\sin(\psi_r)$.

Buncher The first thing we need to do is bunch the beam so that all the electrons are grouped near ψ_r and can be captured by the resonant harmonic. At the same time we need to transversely capture the beam and so we would like to do the bunching with the non-resonant harmonics turned up as strongly as possible. We can do this by setting $a = 0$ so that all the power is in the $(m - r) = -1$ harmonic and then letting the longitudinal ponderomotive force do the bunching [36, 37].

This is a second order effect driven by the quiver motion (Eq. 2.31). Each particle will oscillate around a different central energy determined by its initial phase ψ and so the higher energy particles will begin to outrun the slower energy particles. In the language of the ponderomotive expansion, the oscillator strength f is a constant so this second order motion shows up as a drift:

$$\psi^{(2)} = \psi + \frac{\alpha_m}{\gamma_r^3 \beta_r^3} \frac{k_m}{\delta_k} \sin(\psi) z \tag{2.34}$$

The drift leads to bunching when $\partial_\psi \psi^{(2)} = 0$. For negative α_m the bunching is centered around $\psi = 0$, which is a perfect location to be trapped and then accelerated by the resonant harmonic. When bunched, the energy spread will be exactly the quiver amplitude, $\alpha_m \frac{k_m}{\delta_k}$, and so we see that there is a trade-off between the buncher length and the amount of heating imparted by the buncher.

As a compromise between the buncher length, the heating, and the ponderomotive focusing (which favors small δ_k), we start with $\delta_k(0) \approx k_g/150$. This makes the initial transverse acceptance 0.15 mrad and causes the beam to bunch in about 2.5 mm.

Acceleration Immediately after bunching we increase a (seen as a discontinuity in the plots of Fig. 2.15) to establish a resonant harmonic at ψ_r=0, where the beam has bunched and where the seperatrix is largest. The initial value of a is set to make the bucket height (Eq. 2.23) more than large enough to include the entire energy spread. Then we slowly, but arbitrarily, increase ψ_r to turn on an accelerating gradient as shown in Fig. 2.15(a). Increasing ψ_r would tend to shrink the bucket, so to keep the particles trapped we gradually increase a in order to keep the bucket height constant. After $\approx 1\,\mathrm{mm}$ we reach the maximum accelerating gradient of 150 MV/m at $a = 0.18\,\mathrm{mrad}$ and $\psi_r = -\pi/3\,\mathrm{rad}$. At this point we have defined enough free parameters to calculate the resonant energy:

$$\gamma_r = \int_0^z \frac{q_e E_1(s)}{m_e c^2} \sin(\psi_r(s)) J_{-1}(a(s)) ds \tag{2.35}$$

Stable acceleration The only remaining task is to choose δ_k to stabilize the trajectories. We want to decrease δ_k in order to increase the ponderomotive focusing, but doing so also makes the phase velocity of the non-resonant harmonic closer to the resonant velocity. At some point we will violate the Chirikov criterion and the longitudinal motion will become chaotic.

We can see this balancing act in Fig. 2.15(b,c). In (b) we try to keep the buckets from overlapping so that the longitudinal motion remains predictable and the particles stay trapped around the resonant phase. In (c) we plot the angle y' of a particle whose maximum position lies at the edge of the DLA gap. The result is an approximate picture of the acceptance area of the DLA which we try to keep above zero along the accelerator.

Maintaining resonance In order to complete the design, we have to go back and calculate the taper phase, ψ_t, which is used to keep the resonant particle synchronous with the same sideband throughout the accelerator. To do so we track a test particle sitting at ($\psi = \psi_r(z)$, $y = 0$). By construction it will have energy γ_r calculated from Eq. 2.35 and knowing this we can calculate its time of arrival, t_r at position z as $ct_r = \int_0^z (1/\beta_r) ds$. Then we construct ψ_t to keep the total phase in Eq. 2.27 constant: $\psi_t = \omega t_r - k_g z + \psi_r$.

2.6.3.1 Dynamic Aperture

To simulate the performance of this accelerator design we track 1500 particles through the DLA under the forces of Eq. 2.27, without consideration of collective effects. The ensemble

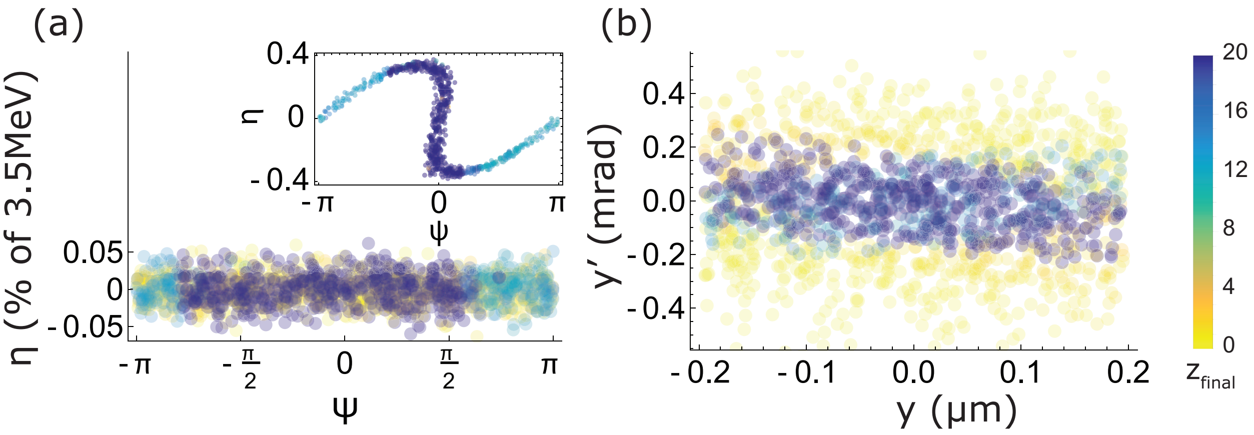

Figure 2.16: Dynamic aperture: (a) longitudinal phase space and (b) transverse phase space at the input to the DLA. Particles are color-coded by how far they make it into the accelerator (up to a maximum of 2 cm). The inset to (a) shows a phase portrait after the buncher (z=2.5 mm).

we sample from is representative of the Pegasus beam: it has a mean energy of 3.5 MeV, it starts unbunched with negligible energy spread, and it has a normalized transverse emittance of 20 nm. The vacuum gap of the DLA is only 0.4 µm, so for simplicity we stop tracking particles which cross $|y| \geq 0.2$ µm.

The resulting phase space acceptance is illustrated by color-coding the initial phase space based on how far the particles travel (Fig. 2.16). Most of the particles are lost instantaneously (yellow) because their initial angle carries them into the wall of the DLA. Of the rest, there are two groups: those that get bunched (purple), and those that remain out of phase (blue), as shown in the inset. The bunched particles will end up seeing a net defocusing force, while the opposite is true for the out of phase particles, thus the transverse acceptance for these two groups is slightly skewed (as can be seen in Fig. 2.16(b)). The out of phase particles are eventually lost when the resonant phase ramps up and the defocusing force becomes stronger.

In total, we capture 70% of the particles that survive the first 1 mm, and about 30% of the all the particles shown in Fig. 2.16. If we account for particles which never made it into the structure, either because they started outside the vacuum gap or because they were

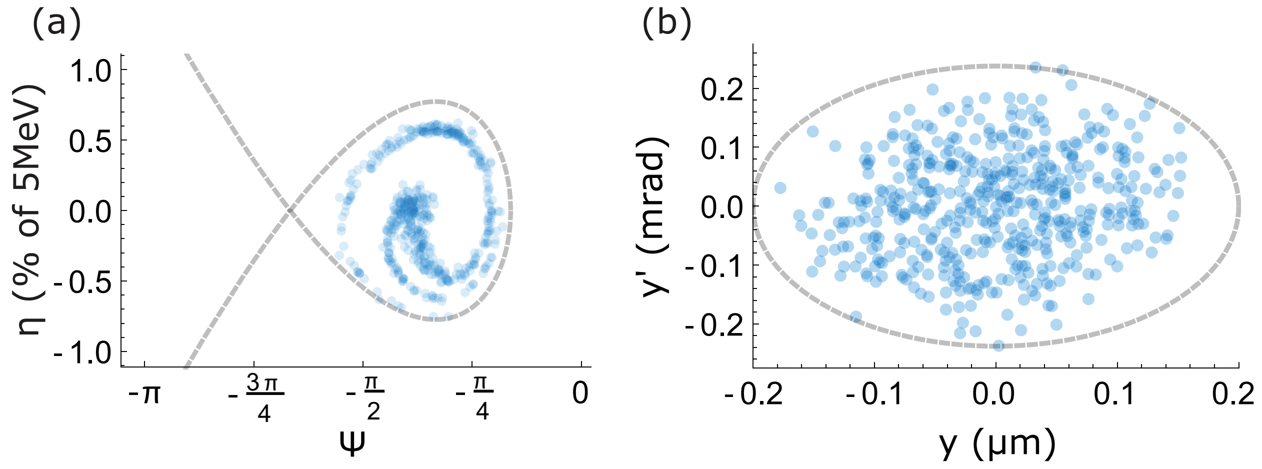

Figure 2.17: (a) longitudinal and (b) transverse phase-spaces averaged over the last 0.4 mm of the DLA. Dashed lines show the linearized buckets of section 2.6.1.

temporally mismatched (the electron beam is 100 fs long, but the laser is only 60 fs long) then we get a total transmission on the order of 0.1%. Assuming an initial charge of 50 fC, we can expect roughly 20 electrons/bunch for 15 bunches. This number can be improved by several orders of magnitude if we use a flat beam transform to improve the 1D brightness [121] and if we velocity bunch the electron beam enough to eliminate the temporal mismatch [82].

2.6.3.2 Output phase space

We can compare the numerical solutions to the secular dynamics obtained by time-averaging the particle motion. In Fig. 2.17 we show a phase space portrait of the beam averaged over the last 0.4 mm of the DLA (i.e. the length of the final beat period, $2\pi/\delta k$). We can see that the longitudinal dynamics are very well described by the time averaged bucket, while the transverse dynamics appear more tightly bound (in y) than might be expected since the quiver motion kills particles near the boundary.

If we were to show an instantaneous phase portrait, then the non-resonant harmonics would significantly alter the distribution (see the video in [118]). For example, the transverse dynamics have a large $y - y'$ correlation which oscillates as the nonlinear harmonics slip by. A similar effect happens longitudinally, where the beam centroid oscillates up and down relative to $\eta = 0$ and also the distribution gets sheared by nonlinear focusing. Nonetheless, the motion is stable and can be profitably regarded in the time-averaged sense as analogous to a conventional accelerator undergoing both synchrotron and betatron oscillations.

2.6.4 Comparison to alternating phase focusing

At the beginning of this section we mentioned that there is another scheme called alternating phase focusing (APF) which is exactly analogous to the FODO lattice. It works by making discrete phase jumps between transversely stable and longitudinally stable regions (instead of smoothly varying the phase, as in our ponderomotive scheme). Niedermayer [37] analyzes this system by linearizing all the forces and then using a matrix based analysis, but it is also possible to analyze the motion using the ponderomotive method. We will do that here in order to more easily compare the techniques.

Firstly, we need to define the phase modulation. We use a square wave S with amplitude a and period δ_k so that $E_z(y=0) \propto \exp(i(k_0 z - \omega t + \pi/2 + S))$. Next we want to break this into plane waves. We can do this by first expanding $\exp(iS) = \cos(S) + i\sin(S)$ and then noticing that S is symmetric about zero such that $\cos(S) = \cos(a)$ and $\sin(S) = \sin(a)S/a$. As a result:

$$E_z(y=0) \propto \exp(i(k_0 z - \omega t + \pi/2)) \left(\cos(a) + i\sin(a) \sum_{m=0}^{\infty} \frac{4}{\pi} \frac{1}{2m+1} \sin((2m+1)\delta_k z) \right) \tag{2.36}$$

The $\pi/2$ is inserted so that the $\cos(a)$ term is accelerating. The $\sin(a)$ terms are then a set of non-resonant harmonics with mode spacing $(2m+1)\delta_k$.

This structure is very similar to the one we have already used (Eq. 2.26). The amplitude of the phase oscillation, a, again controls the trade-off between acceleration and focusing. And the mode spacing δ_k should again be able as small as feasible to improve the focusing strength. The main differences from our scheme is that APF has many harmonics of similar amplitude and that that APF puts its “resonant” harmonic at $\pi/2$ (max acceleration), while ours is centered around ψ_r. This means that the APF scheme uses ponderomotive focusing in the longitudinal direction as well as the transverse direction.

Of course, strictly speaking, the square wave would require an infinite amount of energy to create. That is because the large m terms will have a steep exponential growth away from the middle of the gap (where we are specifying their amplitude). In practice, however, we can always truncate the series to make a practical mode. And, from the ponderomotive analysis, we know that it is mostly the nearby harmonics (small δ_k) that cause focusing, so this shouldn't effect the motion very much.

There is one other major difference between APF and the ponderomotive scheme, and that is the approximation each uses to simply the motion. For one thing, the APF linearizes both the longitudinal and transverse forces, while the ponderomotive scheme only linearizes the transverse force. For another, and more importantly, the APF also makes an assumption about how the square waves work: namely it assumes that both the z kick and the y kick are perfectly square. But actually it is only the *on-axis z kick* that we have engineered to be square. And, since every z harmonic is multiplied by a different amplitude ($\cosh(\Gamma_m y)$), the sum won't build a square wave off-axis. The consequences are even more important for the transverse kick, where each harmonic is multiplied by $\sinh(\Gamma_m y)(1-\beta\beta_m)$ [17]. Even on axis this is not a square wave! The assumption [37] makes to simply the APF scheme amounts to assuming that the oscillator strength from Eq. 2.25 is equal to the resonant oscillator strength for every mode.

This assumption is justified at low energy or at small δ_k. To see this we can write the oscillator strength (in the limit of small δ_k) as:

$$\begin{aligned} \kappa_r &= -\frac{\alpha_r \cos(\psi) k_r^2}{\gamma_r^3 \beta_r^2} \\ \kappa_m &\approx \kappa_r \frac{4}{\pi}\frac{1}{2m+1}\left(1 + \frac{(2m+1)\delta_k}{k_r}\left(1+\gamma^2\right)\right) \end{aligned} \tag{2.37}$$

where the prefactor $\frac{4}{\pi}\frac{1}{2m+1}$ comes from the square wave amplitude. Looking at this expression we can see that the assumption in [37] amounts to neglecting the term $\left(\frac{(2m+1)\delta_k}{k_r}\left(1+\gamma^2\right)\right)$. Since δ_k is small by assumption (and assuming we have truncated the square wave), this assumption will be valid when $\frac{\delta_k}{k_r}\gamma^2 << 1$.

---

[17] The term $(1-\beta\beta_m)$ comes from the cancellation of the electric and magnetic forces.

A major advantage of the APF scheme, however, is that we can use a Courant-Synder analysis to find the matched (electron) beam envelope. The envelope will oscillate throughout the lattice, and our goal is keep its maximum size smaller than the gap of the DLA. This quantity is discussed and plotted in [37], but we can go a little further and derive an analytical expression to show how the envelope will scale. The derivation is a common beam physics problem [122] and the algebra is not enlightening, so we will only outline the steps. First, we observe that the envelope will reach its maximum size in the middle of the lenses. Second, we combine thick-lens matrices to describe the transport of a full FODO cell starting and ending at the middle of a lens (actually we should call it an "FD" cell since we are using two thick lens and no drifts). From this matrix we can calculate the $\hat{\beta}$ function as $\hat{\beta} = R_{12}/\sqrt{1-R_{11}^2}$. $\hat{\beta}$ will be a function of the FODO cell length, $L = 2\pi/\delta_k$, and the oscillator strength κ_r. We can then minimize this $\hat{\beta}$ with respect to L. This requires solving the transcendental equation $\mathrm{sech}(\sqrt{\kappa_r}L/2)^2 = \cot(\sqrt{\kappa_r}L/2)\tanh(\sqrt{\kappa_r}L/2)$.

Numerically we find that $L \approx 2.5/\sqrt{\kappa_r}$ and $\hat{\beta} \approx 4.2/\sqrt{\kappa_r}$. We can convert this to a beam envelope using $\sigma = \sqrt{\hat{\beta}\epsilon_{\text{geo}}}$, where ϵ_{geo} is the geometric emittance. The geometric emittance decreases as γ when we accelerate a beam, but even so we can see that that σ will diverge as $\sqrt{\gamma}$. This may be true for the longitudinal dynamics, but it actually isn't true for the transverse dynamics because it is precisely at large γ where the approximation $\frac{\delta_k}{k_r}\gamma^2 << 1$ breaks. In fact, with $L = 2\pi/\delta_k$ we can see that the left side of that inequality will also diverge with $\sqrt{\gamma}$. In the ponderomotive analysis, however, the transverse motion won't diverge because $\kappa_p \propto 1/\gamma^2$, which scales perfectly to keep $\hat{\beta}$ constant.

We can show this behavior in Fig. 2.18 where we are comparing it to the ponderomotive scheme. We have plotted $\hat{\beta}/\gamma\beta$ to account for adiabatic damping. This way we can see the envelope of an accelerator "design" by charting a path from low to high energy. The divergent behavior of the APF scheme then shows up in panel (b) as an increase of $\hat{\beta}/\gamma\beta$ along all paths, including on the optimal path (which we mark with the dotted a line). By contrast, we can see that the ponderomotive scheme has a clear path to high energy at constant $\hat{\beta}/\gamma\beta$ and can thus remains stable over a large distance.

To gain this stability, however, the ponderomotive scheme pays a steep price. Both of

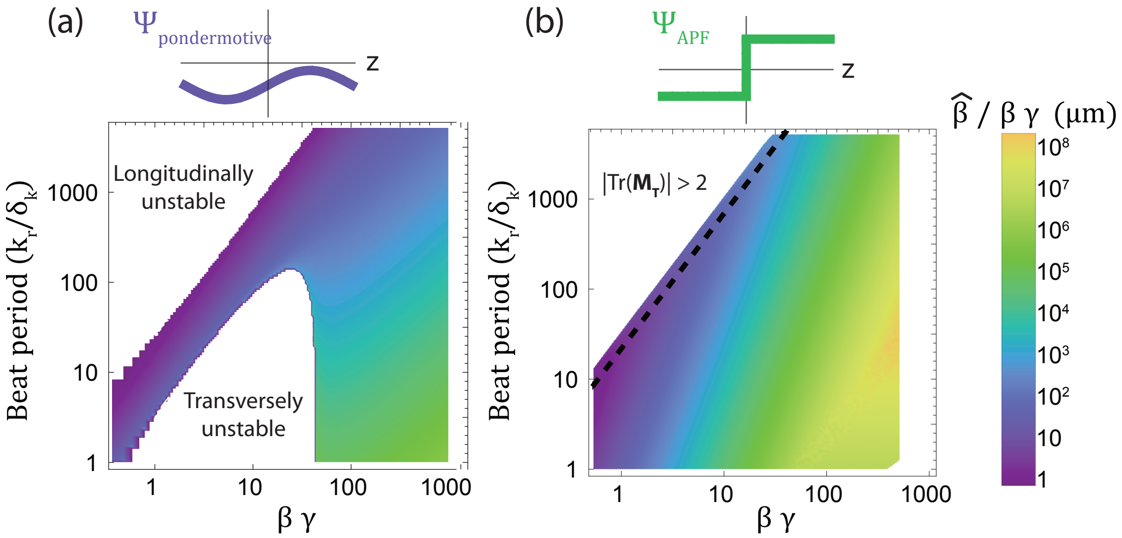

Figure 2.18: Stability of two DLA schemes. (a) The ponderomotive scheme and (b) The APF scheme. Both schemes are driven by a 2 GV/m mode at 800 nm, but the cartoons at the top show how their phase variations differ. The APF envelope eventually diverges, but before that it makes much more efficient use of the available electric field.

these plots are for 800 nm lasers with $E_1 = 2\,\text{GV/m}$. But the APF scheme (working at $a = \pi/4$) has an accelerating gradient of 1.4 GV/m, while the ponderomotive scheme we show here has a gradient of only 40 MV/m[18]. Moreover, the ponderomotive scheme has a complicated stability diagram: make δ_k too small and you violate the Chirikov criteria, but make it too large and the resonant defocusing dumps the beam into the wall. By contrast, the APF dynamics are perfectly symmetric, so that all you need to do is keep the cell length short enough that the phase advance is real (i.e. the trace of the transfer matrix is less than 2).

The APF scheme is so much more successful at getting high gradients because it symmeterizes the problem. In terms of the ponderomotive analysis we would say that it is operating around a resonant phase of $\psi_r = -\pi/2$. This means that, on-average, there is no resonant defocusing to overcome. On the other hand, there is no longitudinal stability from

---

[18] This comes from setting the ponderomotive amplitude $a \approx 0.05$ and thus splitting almost all of the power to the $m = 0$ non-resonant mode. We also have $\psi_r = -\pi/3$. It is lower than the 150 MV/m design we showed earlier because we wanted the stable region to be large enough to see on the contour plot.

the resonant harmonic, only the non-resonant terms. It is possible that we can combine the best of both schemes by operating very close to $\pi/2$—closer than would seem kosher from the resonant bucket—and thus have stability at high energy, but without paying such a high price to overcome the resonant defocusing.

2.7 Energy and momentum conservation in a DLA

So far we have shown how it is possible for a DLA to accelerate electrons without pausing to consider where the energy comes from. But, if we look back at the DLA geometry from Fig. 2.1, we can see an apparent contradiction: the drive laser, which is carrying momentum in the y direction, is somehow supposed to give z momentum to the electron beam. Clearly the DLA must absorb the missing momentum and the laser must loose some energy, but how it happens takes some consideration.

The basic idea is that the DLA diffracts or scatters the light into a mode which carries momentum in the z direction. To see that this is true, we can calculate the Poynting vector for the TM mode. Note that we have to be careful when multiplying the field components, because in Eq. 2.3 we have written them as complex quantities and implied that we should take the real part. We can account for this by adding a complex conjugate to each term before multiplying. If we specialize to a cosh mode the result is:

$$\boldsymbol{S} = \frac{E_m^2}{2Z_0}\Re\left[\gamma_m\beta_m\cosh(\Gamma_m y)\sinh(\Gamma_m y)\left(-ie^{2i\psi}+i\right)\hat{y} + \beta_m\gamma_m^2\sinh^2(\Gamma_m y)\left(e^{2i\psi}+1\right)\hat{z}\right] \tag{2.38}$$

Since this mode is evanescent, the y component must time-average to zero ($\langle\exp(i\psi)\rangle = 0$), but the z component does not. It has a static component carrying momentum in the z direction which we can use to accelerate the electron beam. This is not some fluke of the modal analysis: with a bit of algebra we can verify that this mode conserves conserves energy ($\partial_t u + \nabla\times\boldsymbol{S} = 0$) and momentum ($\partial_t\frac{\boldsymbol{S}}{c^2} = \sum_\beta \partial_{x\beta}T_{\alpha\beta}$) inside the vacuum gap, and so it is valid to think about it isolation. That means it must be the DLA which does the trick of converting y momentum to z momentum.

That the dielectric is able to change the momentum of light shouldn't be surprising. Every mirror, every diffraction grating, and even every flat surfaces does this. In fact, it is a general feature of evanescent modes that they carry momentum parallel to the surface of the material. During total internal reflection this causes a lateral shift of the reflected beam (the Goos-Hänchen effect [103]). It also shouldn't be surprising that the evanescent mode can transfer energy because we often see evanescent modes converted back to propagating modes by inserting another piece of dielectric (such as in frustrated total internal reflection, or such as between the two halves of our DLA).

What is, perhaps, mysterious is where the reaction force on the DLA come from. We could just assert conservation on the macroscopic scale and say the electron beam pushes the DLA backwards and the laser pushes it forwards, but that doesn't explain what is happening. We can demystify this by drawing parallels to other more familiar examples, like the radiation pressure on a mirror. In classical electromagnetism this pressure comes from the Lorentz force of the light pushing on the bound surface current in the mirror [123]. The same thing happens in the DLA: we can think of the interaction as occurring between the electron beam and a bound surface current at the boundary of the DLA.

Another way to think about the problem is to realize that an electron beam going through the structure will radiate away energy. If we could solve the Green's function problem for an electron in a DLA then we would be able to calculate this "wakefield", E_w, and see how it contributes to the stress-energy tensor. When we go to calculate $\boldsymbol{T}$ (or $\boldsymbol{S}$ or u) we will get interference with terms like $|E_0^2 + 2E_0E_w + E_w^2|$ and the cross term $2E_0E_w$ can account for the interaction by removing energy wherever E_w overlaps with E_0. In DLA $E_0 >> E_w$, so we don't have to account for that change in energy, but in principle putting lots of charge into the DLA will "load the wake" and reduce the accelerating gradient.

However we look at it, the result is always that electromagnetism conserves energy for this system. If we had a different system, with $E_w >> E_0$ then we could fret about the radiation reaction, but even that would not last long as E_0 is fed by the electron beam. On a microscopic level, though, we can still ask how the dielectric is able to scatter light. Here we run out of cute answers, because any analysis of scattering assumes the conservation laws (or motivates them with Noether's theorem). At some level the conservation laws are just empirical fact. And in the end, DLA exchanges such little energy, that we can simply ignore them at let nature take care of the details.

CHAPTER 3

DLA measurements

The DLA measurements at Pegasus were performed with the beam-line configured as in Fig. 3.1. The basic idea is that the Pegasus beam will sample the fields inside the DLA, and then we will measure the beam's energy spectrum to see what happened. This is easier said then done, because we will have null results (no change to the spectrum) unless we overlap the electron beam in space and time with an accelerating mode having the same phase velocity as the electron beam. Thus measuring any interaction at all requires a great deal of attention to the accelerator, the optics, and their relative alignment.

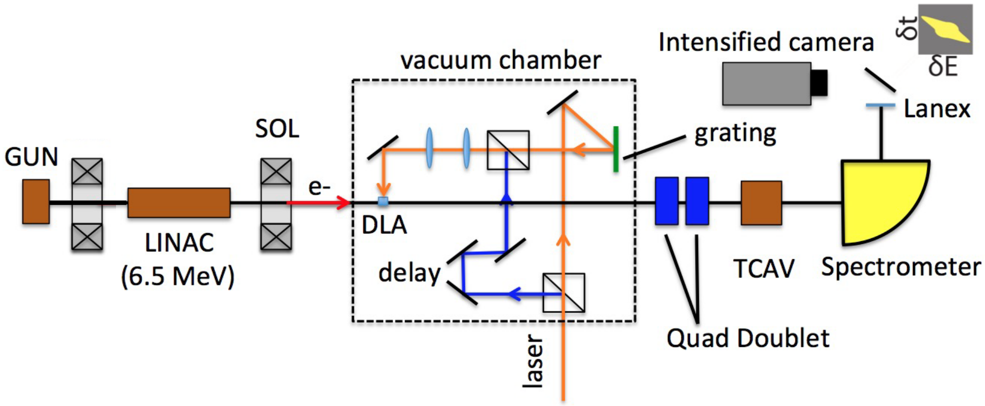

Figure 3.1: Experimental design for a spectroscopic DLA measurement. The electron beam moves from left to right starting at the gun and ending at an electron spectrometer. The laser can be set up with or without pulse front tilt.

To simplify things, we have a made an optical design which allows us to easily switch between flat and pulse front tilt configurations without loosing the alignment (blue and orange lines in the cartoon). We can also easily scan the laser energy, group velocity dispersion, and angle of incidence. Having so many parameters to scan helps us gather a significant amount of data on the DLA structures, which we can then use to confidently fit the diffraction efficiency ($|c_1 + c_2|$).

Alternatively, we can calculate the diffraction efficiency using our simulations from section 2.4.2. In this case, we can use the measured spectra to help reconstruct what happened to the beam inside the accelerator. This gives us a way to connect the measurements with the physics we established in the previous chapter.

In order to fully explain our observations we will start this chapter with a discussion of experimental techniques. We will discuss the optical design, what parameters we chose for the Pegasus beam, and how we were able to get spatio-temporal alignment between the two. Then we will turn to actual DLA measurements, where we will compare the flat and PFT measurements in a variety of circumstances. This will help illustrate the physical mechanisms at work inside the DLA and thus build the case that DLA can be extended to a longer, more stable, accelerator.

3.1 Laser beamline

The optical design for DLA at Pegasus [94] is illustrated in Fig. 3.2 (In Fig. 3.1 we show the same path, but folded up to fit inside the vacuum chamber). The basic scheme consists of two paths: one with a tilted pulse front and one without. The splitting between the two is done with polarizing beamsplitters so that the ratio of light along each arm can be adjusted by a waveplate (although only one arm at a time has the correct polarization for accelerating electrons). In addition to these two configurations, we can switch to a third by inserting a lens to focus the (non-tilted) beam down to a small spot size for use in a timing diagnostic.

Most of the complication in this design comes from trying to create a laser with tunable pulse front tilt. The rest of the optics are simply there to control the spot-size and radius of curvature of the laser. Thus we will start by explaining how the pulse front tilt is created, and then the rest of the design will fall into place.

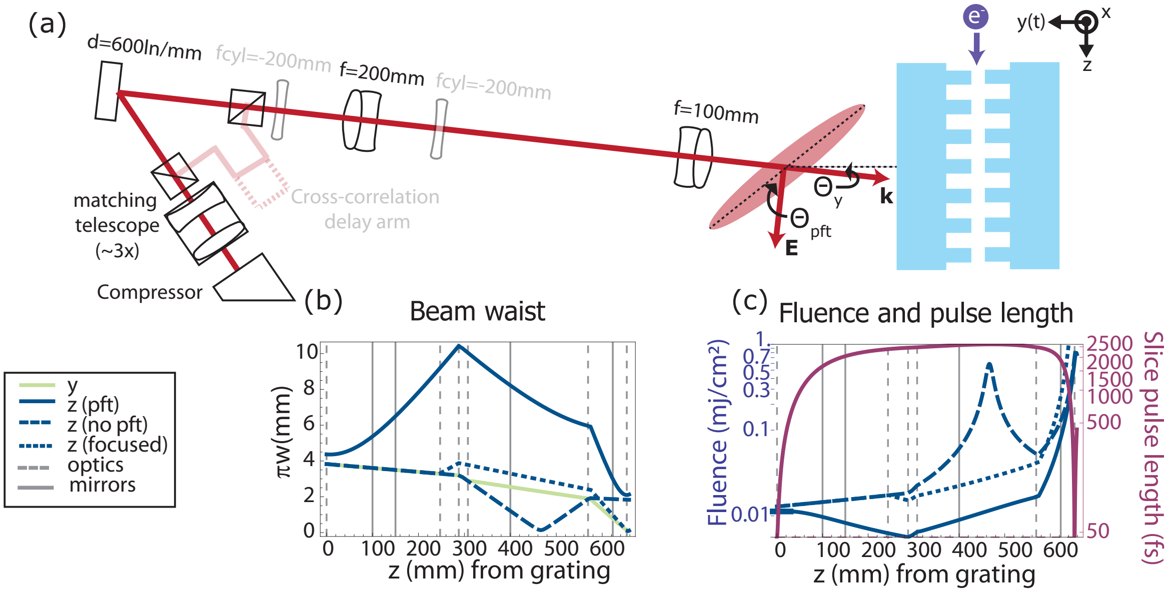

Figure 3.2: Illustration of the laser beamline. (a) Cartoon showing the optical layout. A two lens telescope images the PFT grating onto the DLA with variable angular magnification. Cylindrical lenses control the laser in the non-dispersive plane. (b) Spot size in 4 operating configurations. (c) Fluence (for 300 mJ) and slice pulse length (on-axis) used to check damage thresholds.

Pulse front tilt The technique for creating PFT with a diffraction grating is borrowed from the single-cycle THz community [106, 124]. It works because a diffraction grating causes the angle of incidence to differ from the angle of reflection, and thus the path length of a ray depends on its transverse position. This causes the pulse envelope to develop the z–t correlation we call pulse front tilt (while the phase fronts remain perpendicular to the direction of travel).

Unfortunately, that is not all the diffraction grating does. It also causes the beam to disperse, and as the colors spread out we get a z–ω correlation which combines with the PFT to make an ω–t correlation. Consequently, our femtosecond laser pulse balloons to be a picosecond laser pulse (see Fig. 3.2(c)). This is not at all desirable for DLA, where we need a short pulse in order to outrun material damage.

To eliminate the dispersion, we can image the plane of the diffraction grating onto the

DLA. This causes all the colors, which had spread out after the grating, to be recombined at the image plane. At the same time, it scales the angular magnification of the beam by a factor of m_θ, resulting in a pulse front tilt angle of:

$$\theta_{\mathrm{PFT}} = \arctan\left(\frac{\lambda_0 m_\theta}{d\cos(\theta_d)}\right) \tag{3.1}$$

where d is the grating pitch, and θ_d is the diffraction angle which satisfies the condition:

$$d\left(\sin(\theta_i) + \sin(\theta_d)\right) = N\lambda_0. \tag{3.2}$$

Finally, the pulse front tilt angle is related to the pulse front velocity by $\beta_{\mathrm{PFT}} = \cos(\theta_{\mathrm{PFT}})\csc(\theta_{\mathrm{PFT}} + \theta_y)$, where θ_y represents laser's angle of incidence with respect to the DLA (see Fig. 3.2). The incidence angle θ_y is used to control the phase velocity of the mode, and it is usually small enough to have negligible effect on the PFT.

Design procedure Now that we know how to create the PFT, we are ready to design an optical system for the DLA experiment. The minimal design must create a high intensity laser with a flat phase and a pulse front tilt of 45.1° (to match an electron energy of 6.5 MeV). Specifically, we want to be able to control 5 parameters: the spotsize in x, the spotsize in z, the radius of curvature in z, the PFT angle, and the dispersion. Several of these can be controlled by the "matching optics" at the start of the setup, but the PFT angle and the dispersion can only be controlled by imaging the diffraction grating onto the DLA.

In theory we could image the diffraction grating with one lens, but then the PFT angle would be fixed. To create a more flexible system we use a two-lens telescope (the solid black lenses in Fig. 3.2). We can then tune the PFT by adjusting the spacing between the lenses to control the angular magnification (and thus the PFT, via Eq. 3.1).

In addition to the angular magnification, the PFT also depends on our choice of diffraction grating and diffraction angle θ_d. But there is not so much freedom here as might initially be supposed, because the only way to image the surface of the diffraction grating is to make sure the laser exits perpendicular to the surface of the grating (i.e. $\theta_d = 0$) [125]. We also want the angle of incidence to be large enough to easily separate the incoming and outgoing paths.

Among commercially available gratings, the best we can choose is a 600 ln/mm diffraction grating which will work with an incidence angle of 29°. On its own, it produces a pulse front tilt angle of 26°, and thus the two-lens telescope should have a magnification of roughly 2.

Since we are imaging the grating, the spot size of the laser at the DLA is tied to its size at the diffraction grating. Hence the need for matching optics (i.e. another telescope) to control the size at the diffraction grating. The matching optics aren't quite enough, however, to finish the design. We actually want to insert cylindrical lenses in order to make a tight focus in the x (non-dispersive) plane (where the electron bunch is only $\sigma_r = 10\,\mu\text{m}$ wide). Such a focus allows us to have drastically different fluences at the diffraction grating and at the DLA so that we can take full advantage of the large damage threshold of fused silica [96] compared to the diffraction grating [126].

Once we have chosen focal lengths for the lenses and set the total path length, then the system is fully defined by the target laser parameters. In order to solve for all the lens positions and calculate the spot-size throughout the beam-line we turn to a matrix-based ray-tracing model. Then, to avoid problems from clipping, we demand that $\pi w < d_{\text{optic}}$ throughout the beamline (where d_{optic} is the clear diameter of the optic).

3.1.1 Kostenbauder matrices

The matrix approach to beam optics is familiar to both laser physicists [104] and accelerator physicists [34]. For lasers, the conventional approach uses 2×2 ("abcd") matrices to track the transverse position position and angle of rays through a series of lenses. Kostenbauder [127] extended this approach to 4D "ray-pulse" matrices which can track ultrashort pulses. Since Kostenabuder, the Ray-pulse matrices have been extended to the full 6D case [128], however our optics are only coupled in the z, t planes (recall that z is a *transverse* coordinate, since it gives the direction along the DLA accelerator — see Fig. 3.2), and so we can get away with a 4D description (with an additional 2D matrix for the x plane).

In 4D, a ray is defined by its transverse position, transverse angle, longitudinal position (relative to a well-defined reference trajectory), and frequency: (z, θ, t, f). The canonical example of a laser envelope which can be characterized by such a ray is the bi-variate Gaussian: $e^{a(z)^2+2b(zt)-c(t)^2}$, but the approach works for any envelope which can be made of "paraxial" rays. The entire optical system can then be reduced to a set of linear relations:

$$(z, \theta, t, f)^T = \boldsymbol{M}(z, \theta, t, f)^T \tag{3.3}$$

The matrix $\boldsymbol{M}$ tells us how a ray propagates through the system. Thus arguments like the ones that led us to the pulse front tilt angle (Eq. 3.1) would allow us to derive the elements of $\boldsymbol{M}$ for common systems. For example [127]:

$$\begin{array}{|c|c|c|}\hline \begin{pmatrix} A & B & 0 & 0 \\ C & D & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & \frac{L}{n(\lambda_0)} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\partial_f v_g \frac{L}{v_g^2} \\ 0 & 0 & 0 & 1 \end{pmatrix} & \begin{pmatrix} -\frac{\cos(\theta_d)}{\cos(\theta_i)} & 0 & 0 & 0 \\ 0 & -\frac{\cos(\theta_i)}{\cos(\theta_d)} & 0 & \frac{N\lambda_0^2}{cd\cos(\theta_d)} \\ \frac{N\lambda_0}{cd\cos(\theta_i)} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ \hline \end{array} \tag{3.4}$$

On the left we have $\boldsymbol{M}$ for a generic non-dispersive system, in the middle we have a dispersive slab with index of refraction $n(\lambda_0)$ and group velocity v_g, and on the right we have a diffraction grating of order N with pitch d, at incidence angle θ_i and diffraction angle θ_d. λ_0 is the vacuum wavelength of light.

Most of the optics in our setup are non-dispersive, and so we can start with a generic 2D analysis. For example, the telescope which images the DLA is approximately described by:

$$\begin{aligned} \boldsymbol{M} &= \begin{pmatrix} 1 & f_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f_2} & 1 \end{pmatrix} \begin{pmatrix} 1 & f_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & f_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f_1} & 0 \end{pmatrix} \begin{pmatrix} 1 & f_1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & f_2 \\ -\frac{1}{f_2} & 0 \end{pmatrix} \begin{pmatrix} 0 & f_1 \\ -\frac{1}{f_1} & 0 \end{pmatrix} = \begin{pmatrix} -\frac{f_2}{f_1} & 0 \\ 0 & -\frac{f_1}{f_2} \end{pmatrix} \end{aligned} \tag{3.5}$$

where f_1 and f_2 are the focal lengths. This design, where we use two focusing lenses separated by $f_1 + f_2$, is called a Keplerian telescope[1]. At a distance f_1 after the first lens the beam reaches a 'Fourier plane' where the M_{11} and M_{22} elements go to 0, such that the angular distribution from the diffraction grating gets imaged into the spatial dimension. At this

---

[1] Notice that a Keplerian telescope flips the image, so care must be taken to orient the PFT in the correct direction.

point we can see the spectrum of the laser spread out in space. Then, in the second half of the transport, the Fourier transform gets inverted, and so we end up imaging the grating with an angular magnification of $-\frac{f_1}{f_2}$. Obviously this is a pleasingly symmetric setup, and it turns out that it is an ideal choice in the sense that it simply scales the beam (in both size and radius of curvature).

In a realistic setup we can change the magnification by varying the lens spacing. This breaks the symmetry of the arrangement, and as a result the final matrix looks like $\boldsymbol{M} = \left(\begin{smallmatrix} M_{11} & 0 \\ M_{21} & 1/M_{11} \end{smallmatrix}\right)$. The effective focal length $-1/M_{21}$ adds a radius of curvature to the beam which will need to be pre-compensated for by the matching optics. We can then multiply the matrix for a diffraction grating by this generic imaging matrix to get the ray transfer for the PFT system:

$$\boldsymbol{M} = \begin{pmatrix} M_{11} & 0 & 0 & 0 \\ M_{21} & \frac{1}{M_{11}} & 0 & \frac{\lambda}{\beta_{\mathrm{PFT}} c} \\ \frac{M_{11}}{\beta_{\mathrm{PFT}} c} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.6}$$

where we have used Eq. 3.1 to replace the grating parameters with the velocity $\beta_{\mathrm{PFT}}\, c$.

In order to use these rays to propagate an optical field, we have to solve the Huygens kernel [127]:

$$\begin{aligned} E\left(z_{out}, t_{out}\right) = \eta \int\int dz_{in} dt_{in} E\left(z_{in}, t_{in}\right) \times \ldots \\ \exp\left[-\frac{i\pi}{\lambda_0} \begin{pmatrix} z_{in} \\ z_{out} \\ t_{in} - t_{out} \end{pmatrix}^T \begin{pmatrix} \alpha & \beta & \gamma \\ \beta & \delta & \epsilon \\ \gamma & \epsilon & \xi \end{pmatrix} \begin{pmatrix} z_{in} \\ z_{out} \\ t_{in} - t_{out} \end{pmatrix}\right] \end{aligned} \tag{3.7}$$

where the six coefficients $(\alpha, \beta, \gamma, \delta, \epsilon, \xi)$ come from the matrix $\boldsymbol{M}$. The appearance of the combination $(t_{in} - t_{out})$ is a result of the absence of time-dependent optical elements, while the symmetry of the coefficient matrix is a general property of quadratic forms. Only six coefficients are needed to describe the optical transport, which suggests that the 4×4 matrix $\boldsymbol{M}$ only has six independent elements. In fact, seven of the elements in $\boldsymbol{M}$ are trivial (the output frequency is identical to input frequency and the input time only affects the output time), while the three nontrivial relations are given by Eq. 15 of Ref. [127].

In cases where there is imaging ($M_{12} = 0$), the Huygens kernel is becomes a delta function, so that Eq. 3.7 reduces to [127]:

$$\begin{aligned} E\left(z_{out}, t_{out}\right) \propto \int\int dz_{in} dt_{in}\, e^{(-I\pi M_{21}/\lambda_0 M_{11})} \times \dots \\ \delta(z_{in} - z_{out}/M_{11})\delta\left(t_{in} - \left(t_{out} - M_{31} z_{in}\right)\right) E\left(z_{in}, t_{in}\right). \end{aligned} \tag{3.8}$$

When $M_{31} = \frac{M_{11}}{\beta_{\mathrm{PFT}}\, c}$ is matched to give the pulse the same velocity as the electrons, the resulting field has the easy interpretation that the electrons travel through the transverse profile of the laser pulse (while remaining stationary in the laser pulse's time dimension). Outside of perfect imaging the full kernel can still be evaluated analytically provided the laser is Gaussian [127, 129] (and this is how we made Fig. 3.2).

3.1.2 Laser measurement

In order to actually align the optics and deliver a known pulse to the DLA, we need make careful measurements of the laser envelope (both its amplitude and phase). Rather then attempt to measure the full 6D space of the laser (as in [130]), we measure individual projections and then reconstruct the relevant portions of the envelope. This relies on the assumption that our laser starts out uncoupled between its longitudinal and transverse directions and that we are able to understand how the envelope propagates through the PFT optics. To check this, we have developed a cross-correlation technique to measure (and align) the PFT angle. Consequently, after making several independent laser measurements, we will be able to reconstruct a laser envelope that can be used as input for simulations.

The easiest parameter of the laser to measure is the pulse energy. Simply stick a meter in the beamline and get a number. Simplicity aside, the pulse energy is quite important because it normalizes all of the envelope measurements. We find that our laser can deliver a maximum of 180 µJ to the DLA (or about double that when bypassing the diffraction grating), and we want to be able to connect this to E_0.

To start, recall that the electric field is given by $\mathbf{E} = \Re\left(E_0 \mathcal{E} e^{i(ky-\omega t)}\right)\hat{z}$. The envelope is a function of the transverse coordinates (x, z) and a longitudinal coordinate moving at

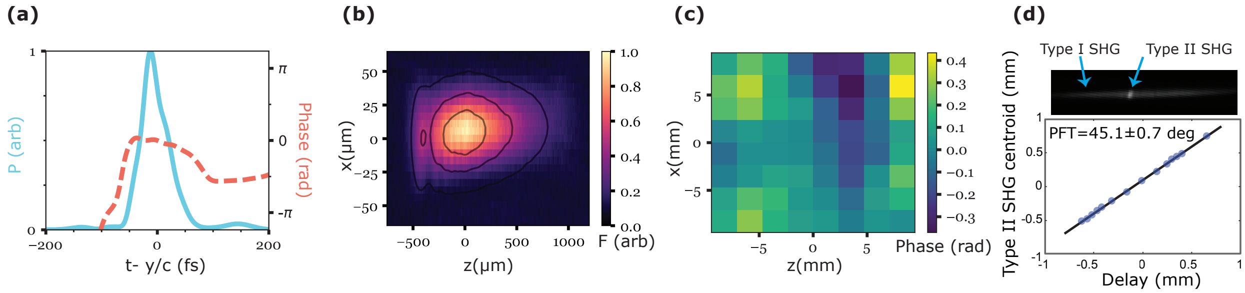

Figure 3.3: Laser measurements: (a) FROG gives the temporal amplitude and phase before the PFT setup; (b) A camera gives the fluence at the plane of the DLA; (c) A wavefront sensor gives the phase before the PFT setup(shown here without the quadratic component); and (d) A cross-correlation measurement of spatio-temporal couplings. At top is an individual correlation and at bottom is a measurement of the PFT angle.

the group velocity, which we will call τ. The intensity carried by this laser is given by its Poynting vector $\mathbf{S} = \frac{E_0^2}{Z}\left(\Re\left(\mathcal{E}(\tau,x,z)e^{i(ky-\omega t)}\right)\right)^2\hat{y}$, where $Z = \sqrt{\frac{\mu}{\epsilon}}$ is the electromagnetic impedance. By definition $\mathcal{E}$ is slowly varying, so when we time-average this intensity over the optical scale we find $\langle\mathbf{S}\rangle = \frac{E_0^2}{2Z}\mathcal{E}(\tau,x,z)\mathcal{E}^*(\tau,x,z)\hat{y}$. Finally we can connect the intensity to the often used quantities of fluence(F), power(P), and energy(En) by integrating out the relevant dimensions:

$$\begin{aligned}
\mathrm{F}(x,z) &= \frac{E_0^2}{2Z}\int_{-\infty}^{\infty} d\tau\mathcal{E}(\tau,x,z)\mathcal{E}^*(\tau,x,z)\\
\mathrm{P}(\tau) &= \frac{E_0^2}{2Z}\int_{-\infty}^{\infty} dxdz\mathcal{E}(\tau,x,z)\mathcal{E}^*(\tau,x,z) \qquad (3.9)\\
\mathrm{En} &= \frac{E_0^2}{2Z}\int_{-\infty}^{\infty} d\tau dxdz\mathcal{E}(\tau,x,z)\mathcal{E}^*(\tau,x,z)
\end{aligned}$$

Since our laser travels in and out of a dielectric structure we should be clear about where we are specifying the fields. In this work, we define E_0 to be the amplitude of the electric field at the plane of the DLA in the absence of non-linear pulse propagation (thus $\mathcal{E}(\tau,x,z)\mathcal{E}^*(\tau,x,z)$ has a maximum value of 1 in the absence of non-linear pulse propagation). When crossing a dielectric (at normal incidence) the field needs to be multiplied by the Fresnel factor $t = 2n_1/(n_1+n_2)$ [103]. The reflectance of about 3.5% per surface is accounted for when we list E_0.

Frog The first measurement we will consider is the FROG measurement of the longitudinal beam envelope . We use a commercial variant of FROG called GRENOUILLE [131] to record a spectrogram (time and frequency of the laser) and then we use a phase retrieval algorithm to reconstruct the amplitude and phase of the laser. The GRENOUILLE encodes the time coordinate by splitting a single beam with a bi-prism. The bi-prism causes the two beams to cross at a large angle such that the temporal separation of the two beams changes with transverse position. Thus cross-correlation using a bi-prism interferes pieces of the laser which came different transverse positions. As a consequence, GRENOUILLE is not suitable for measuring lasers with large $z - t$ correlations and so it is not possible to measure the Kerr phase or the 45^o PFT. Instead, we obtain accurate measurements by locating the GRENOUILLE before the diffraction grating, where the beam envelope is still separable.

A typical reconstruction, taken near full compression, is show in Fig. 3.3(a). The pulse consists of a short peak of FWHM 45 fs with satellite lobes caused by residual higher-order (3+) dispersion. To normalize this shape by the laser energy we can define a the full pulse width, τ_{full} as $\text{En} = \text{P}_{\text{max}}\tau_{\text{full}} = 60\,\text{fs}$. The effective length of an accelerator along this coordinate would be (see § 2.2) $|\int \mathcal{E}(\tau)d\tau| = 75\,\text{fs}$ (in the absence of nonlinear effects). We repeat this measurement for a variety of compressor settings (adding and removing dispersion) and confirm that the envelope changes as expected [109]. Furthermore, this shows that we can propagate the laser from the GRENOUILLE to the DLA, provided we know the dispersion (if we are perfectly imaging the diffraction grating, then the dispersion is ≈ 0 and the propagation is trivial).

Camera The second measurement, shown in Fig. 3.3(b), is a camera image of the laser spot size at the DLA (with no DLA present)[2]. The charge collected on the camera sensor is proportional to the laser fluence. We can fit a Gaussian to the projections of this laser along the x and z axes and we find that $w_z = 650\,\mu\text{m}$ and $w_x = 35\,\mu\text{m}$ at the $1/e^2$ intensity. As with the time domain, we can also define a full beam area as $\text{En} = \text{F}_{\text{max}}\text{A}$ and we find $\text{A} = 36000\,\mu\text{m}^2$ (which is close to the Gaussian approximation $\text{A} = \pi w_x w_z/2$). We have now

---

[2] This image is for the PFT beam. The flat beam, not shown, has a nearly identical shape.

calibrated E_0 via the intensity $S_{\max} = \mathrm{En}/\mathrm{A}\tau_{\mathrm{full}}$ and $E_0 = \sqrt{2ZS_{\max}}$.

Wavefront sensor While the FROG measurement gives us both amplitude and phase, the camera only gives us the amplitude of the envelope. To get the phase, and thus the effective length of a PFT accelerator, we use a wavefront sensor (which is analogous to a pepper pot in electron beam optics). The wavefront sensor is too large to be mounted in-situ, so we measure the wavefront immediately before the pulse front tilt grating. At this point the beam is round, but, as shown above, it is trivial to transform the envelope from there to the DLA. Thus we use the matching optics to set the radius of curvature at the wavefront sensor ($w = 1.3\,\mathrm{mm}$ and $R = -1.8\,\mathrm{m}$), since the downstream optics will convert this to $R = \infty$ at the DLA. In Fig. 3.3(c) we plot the wavefront with the quadratic term removed so that the image shows the wavefront error. The wavefront error is small (in the absence of Kerr dephasing), and the effective length of the laser is $|\int \mathcal{E}(z)dz| = 1.1\,\mathrm{mm}$ (in reality we will have a shorter length, truncated by the boundaries of the DLA structure).

PFT Measurement The pulse front tilt angle can be measured by cross-correlating the envelope of a known pulse with the tilted pulse. For this we can use the untilted laser pulse that we split off before the diffraction grating (see Fig. 3.2). Then, to do the cross correlation, we interfere these two pulses inside an α-beta barium borate (BBO) crystal and take a picture of the resulting second harmonic generation (SHG) as a function of the delay stage.

BBO actually allows two types of SHG (at 800 nm), and we will use both types for different purposes. In type I SHG two photons of ordinary polarization combine to produce one (doubled) photon of extraordinary polarization. In type II SHG photons of ordinary and extraordinary polarization combine to produce one (doubled) photon of extraordinary polarization. Type I is thus useful for learning about an individual laser pulse (flat or PFT), while type II is used to observe the overlap (or cross-correlation) between two pulses. We can switch between the two by rotating the crystal (or the polarization of the lasers) to match up with extraordinary and ordinary axes[3]. For intermediate settings we can see both types of SHG at the same time.

We can spatially resolve the SHG signal by imaging the BBO with a camera (after using a dichroic mirror to dump the 800 nm light), like we have shown in Fig. 3.2(d). The type I signal shows us the cylindrical shape of laser (the PFT and reference pulse are indistinguishable when overlapped), and the type II pulse shows us the position in space where they temporally overlap.

We use the type I signal to optimize the laser intensity. This works because the SHG intensity scales like $I(2\omega) \propto I(\omega)^2$. It makes an easy way to check that the telescope is imaging the diffraction grating. Often the flat and PFT arms won't be maximized at the same time, indicating that the dispersion isn't the same on both arms. When switching between configurations we can make up for the small difference by changing the input dispersion with a grating compressor.

We can then use the type II signal to measure the PFT angle by varying the cross-correlation delay arm. When we do this, the type II spot will move across the image in panel (d) at a rate proportional to the PFT velocity. To prove this, we use the Kostenbauder matrices to derive the cross-correlation intensity as:

$$\begin{aligned} \langle S \rangle &\propto \left| \chi^{(2)}_{z,y,z} \left(2\omega, \omega, \omega\right) E_x\left(\omega\right) E_y\left(\omega\right) \right|^2 \\ &\propto \exp\left(-4 \left(\frac{z}{M_{11} w_0} \right)^2 - \left(\frac{z + \beta_{\mathrm{PFT}} c \delta t}{\beta_{\mathrm{PFT}} c \tau_0} \right)^2 \right) \end{aligned} \tag{3.10}$$

where $\chi^{(2)}$ is the nonlinear susceptibility, δ_t is the cross correlation delay, $M_{11}w_0$ is the magnified spot size, τ_0 is the initial pulse length, and β_{PFT} is the pulse front tilt "velocity". The first term in the exponential limits the intensity to fit inside the transverse spot size of the pulse. The second term tells us that the centroid of the type II SHG traces out the line $(z + \beta_{\mathrm{PFT}} c \delta t)$, allowing us to determine the pulse front tilt between the two pulses. We use this measurement, shown in Fig. 3.3(d), to adjust the angular magnification of the PFT

---

[3] Note that this works well for a thin crystal, but for a thicker crystal we would also need to optimize θ, the angle between the laser propagation direction and the crystal axis (a.k.a. birefringent axes). The type I phase-matching is optimized at an angle $\theta = 29.2^o$ and type II at $\theta = 42.4^o$. Our crystal is cut such that type II is optimized at normal incidence, while type I would require hitting the BBO at an angle.

telescope and optimize the PFT angle. The largest systematic error in this measurement is the angle between the propagation direction of the two arms, which we align to less than 1^o by minimizing interference fringes in the type I SHG (caused by leakage through the polarizing beam-splitters).[4]

Envelope reconstruction Now we can reconstruct the beam envelope $\mathcal{E}(\tau, x, z)$ by stitching together the four measurements. Assuming that our measurement of the spatio-temporal coupling is accurate, then we know the Huygens integral is close to that in Eq. 3.8 with β_{PFT} from the cross-correlation measurement. Thus we can write the field envelope at the DLA as $\mathcal{E}(\tau, x, z) = \mathcal{E}(\tau)\mathcal{E}(x, z)$ where $\mathcal{E}(\tau)$ comes from the FROG trace with the substitution that $\tau \rightarrow \tau - z/\beta_{\mathrm{PFT}}$ (i.e. electrons don't move through this coordinate) and $\mathcal{E}(x, z) = \sqrt{\mathrm{F}(x, z)} * e^{i(\frac{k_0 x^2}{R_x} + \frac{k_0((x/M_x)^2 + (z/M_z)^2)}{W(x,z)})}$ gets the fluence from the camera and the wavefront errors (W) from the wavefront sensor. The magnifications M_x and M_y are measured to be roughly 1/35 and 1/2, while the additional radius of curvature R_x is not directly measured but can be calculated from the Kostenbauder matrices. Since individual electron trajectories do not cross the laser in the x direction this phase is generally ignored anyways.

3.2 Accelerator beamline

The role of the accelerator in DLA measurements is to provide a microscopic probe with good energy resolution. This places a huge premium on the transverse beam brightness, because both the DLA and the spectrometer require us to focus the beam to a small spot size. The best way to do this is to have a tiny source size, like the tip-emitters used in sub-relativistic measurements [19, 20, 21], but in our RF photoinjector the best we can do is focus the laser down to a small spot-size. There is another limit, however, because if the charge density at the cathode is too high then the pulse will spread out temporally and end up being too long to overlap with the laser.

---

[4] Over the 0.5(1mm) length of the DLA structures an error of 1^o will cause a 30(60) fs error in electron to laser time-of-arrival, leading to a 14(44)% reduction in the accelerating gradient.

Sub-relativistic experiments avoid battling space-charge by sending only a single electron at a time (see section 1.2). But Pegasus's duty cycle is too small (and the beam to jittery) to average over a billion exposures, and so we need enough charge to get a measurement in a single-shot (or close to it). Consequently, out experiments are often signal starved, and it is worthwhile to discuss how we can optimize the beamline.

The signal bottleneck entirely revolves around getting charge through the DLA within the temporal window set by the drive laser. So we will start by optimizing just the first half of the beamline (sketched in Fig. 3.1). After we have enough charge to work with, then we will be ready to discuss the spectroscopy part of the measurement. We will discuss how the spectrometer works and then show how it is calibrated.

3.2.1 Optimizing transmission

For simplicity we can imagine our DLAs to be hard-edged slits which poison any electron that touches their walls. The smallest slit is 400 nm tall by 1 mm long, and the largest is 800 nm by 1 mm. That means the DLA phase-space acceptance is a diamond (in $y - y'$) with an area as small as 0.08 nm-rad. This is 4 times smaller than the best rms (geometric) emittance achieved at Pegasus [82], and so no matter how hard we focus, some of the beam will always hit the DLA.

In the ultrafast electron microscope we also have a very strict requirement on the transverse beam brightness, but in that case we are free to stretch the electron beam to 10 ps and operate the RF photoinjector in the "cigar" regime. For the DLA, however, most of that charge would be useless since the accelerator is only turned on for 45 fs by the tilted pulse front and for 1 ps by the flat pulse.

If we want to increase the number of interacting electrons we will need to add optics outside the gun. One trick is to take advantage of the planar structure of the DLA. We can make a flat beam by a magnetizing the cathode and then using three skew quads [41]. The advantage of the flat beam is that its small emittance is independent of the laser spot-size on the cathode, and so we are free to make it large and collect lots of charge. Unfortunately

Q (fC)	ϵ_n (nm)	σ_r (μm)	$\sigma_{\beta r}$ (mrad)	α_{twiss}	σ_t (ps)	$\gamma m_e c^2$ (MeV)	$\sigma_\gamma m_e c^2$ (keV)
300	35	10	0.5	1*	1	6.5	0.5*

Table 3.1: Working point for DLA measurements. Numbers with an asterisk are known only through simulation.

the flat beam transform is rather complicated, and so it is still a work in progress that has not yet been used for DLA measurements.

Another way to increase the number of interacting electrons is to give up on some energy-resolution. If we run the linac off-crest (giving up about half the total beam energy), then we can add a chirp and a compress the beam at the DLA. In [82] we added a 50 keV correlated energy spread to the bunch and compressed the beam to less than 10 fs [82]. The problem with this scheme is that simulations show the time-of-arrival fluctuations would be a little over 1 ps, meaning that many shots would show no interaction at all[5].

Rather than suffer through many missed shots, we operate the linac close to on-crest. The resulting pulse length is about 1 ps, which is too long to fit inside the tilted pulse front. But, since it is longer than the time-of-arrival jitter, it makes spatio-temporal alignment much easier.

The rest of the beam parameters are listed in Table 3.1. They come from operating the gun in-between the pancake and cigar regimes and then focusing tightly into the DLA. With this beam we measure a transmission of 3% through the 800 nm×0.5 mm channel, 1.5% through the 800 nm×1 mm channel and 0.6% through the 400 nm×0.5 mm channel. We typically operate slightly over or under-focused in order to limit the effects of beam-jitter, but we still see shot-to-shot transmission fluctuations of 10-20%.

These considerations all apply to DLAs which are turned “off”. But when we power them with a laser there will be strong transverse forces affecting the bunch. It turns out that this isn’t a big concern, even for the 1 mm structure. To be sure, at some phases the

---

[5] The culprit is almost entirely fluctuations of the RF power. This causes the beam energy to jitter, and thus the time-of-arrival at the linac fluctuates. Only occasionally does the beam compress at the right spot and the right time.

transmission would be reduced by a factor of 5 or more, but a 6.5 MeV beam is stiff enough that the particles which start on crest, stay on crest (where there is no transverse kick), and so we can still measure the peak energy gain. To measure even longer structures, however, will require that we add focusing to keep the beam confined to the channel [38].

3.2.2 The spectrometer

To measure the spectrum we use a dipole magnet which bends the beam. Higher energy particles bend less, and so if we look at the beam after the bend then we can infer the particle energy. But this is *not* a perfect mapping, it still has contributions from other terms. For example, if two particles have the same energy, but one started to the left of the other, then it will end up to the left after the bend, and so we might mistakenly conclude that it had a different energy. This presents a problem for interpreting DLA results, because in addition to changing the particle energies, the DLA changes their trajectories. To convince ourselves that we will only notice a change in energy, we need a model for the spectrometer.

Let us consider a simple bend with gyroradius $\rho_0 = \frac{\gamma m_e \beta c}{qB} \approx 3.3[\mathrm{m}]\frac{\gamma m_e c^2[\mathrm{GeV}]}{B[\mathrm{T}]}$, followed by a drift of length L. The linear electron transport is given by:

$$\begin{pmatrix} x & x' & \eta \end{pmatrix}_f = \begin{pmatrix} 1 & L & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \rho_0 \sin(\theta) & \rho_0 \left(1 - \cos(\theta)\right) \\ -\frac{1}{\rho_0}\sin(\theta) & \cos(\theta) & \sin(\theta) \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ x' \\ \eta \end{pmatrix}_0 \tag{3.11}$$

where θ is the bend angle of the reference trajectory. We can write the position of an electron at our detector as $x_f = Mx_0 + lx'_0 + D\eta_0$, where the magnification M, effective length l, and dispersion D can be identified from Eq.3.11. To make $Mx_0 + lx'_0$ smaller than $D\eta_0$ we use a quadrupole doublet to focus the beam as tightly as possible on the spectrometer screen. Because the beam is at waist at the DLA (in order to fit electrons through the structure), this procedure will set the doublet to roughly image the electron beam from the DLA to the spectrometer (in the dispersion plane). This arrangement has the benefit of making $l \approx 0$ such that we are insensitive to any changes in x'_0 that may be applied by transverse kicks from the DLA. Consequently, we can conclude that any changes in x_f between laser-on and laser-off shots are caused by (de)acceleration in the DLA.

From the spectrometer we don't get x_f directly, but rather we record a 2-D intensity distribution $f(x, y)$ of the light from a screen which scintillates when struck by electrons. We can then integrate $f(x) = \int f(x, y)dy$ across the non-dispersive direction[6] to find what we often, reductively, refer to as the electron spectrum. In fact, what we measure still has some influence from the electron spatial distribution as well as some blurring caused by the fluorescent screen and by the camera-lens system imaging that screen. Thus, even with the laser off, we typically measure an "energy-spread" of $> 3\,\mathrm{keV}$ (rms) even though simulations suggest the beam should have an energy-spread of less than 1 keV (rms). We refer to the 3 keV as the resolution of the spectrometer and the laser-off distribution as its point-spread function.

We can improve the resolution of the spectrometer by increasing the dispersion. The easiest way to do this is to increase the drift L after the dipole kick,[7]. In fact, we use two different drifts $L1$ and $L2$ (as shown in Fig. 3.4) to provide better resolution and field-of-view respectively. We can also improve the resolution by using a thin, single-crystal, screen (made of cerium doped yttrium aluminium garnet, known as YaG:Ce) however we can collect more light if we use thicker screens or poly-crystalline screens. The best resolution we achieved for DLA was 3 keV on screen L_2 using a 100 µm thick YaG:Ce screen.

3.2.3 Dipole calibration

We use a round-pole dipole like the one pictured in figure 3.4. The dipole consists of a bar shaped current source coupled to a cylindrical yoke. The dispersion (D) is largely determined by the fact that the downstream beam-pipe is rotated by an angle of 45°, but some uncertainty is introduced by the extent of the fringe field. The accuracy of the energy

---

[6] By scanning the dipole current we can see the dispersion direction (which may be slightly rotated relative to the camera coordinate system). And by scanning the quadrupole currents we can see if the quadrupoles are rotated (and thus introducing a $y - x'$ correlation) and find the shear to be removed.

[7] Note that there is a limit to this technique, since eventually we are unable to focus the beam tightly onto the spectrometer screen.

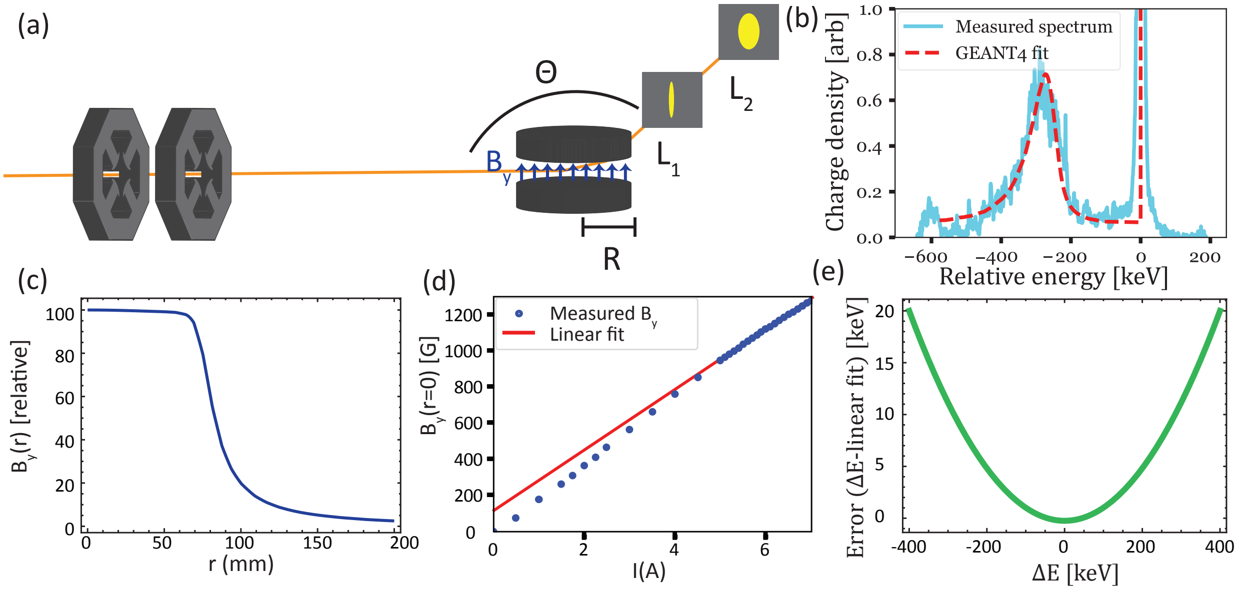

Figure 3.4: Spectrometer geometry and calibration data. (a) Cartoon of the spectrometer. (b) Calibration of a spectrometer via the known energy loss spectrum of a fused silica wafer. (c) Simulated map of the dipole field. (d) $B(I)$ curve showing saturation of the field in the spectrometer. (e) Error due to nonlinear dispersion.

gain measurement is proportional to the error in the dispersion, so we reduce this systematic uncertainty by calibrating the dipole in three separate ways: by modeling the magnetic field, by using a beam based measurement, and by measuring the energy loss of electrons traveling through the DLA. The measurements are summarized in Table 3.2 and described in some detail below.

Field based calibration A hard-edged dipole is characterized by its pole-face rotation ϕ and its bending angle θ. Round-pole dipoles are designed to have zero pole-face rotation (i.e. the reference electron trajectory enters and exits the magnet perpendicular to the field boundary), but if the beam is offset from the center of the beam pipe then we need to include the edge fields [132]. For our case, the resulting correction is negligible and the dispersion of a dipole followed by a drift of length L can be calculated from Eq. 3.11 as:

$$D = L\left(\sin(\theta)\right) + \rho_0\left(1 - \cos(\theta)\right) \tag{3.12}$$

where $\theta = \arctan(2R/\rho_0) \approx 45^\circ$ is the bend angle in a dipole with effective yoke radius R. To account for the fringe field of the dipole, we use an effective pole radius defined as $R = \int_{-\infty}^{\infty} B/B_{\max}\, ds$. We use a field profile simulated in Maxwell (Fig. 3.4(c)) to calculate that $R = 9.3\,\mathrm{cm}$. Knowing R, we can measure the drift distances L_1 and L_2, however we have only achieved approximately 1 cm precision due to uncertainty about the screen location relative to the vacuum cubes. There is an additional uncertainty in the final dispersion on the order of 1 mm due to beam misalignment (which changes ϕ, R,θ, and ρ_0).

Beam based calibration A dipole magnet can also be calibrated by using a beam of known energy[8]. The idea is to measure the centroid displacement on the spectrometer screen as a function of the magnetic field in the dipole. Since the magnetic force is proportional to p_z we know that changing the field strength is equivalent to changing the beam energy and (in the linear approximation) we can calculate the shift seen on the spectrometer screen as: $\delta x = D\frac{\delta p_z}{p_z} = D\frac{\partial_I B|_{I0}}{B(I_0)}\delta I$. Given a known curve $B(I)$ which centers the beam at current I_0 we can calculate the dispersion. The accuracy of the dispersion based on this technique is limited to ±5 to 10 mm by hysteresis in the dipole magnet[9].

Energy loss calibration Finally, we can calibrate the dispersion by measuring the distance between two beams of known energy. This is most easily done by measuring a beam which is partially clipped by the fused silica wafer that makes up the DLA. The dimensions of the wafer are known very precisely and so we can simulate the energy loss spectra in GEANT4 [133] and then compare that spectrum to measurements. We get a calibration by fitting the dispersion such that the simulated spectrum matches the measured spectra (see 3.4(b)).

Note that at Pegasus energies there is significant radiative energy loss which has a long tail (i.e. including MeV photons), such that the mean energy loss (given by the total stop-

---

[8] We can measure the beam energy using an upstream spectrometer which has been calibrated via field maps and against the diffraction angle of known crystals.

[9] Hysteresis would be negligible if $\partial_I B$ were constant, but since our magnet saturates (see Fig. 3.4) it becomes important.

ping power from E*STAR [134]) is much larger than the peak of the energy loss spectrum. Moreover, the peak energy loss is more-or-less independent of beam energy above a few MeV, so that we aren't very sensitive to the calibration of the upstream spectrometer. The main inaccuracy of this method is actually due to non-linear beam dynamics in the spectrometer (and chromatic aberration before it).

Nonlinear beam dynamics Given a field map we can calculate particle trajectories as a function of energy. Such a calculation directly yields a dipole calibration curve, which can be fit to find the linear dispersion. By examining the residuals of the fit we can see a nonlinear contribution to the dispersion (Fig. 3.4(e)). In principle we should account for this, however for the energies of interest ($\lessapprox$ 400 keV) the correction is less than 3%, which is negligible compared to the uncertainty of our DLA measurements.

Summary of dipole calibrations Table 3.2 summarizes the dipole calibration. The spread of measured dispersions indicates a roughly 20% systematic uncertainty in the spectrometer calibration. However, we believe the energy loss method to be more accurate than the others. The field-based measurement requires precise field maps and the current scan requires precise knowledge of the $B(I)$, but the energy loss method only requires us to calculate the energy loss due to fused silica. Thus when reporting results we use the calibration from the energy loss method, with the understanding that there may be a systematic error in all reported energy changes (equivalent to a change in the DLA diffraction efficiency, (c_1, c_2)).

3.3 Electron-laser alignment

The DLA interaction takes place in a spatio-temporal volume of 1 µm × 50 µm × 1 mm × 45 fs. To make the laser and electron beams overlap inside this volume we built a sample holder (Fig. 3.5) which allows us apply a series of in-situ alignment techniques. The whole assembly is located on a four axis stage which allows us to move the DLA in x, y and to rotate the DLA about the x, y axes.

Method	Screen	$\gamma_0 m_e c^2$ (MeV)	D (cm)	Source of systematic error
Field based	1		13.6	Measurement of L_1
Current scan	1	6.5	10.4	Hysteresis
Energy loss	1	6.5	12.6	Nonlinear dispersion
Field based	2		27.6	Measurement of L_2
Energy loss	2	8	31.6	Nonlinear dispersion

Table 3.2: Summary of the dipole calibration. At screen 2 the dispersion is large enough that the main and scattered beams are not visible at the same time for 6.5 MeV beam energy, and so the calibration is done at 8 MeV beam energy.

We start the alignment by positioning the DLA according to a He:Ne laser which propagates in line with the electron beam. Then we try to move and twist the sample holder until the electron beam can propagate through the alignment channels (Fig. 3.5(l)). To help orient ourselves during this process, we start with an over-focused electron beam so that we can see a shadow of the DLA. When the sample holder has the correct y' angle (rotation about the x axis), then we will see a bright spot in the middle of the DLA shadow corresponding to the alignment channel. We can then focus the beam down into the channel.

In order to find the tightest focal spot we can look at the brightness of an in-situ YaG:Ce screen (Fig. 3.5(g)), but the spatial resolution of the screen is too poor to measure the spot size. For that we use a different YaG:Ce screen located about 20 cm upstream. It is imaged by an in-vacuum microscope objective and has a few micron resolution. We show such measurements in Fig. 3.6(a), in which we can see that the shot-to-shot jitter exceeds the rms beam size. Since this jitter would cause nearly 100% charge fluctuations through a DLA, we actually want to de-focus the beam a little bit. We can make a quantitative check of the spot-size at the DLA by using a piece of tantalum (3.5(h) to do a knife-edge scan as shown in Fig. 3.6(b). [10].

---

[10] Normally we think of knife edge scan as the convolution of a Gaussian beam distribution distribution with a step function and so we fit the scan with an error function. In this case there is also significant centroid jitter and so the knife-edge will systematically overestimate the spot-size.

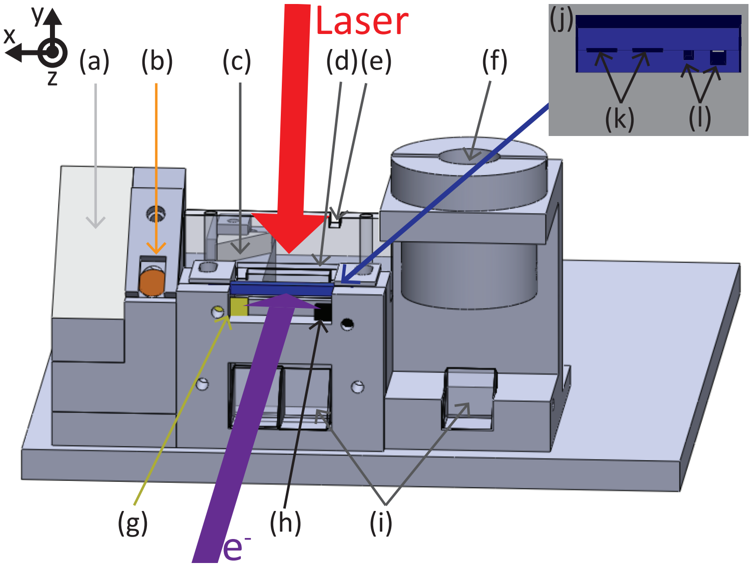

Figure 3.5: DLA sample holder. (a) Scintillating screen at 45° for simultaneous imaging of the laser and electron beam (b) TEM grid at 45° for time-of-arrival diagnostic. (c) Mirror for the light from g, a yag screen. (d) Collimator to reduce the number of scattered electrons visible on the spectrometer. (e) Alignment hole in the collimator. (f) BBO mount for measurement of laser compression. (g) YAG:Ce screen to image electron beam. (h) Tantalum for knife-edge measurements of the electron beam size. (i) Glass prisms which serve as laser beam-samplers. (j)The DLA. (k) The acceleration channels. (l) The alignment channels.

Next we want to find transmission of the electron beam through the DLA accelerating channels. Since we have a very exact, lithographic, map of the DLA, we know how to translate the sample holder from the alignment channels to the accelerating channels (from (g) to (e) in Fig. 3.6). But the angular acceptance of the DLA is much smaller than that of the alignment channel and so it can be a long search to find transmission through the channel. With practice, it is possible to do the search by hand, but the first time we set up a DLA we often need more guidance than the alignment channel can give us. The main trick we can use, is that the accelerating channels are located directly in the middle of an *exactly* 1 mm thick DLA. So if we translate the sample stage up and down then we can map out the top and bottom of the DLA and then return the to the middle. If the DLA is tilted (relative to the electron beam), then we will measure the height to be slightly more than 1 mm, since

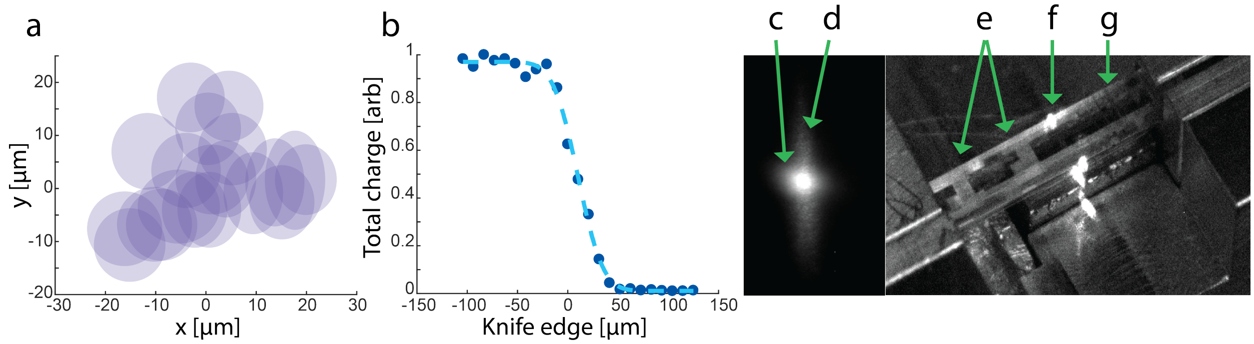

Figure 3.6: DLA spatial alignment: (a) Beam jitter and spot-size near the DLA (rms ellipses); (b) Knife edge scan at the DLA ($\sigma_y = 20\,\mu\text{m}$); (c) Electron and (d) laser spot size seen simultaneously on the scintillator in Fig. 3.5 a; (e) DLA channels; (f) laser spot reflecting off of the aluminum mount; (g) alignment channels.

the edges of the DLA will be elongated (this will also make the beam look bigger during knife-edge scans using the top edge of the DLA). Using these two cues we can align both y and y' (this is enough guidance to align a DLA whose alignment channels are missing).

Eventually the alignment is good enough that we can move the stage back and forth to take a "scanning electron microscope" image of the DLA slits. We carefully record a map of transmission through the DLA and match it up to the known pattern of the channels. This fixes the DLA in space (within a few µm and a few mrad), and so we are ready to go back into the box and align the laser to the DLA.

When aligning the laser, we want to set the angle of incidence based on the phase-matching condition in Eq. 2.2. It is easy to get an angle of zero by aligning the back-reflection from the DLA such that it propagates far upstream. In order to get a non-zero angle, we use the rotation stage to tilt the DLA before aligning the back-reflection. The rotation stage is fairly precise, so most of the uncertainty will be from aligning the back-reflection (typically a few mrad). For the PFT laser we have to be extra careful, since the angle of incidence (and thus the angle of back-reflection) depends on the wavelength. To simplify things, we stick a pinhole in the middle of the Keplerian telescope, where the laser has been maximally dispersed by the diffraction grating. This picks out a single wavelength, which we measure with a spectrometer, and then we can align the back-reflection accordingly.

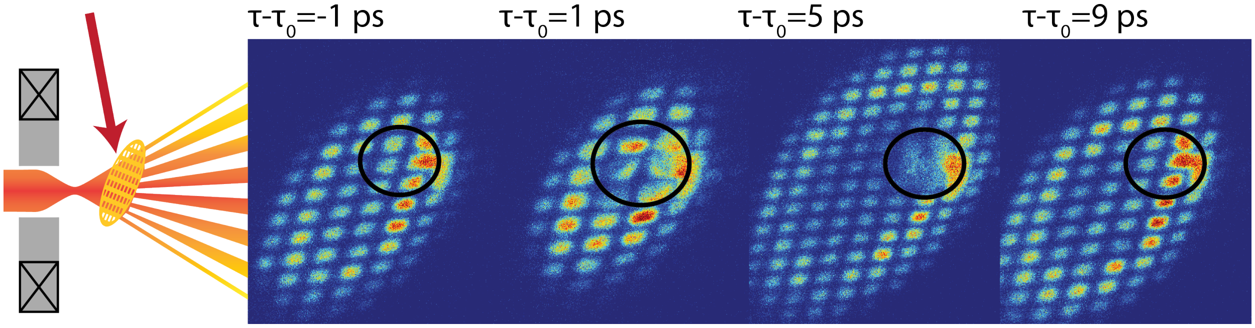

Figure 3.7: TEM grid timing measurement. At left is a cartoon of the electron beam forming a point-projection image of the grid. When the grid is hit by a laser, it creates a cloud of electrons which distort the image and serve as a timing diagnostic. The current location of the laser is shown by circle. A burn spot from a previous scan is visible at the right edge of the circle.

At the same time, we need to spatially overlap the laser with the electrons on the DLA. Since we have recorded the positions of the sample holder which correspond to various accelerating channels in the DLA, then we steer the laser spot (Fig. 3.6(f)) to overlap with the appropriate grating (Fig. 3.6(g)). Once the position and angle of the laser have been coarsely aligned by hand, then we can use piezo-actuated mirrors to make fine adjustments. To check the overlap we use a DRZ scintillating screen, mounted at $45\,^\circ$ (Fig. 3.5(a)), to view the electron beam and the laser at the same time (as shown in Fig. 3.6). To summarize the process up to this point: we align the DLA to the electron beam, then the drive laser to the DLA, and finally we make micron-scale adjustments optimize the electron-laser overlap.

Next we can synchronize time-of-arrival to the 100 ps level by using a fast photodiode to measure the delay between the drive laser and uv (cathode) laser. To get down to 1 ps timing we use the TEM grid based technique [135] show in Fig. 3.7. The electron beam is focused in front of the TEM grid such that it forms a shadow ("point-projection image") of the grid downstream. The drive-laser is then tightly focused to approximately $5\,\mathrm{J/cm^2}$ (see Fig. 3.2) in order to ionize the copper grid via multi-photon-photoemisson. If the electron beam arrives shortly after the ionizing laser, then it will pass through a cloud of electrons which will distort the grid image. The effect turns on in less than 1 ps and, in our example it

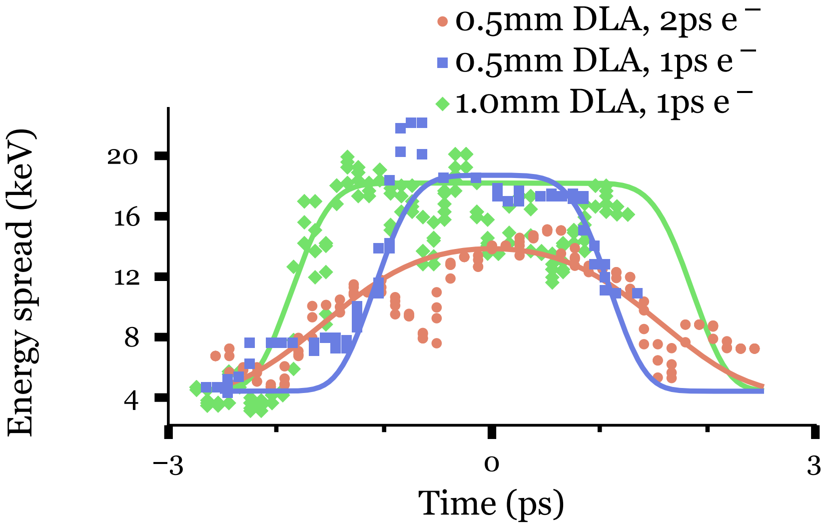

Figure 3.8: Timing scans. The structure length (0.5-1 mm) determines the plateau region while the electron beam length (1-2 ps) determines the roll-off.

lasts for about 7 ps. At higher fluences (e.g. 8 J/cm^2) the effect can last for up to 15 ps, but it will quickly melt the grid-points. This can actually be a useful way to check spatial-overlap, so we often purposefully burn a single grid point before we start the timing scan (we can see such a melted grid-point immediately to the right of the current laser location).

If the TEM grid used for timing synchronization is not quite in the same plane as the DLA, then there will be a few ps error in the time-of-arrival. Thus we typically search for acceleration by scanning the time-of-arrival (and also the rotation about the x-axis) until we find an acceleration signal. At that point, we can optimize alignment to maximize the DLA interaction.

Once optimized, the time-of-arrival scan can acts as a cross-correlation timing measurement [136]. We show a version of this in Fig. 3.8 by plotting the energy spread as function of time-of-arrival. Since this example uses flat-pulse illumination we expect that the length of the DLA structure (0.5 or 1 mm) controls the time window for the interaction. If the beam is entirely inside the structure when the laser arrives then we will see a big interaction, while if it is outside we will see none. For places in-between we will see a cross-correlation of the time-window with the current distribution.

To help see this we plot error function fits to the data. Despite significant jitter we can distinguish qualitative differences between the three curves: the 1 mm long DLA has a longer interaction window, while the longer (2 ps) electron beam has a more rounded shape. These scans help us establish that the DLA is aligned and operational. [11]

3.4 DLA measurements

In this section I will present our measurements of the DLA interaction. Having already introduced the idea of "DLA spectroscopy", it suffices to reiterate that each interaction yields an electron spectra from which we will extract a single number, ΔE, which represents the maximum observed energy gain (or energy loss). We can then study the response of ΔE to changes in the drive laser in order to measure properties of the DLA and test our model of the interaction.

Historically, the goal of these types of spectroscopic measurements have been to determine the peak accelerating gradient and thus demonstrate DLA's place in the advanced accelerator community. This typically means measuring ΔE as function of the laser intensity up until the damage threshold of accelerator, and then dividing by an effective length to estimate the gradient. One of the main contributions of this thesis is to realize the importance of measuring (and/or simulating) the effective length. We find that dephasing plays an active role in determining the effective length (and thus gradient) and is particularly important at high intensities where the Kerr phase modulation can grow rapidly [12].

We can explore the relationship between the gradient and effective length by studying ΔE as a function of both the drive laser intensity and the group delay dispersion. For our experiments with a flat-pulse these parameters are connected through the Kerr effect, such that there is an optimal combination which compensates for the dephasing. However, for the

---

[11] Note that the gap size doesn't effect the timing. But for completeness we note that all 3 curves in this example are from 400 nm gap DLAs.

[12] This was predicted at least as early as 2014 [51], but went unnoticed by our collaboration when planning the experiments.

PFT-laser, they are mostly unconnected, and instead we can learn about the PFT alignment. The main interest with the PFT-laser is extending the interaction to longer distances. And with longer distances we can can easily manipulate the phase-matching condition to study off-resonant DLA interactions and see some simple accelerator physics for the first time in DLA.

At the end of this chapter we will summarize the measurements with a series of records set during these DLA experiments. The records apply only to relativistic DLA experiments, but they are important to the extent that they represent progress in our engineering of high-gradient DLA interactions. This progress had been accompanied and made possible by the physical understanding emphasized in the proceeding chapter.

3.4.1 Analyzing DLA spectra

When we measure a spectrum, what we get is a distribution $f(\Delta E)$ which represents the whole beam. We can't break the spectrum into single-electrons and track what happened to each one, and thus we can't directly apply the acceleration equations from section 2.2. Instead we are going to have consider what happens to a beam distribution when it passes through the DLA.

To start, we need a model of the beam distribution before the DLA. In 1D this is fairly easy: the electron beam is much, much, longer than a laser wavelength and the energy spread is too small to cause dephasing, so we can just treat the beam as uniformly distributed in phase and neglect the energy spread: $f_0(\psi, \eta) = 1/2\pi\delta(\eta - 0)$.

Next we want to see what happens when the DLA kicks this beam. In the kinematic approximation we know that the DLA will add a sinusoidal voltage $\Delta E \approx q_e E_m L_{\text{eff}} \sin(\psi_0) \equiv q_E V \sin(\psi_0)$. Thus any particle which ended up at ΔE came from $\psi_0 = \arcsin(\Delta E/q_e V)$. This tells us how to calculate $F_1(\Delta E)$, the cumulative distribution function of the energy gain, since we know that the probability of finding an energy change less than $|\Delta E|$ is the same as probability of finding a phase less than $|\psi_0|$. The probability of finding a phase less than $|\psi_0|$ is just $F(\psi_0) = 2\int_0^{\psi_0} 1/2\pi$. Then, substituting in $\psi_0 = \arcsin(\Delta E/q_e V)$ and taking a derivative tells us that:

$$f_1(\Delta E) = \partial_{\Delta E}\left(2\frac{1}{2\pi}\arcsin\left(\frac{\Delta E}{q_e V}\right)\right) = \frac{1}{\pi}\frac{1}{\sqrt{q_e^2 V^2 - \Delta E^2}} \tag{3.13}$$

This is sometimes called the "arcsin" distribution, and it is strongly peaked at $\pm q_e V$ since there are more electrons at the turning points than elsewhere (see Fig. 3.9)(a). This feature is so salient, that inside the ACHIP collaboration it has been nicknamed the "double-horn" distribution.

That is not the end of the story, however, because when we measure a spectrum we rarely get to see f_1 directly. What we do see depends on whether we are looking at acceleration from a flat or tilted pulse front, so we will now address each separately.

Flat pulse With a flat pulse, the maximum energy gain ($\Delta E = q_E V$) is small enough that we have to worry about the 3 keV resolution of our spectrometer. Essentially, instead of measuring $f_2(x) = D(\Delta E/\gamma m_e c^2)$ (where D is the dispersion of the spectrometer and x is the location on the spectrometer screen), we are going to sample a distribution $f_2(x) = D(\Delta E/\gamma m_e c^2) + g(x)$, where g is whatever limits the spectrometer resolution. It could be the intrinsic energy spread, the spot size or divergence, the camera point-spread function, or anything else. Regardless of what causes it, we know that it is roughly 3 keV wide and that it is going to complicate the measurement.

In the language of probability theory, what we have is the sum of two random variables. One is proportional to the $f_1(\Delta E)$ we calculated earlier, and the other is spectrometer point-spread-function g. Assuming f_1 and g are independent, then what we see on the spectrometer will be the convolution $f_2(x) = f_1(D(\delta E/\gamma m_e c^2)) \circledast g(x)$. This process is shown in Fig. 3.9, where the arcsin distribution is diluted by a Gaussian spread g.

If we can measure g, for example by turning the laser off, then we can numerically invert (deconvolve) to get f_1. The most straightforward way to do this is via the Fourier convolution theorem: $f_1 = \mathcal{F}^{-1}\left(\mathcal{F}(f_2)/\mathcal{F}(g)\right)$; however, we would quickly find that (due to noise) we are at risk of dividing by zero. A simple solution to this problem is known as Wiener deconvolution, for which we calculate: $f_1 = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(f_2)\mathcal{F}(g)^*}{\mathcal{F}(g)\mathcal{F}(g)^* + \lambda^2}\right)$ where λ is a parameter which

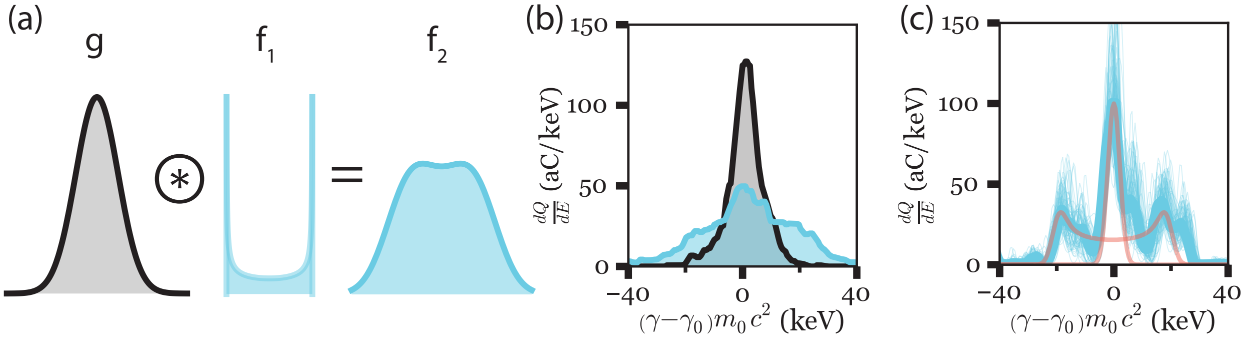

Figure 3.9: Spectrum from a 7 GV/m flat-pulse laser (400 nm gap DLA). (a) Cartoon of the convolution process. (b) Single-shot spectrum. Laser-on is blue and laser-off is black. (c) Deconvolution of the same on-shot with many off-shots. The orange lines are cartoon distributions identifying the unaccelerated population and the arcsin-like accelerated population.

smooths out the deconvolution. If λ is the Fourier transform of an additive noise term, then the Wiener deconvolution is optimal. Our case is somewhat more complicated, because we do not know the exact point-spread function. Consequently, we use a blind Richardson-Lucy algorithm [137] to reconstruct the electron energy gain spectrum. This is an iterative technique which alternately updates our knowledge of the psf and the energy gain spectrum. However, we find that the final result (the peak of the energy gain spectrum) is largely independent of precise details of the algorithm (and indeed that the Wiener algorithm is sufficient provided λ is large enough to avoid division by zero).

We show what this process looks like in Fig. 3.9. In panel (b) we show a sample laser-off spectrum (g) in black and a laser-on spectrum (f_2) in blue. Even though we are using the screen at L_2 to add extra dispersion (see Table 3.2), the laser-off term clearly makes up a large part of the measured spectrum. Thus we do the deconvolution, which we show as a blue curve in panel (c). We can decompose the result into a peak at zero, which corresponds to unaccelerated electrons, and an arcsin term which comes from the DLA.

The main source of uncertainty in this method is determining the correct "laser-off" distribution g to use in the deconvolution. In Fig. 3.9 we illustrate this by deconvolving the same on-shot with many independent off-shots (the individual deconvolutions are shown

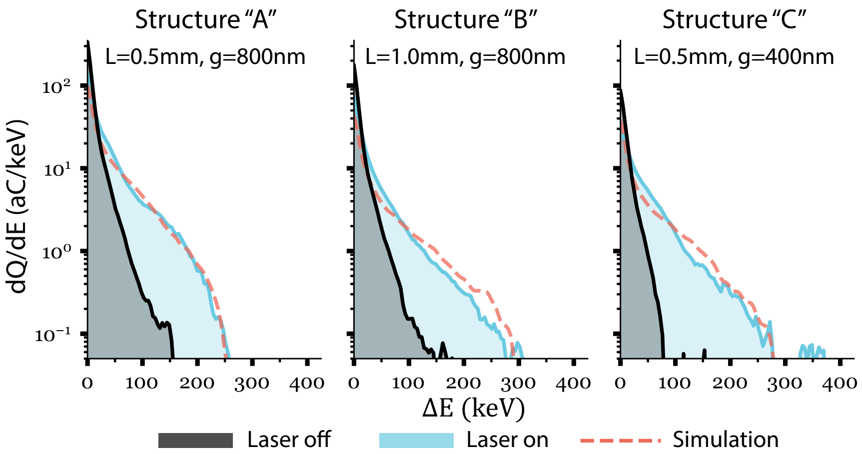

Figure 3.10: Highest recorded energy gain for the three different accelerators of Table 3.3. The electron energy spectra are shown with the laser off (black) and laser on (light blue). Particle tracking simulations used to fit the diffraction efficiency are also shown (red dashed). Each spectrum is an average of 50 shots.

with thin blue lines and their average is shown with a thick line). The variance happens because of shot-to-shot fluctuations in the the camera noise and in the electron beam, so it can't be easily corrected. Nonetheless, the width of the spectrum is consistent enough to determine the energy gain.

PFT pulse The spectrum from a tilted pulse front does not look at all the same as the untilted spectra. First of all, the energy spreads are so large that the resolution of the spectrometer is negligible (a delta-function), so we can just ignore it. And second, we now have so many un-accelerated electrons that we need to put the spectrum on a semi-logarithmic plot in order to see the tails.

What this looks like is shown in Fig. 3.10 for a few different DLA structures. We have only shown the energy-gain side of the spectra because the energy-loss has a higher noise threshold (caused by electrons which were scattered inside the fused silica, see Fig. 3.4). There are so few electrons making it out to the tail of these pulses that we are often trying to detect just 1 or 2 per shot per keV. In order to improve the signal-to-noise ratio ,we have to average roughly 50 shots together (by lining up their $\Delta E = 0$ peaks). To avoid biasing the results we average all shots which show acceleration (due to jitter some “on” shots miss entirely). Then we compile 50 laser off shots in order to have a fair baseline for comparison.

The reason we have so few accelerated electrons to work with is that the electron beam is much longer than the 45 fs slice of the tilted pulse front laser. Thus most of the electrons don’t get accelerated at all. Even worse, a lot of them get accelerated less than the maximum amount, because the are not located at the peak intensity. This completely washes out the arcsin distribution. In fact, the distribution we see is dominated by the shape of the laser rather than by an arcsin shape. We can’t analytically invert the shape of the laser, so instead we run simulations. In many ways the simulations are actually preferable, because they can use measured parameters, include the Kerr effect, and go beyond the kinematic approximation.

3.4.2 As a function of electric field

Now that we know what to make of the spectra, we area ready to just grab the maximum ΔE and go. The first thing we will look at is how the energy gain scales with the input electric field (by changing the laser energy). Naively we would expect the energy gain to be proportional to E_0, because the DLA is a linear dielectric. And, indeed, a factor of $E_m = E_0|c_{1,m} + c_{2,m}|$ comes out of the integral in Eq. 2.7, but there remains a factor of L_{eff} to contend with.

We will see that L_{eff} is not at all constant for the flat pulse front, because the Kerr effect dominates the interaction and causes the energy gain to saturate. This makes measuring the accelerating gradient more complicated since we will need to use NLSE simulations. We also use NLSE simulations for the tilted pulse front, but the dephasing does not cause as much saturation because the edges of the DLA truncate the acceleration before the electrons reach the part of the laser with a rapidly changing phase.

Flat pulse In Fig. 3.11 we show several raw electron spectra and then the extracted electron energy gain as a function of E_0 in a 400 nm gap DLA structure. The energy gain is

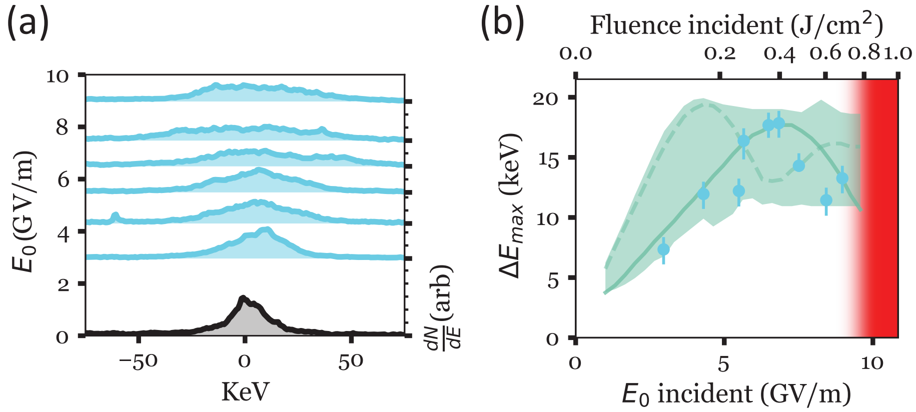

Figure 3.11: Saturation of flat pulse acceleration. (a) Single-shot electron spectrum for increasing laser power (blue) as compared to laser off (black). (b) The peak energy gain as a function of E_0 (data in blue). The error bars indicate a 70% confidence interval based on measured variance. The (green) lines are simulations described in the text. The peak accelerating field before damage (red) is 1.8 GV/m.

so small because the laser is only on for an effective length of 75 fs (before the Kerr effect further shortens it). Thus we have to use the deconvolution technique to analyze the spectra.

To make a robust measurement of the energy gain we define ΔE_{max} as the interval which contains 75% of the charge. To account for jitter we actually take many measurements at each E_0 and what we report is the maximum value. The reported uncertainties are the 70% confidence interval on the order-statistic for a truncated Gaussian (the distribution of energy gains is truncated on the high end).

What we find is that from 0 to roughly 6 GV/m, the energy gain increases linearly, but at higher intensities the energy gain saturates and even shrinks. This is explained by the Kerr effect: by 6 GV/m, the self phase modulation $\Delta\Phi = n_2 I k_0 d > \pi$ such that the electrons which are accelerated by the peak of pulse will be decelerated by the tails. This shortens the effective length of the accelerator, and so it makes it harder to determine the accelerating gradient.

Below 6 GV/m, where the behavior is linear, we could still expect to make an accurate measurement of the diffraction efficiency. But we are stymied by another complication:

namely, that our $\Delta E_{\max}$ statistic averages over all the electrons in the beam, including those that weren't accelerated at all. Fortunately, the nonlinear effect helps us: all of the electrons will eventually saturate at the same level, and so we can use that level to find the diffraction efficiency. The way this works is shown via the NLSE simulations in Fig. 3.11. The dashed green line shows the energy gain of an on-axis particle, while the solid green line shows a synthetic measurement on a realistic beam (the beam has $\sigma_x = 10\,\mu\text{m}$ and a misalignment of $x_0 = 10\,\mu\text{m}$, $\theta_y = 5\,\text{mrad}$). Compared to an on-axis particle, the saturation of the beam is delayed, but the saturation level is largely unaffected. Now we don't know the actual electron beam parameters very well, but this doesn't matter. The range of possibilities is shaded in green, and all of them saturate at the same point. Thus we can measure that $|c_1 + c_2|$=0.2 ± 0.04, in the middle of the range (0.18-0.23) of efficiencies predicted by FDTD simulations as a function of tooth-offset (see Fig. 2.10).

PFT pulse With the PFT-laser the electrons cross the "transverse" part of the laser envelope ($w_z = 550\,\mu\text{m}$) which would have an effective length of 1.1 mm in a suitably long DLA structure. But our structures have lengths of 0.5 and 1 mm such that the tail of the laser will be cut off by the edge of the DLA. This means that the Kerr effect is much less noticeable since the tails have the largest phase change (relative to the peak). It also means that the energy gain will be a few 100 keV, which is so much larger than the spectrometer resolution that the deconvolution is negligible. But the PFT only accelerates a 45 fs slice of the electron beam and so we cannot use the 75% criteria from the flat-pulse analysis (since 95% of the beam is essentially unaccelerated). Instead, we define $\Delta E_{\max}$ as the largest energy for which the spectrum remains 3σ above the detector noise.

With this definition in hand, we can plot for $\Delta E_{\max}$ vs E_0 as we did for the flat case (Fig. 3.12). As mentioned above, the Kerr effect has less importance here so we can easily model the data with a linear fit. We show the result for 800 nm gap channels with lengths of 0.5 and 1 mm. The slope of the 1 mm channel is almost exactly twice that of the 0.5 mm channel, as would be expected for flattop lasers. But our laser is actually Gaussian, and to explain this result we have to point out that the center of the laser is offset as shown in

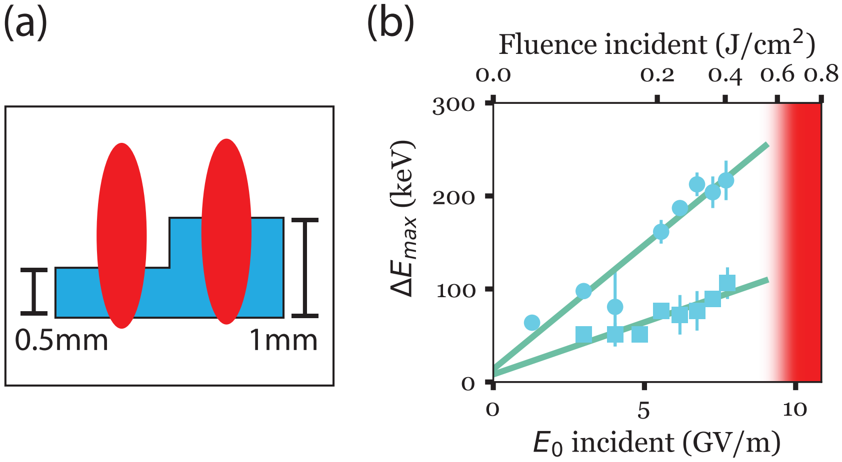

Figure 3.12: Energy gain vs E_0 across the 0.5 and 1 mm channels using the PFT laser. (a) Cartoon showing the laser (mis)alignment relative to two channels. (b) The blue circles show the energy gain across the 1 mm channel and the squares show the energy gain across the 0.5 mm channel. The datum represent the median of the 10 highest shots recorded (out of 75 shots per electric field), while the error bars show the spread. The green lines are linear fits to the data.

Fig. 3.12(a). Indeed after fixing the laser offset in z (and actually overcompensating to the other size of the midline) we were able to achieve even larger energy gains (see Fig. 3.10). In any case, the fact that the two channels are different at all demonstrates that the PFT velocity is well matched to the beam velocity.

3.4.3 As a function of dispersion

We can change the group delay dispersion (GDD) of the laser pulse by adjusting a grating compressor located after the regenerative amplifier of our Ti:Sapphire system. Adding or removing dispersion adds a phase $\phi_2(\omega-\omega_0)^2$ to the spectral envelope $\mathcal{E}(\omega)$. In the temporal domain, this corresponds to stretching the pulse and simultaneously adding or subtracting a quadratic phase (i.e. a chirp) to the electric field.

For the flat pulse front it is the quadratic phase which is most important, because it allows us to compensate for the Kerr dephasing. But for the tilted pulse front it is the stretching which most important and we don't notice the quadratic phase at all.

Flat pulse In the flat-pulse geometry, the acceleration happens so quickly that the electrons hardly move. Thus the laser envelope takes the 1D form $\mathcal{E}(\tau)$. In section 2.3 we used this fact to show that dispersion alone will have no effect on the acceleration. We then showed if a Kerr effect is present, adding negative dispersion can flatten part of the phase and thus increase the energy gain. The way the compensation works is shown in the cartoon of Fig. 3.13: by applying the proper amount of dispersion, the chirp can cancel the Kerr phase (to second order) and help keep the electrons properly phased.

We demonstrate this effect on an 800 nm gap[13] DLA channel by recording a 2D scan of the dispersion and laser energy. As shown in Fig. 3.13(b,c), changing the dispersion has little effect at low intensity (where the Kerr effect is negligible), but a large effect at high intensity (where it can flatten a full π phase change).

The compensation process can be understand intuitively by comparison to solitons, where anomalous dispersion perfectly balances the Kerr effect. It can also be understood as the inverse of a common spectral broadening technique: rather than applying self-phase-modulation to broaden a spectrum and then dispersion to compress the laser to make a *shorter* pulse, we are applying dispersion and then self-phase modulation in order to create a compressed version of a *longer* pulse. The longer pulse is then able to accelerate the electrons to higher energies.

Of course, lengthening the laser sacrifices gradient, and so it is not the ideal way to increase energy gain. The real reason we are interested in compensation is that it shows the power of shaping the drive laser. Thus it helps motivate the SLM-based beam control from § 2.6.

PFT In the PFT geometry, the electrons only see the transverse laser envelope, $\mathcal{E}(z)$. Thus the dispersion only shows up as a change in E_0 and as a change in the number of

---

[13] We had to use an 800 nm gap because the signal from a 400 nm gap channel is to small for most of this plot.

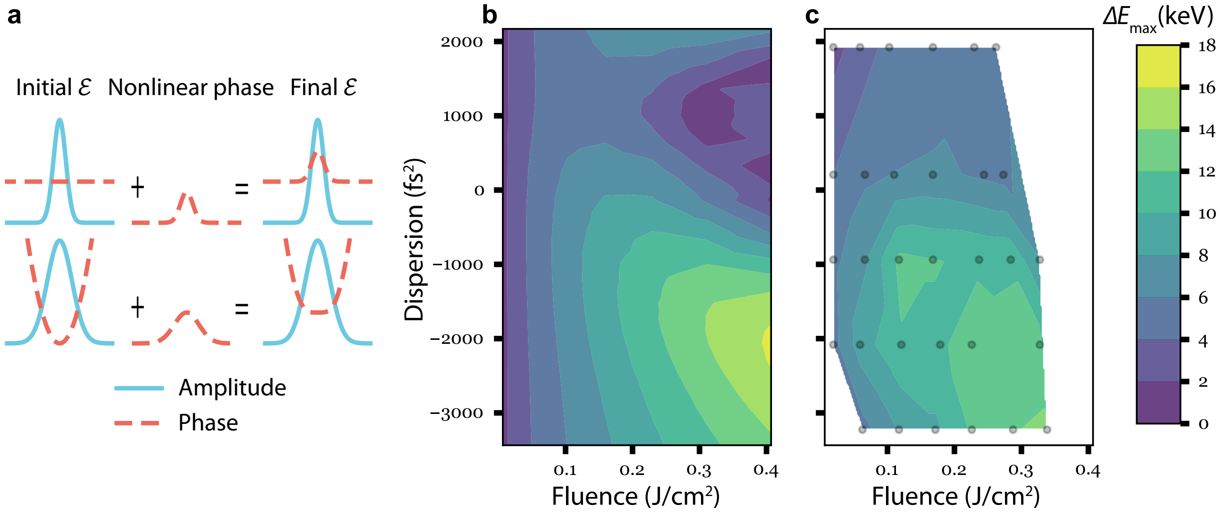

Figure 3.13: Compensation of nonlinear dephasing by tuning the laser phase. (a) Cartoon showing how additional dispersion (bottom row) can flatten the phase and thus increase the energy gain. (b) Theoretical energy gain as a function of fluence and dispersion (as determined at the entrance to the fused silica wafer). (c) Measured energy gain (sampled at the location of the dots) at the same conditions as in (b)

electrons which can be accelerated. These effects both happen because dispersion lengthens the laser pulse.

We can see how the PFT slices the electron beam by recording images of longitudinal phase space. We use a transverse deflecting cavity to streak the beam in y and a spectrometer to disperse it in x (see Fig. 3.1), and the resulting images, shown in Fig. 3.14(a,b), reveal a region of the beam whose energy spread is increased by the laser. The slope (chirp) in this region comes from the 1.5 m drift from the DLA to the deflecting cavity, because the accelerated electrons move faster than the decelerated ones. In some sense, the deflecting cavity ends up working like a time-of-flight spectrometer.

The height of each accelerated slice corresponds to the pulse duration of the laser, and it is what changes with dispersion. To quantify this effect, we measure the fraction of accelerated electrons as the fraction of total charge for which $|\Delta E > \Delta E_{\max}/\sqrt{2}|$. As expected, this fraction (Fig. 3.14(c)) closely tracks the laser-pulse duration as measured via frequency resolved optical gating. Thus we can see how the PFT transforms our accelerator

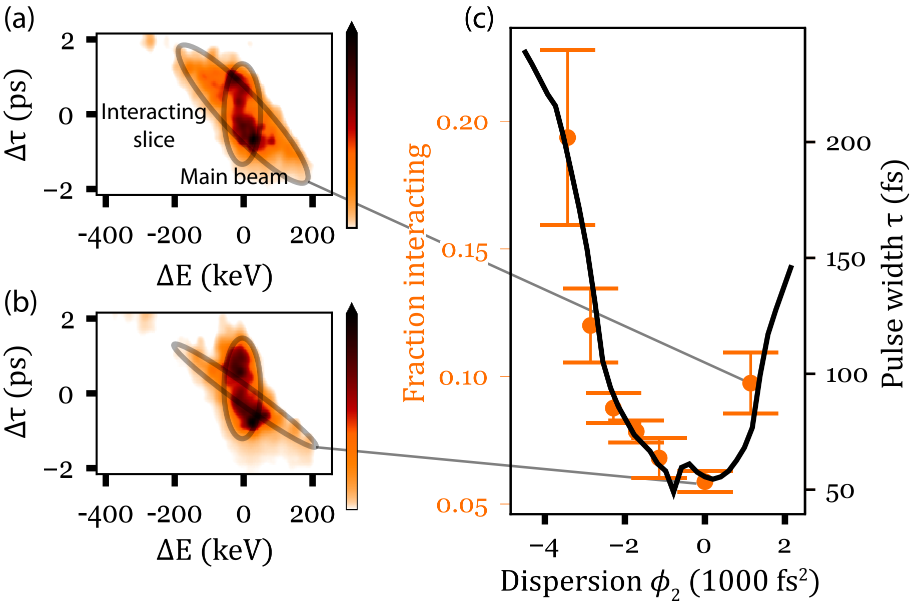

Figure 3.14: Measured phase space images (a, b) show that the DLA accelerates electrons in a temporal "slice" of the transmitted beam. The fraction of interacting electrons (referred to the transmitted beam charge, typically < 5 fC) is shown in (c) to track the measured laser pulse duration τ (in orange) as the laser compressor dispersion ϕ_2 is adjusted.

from accelerating a large number of electrons for a short distance to a small number of electrons over a long distance.

3.4.4 As a function of phase velocity

The longer the interaction inside the DLA, the more stringent the phase matching becomes. In practical terms, this translates in a strict condition on the laser angle of incidence, because, as we discussed in section 2.2, $\beta_m = \omega/(ck_g + \omega \sin(\theta_i))$. Thus we can expect to see interesting dynamics by changing θ_i (which we sometimes label θ_y or "tilt").

We will use the flat pulse as a way to check our control of β_m, but where things really get interesting is with the PFT cast. There the interaction is long enough that we can start to see longitudinal dynamics, and so by changing the phase velocity we can actually cause net acceleration or deceleration of the entire electron bunch. That offers us a first glimpse of "accelerator physics" in a relativistic DLA.

Flat pulse For the flat pulse θ_i is independent of ω, so the phase matching condition $\beta_m = \omega/(ck_g + \omega \sin(\theta_i))$ only accepts a very narrow range of ω compared to the bandwidth of the laser. In Fig. 2.3 this was represented by an integral across the short dimension of a very elongated ellipse. On the one hand, this make the acceleration rather insensitive to θ_i, since if we get it "wrong" we will just be using a different ω than we had planned on. But, on the other hand, this means that at a given θ_i the acceleration should be very sensitive to ω. We can test this by notching a portion out of the laser spectrum and looking for changes in the electron energy gain.

We create the notch by moving a wire through the laser in a dispersive region (e.g. in the middle of the regen compressor). We show the results in Fig. 3.15(a). The wire doesn't create an sharp notch, but it does create a tunable spectrum that we can use to connect pairs of θ_i, ω.

What we do, is we measure Δ_E for each of the 12 spectra in Fig. 3.15(a). We use a low intensity laser ($E_0 \approx 4\,\mathrm{GV/m}$), so that the Kerr effect is small and we can write $\Delta E \propto \sqrt{I(\lambda_0)}$ (see Eq. 2.11). Thus as the wire moves through the spectrum, we should see a reduction in the energy gain proportional to $\sqrt{I(\lambda_0)}$, but independent of the other wavelengths.

To find λ_0, we can plot our 12 measured energy gains against our 12 measured spectral amplitudes for each wavelength in the spectra. The wavelength which best matches the linear relationship $\Delta E \propto \sqrt{I(\lambda_0)}$ will be λ_0. We visualize this process by plotting a goodness of fit parameter ν/χ^2 in Fig. 3.15(b) (ν is the degrees of freedom and χ^2 is a typical test statistic for normally distributed residuals) for the linear fits. The maximum value in the plot then corresponds to the best estimate of the phase matching wavelength. It is not a delta-function, because there are strong correlations between the 12 spectra, and because the energy gain measured at low electric field is noisy. This is a complicated methodology, but the resulting phase-matching angle is robust to changes changes in the test statistic and the model for the energy gain, so we have some confidence in the results.

The solid curve in (b) shows the result for a DLA which we had intended to align for

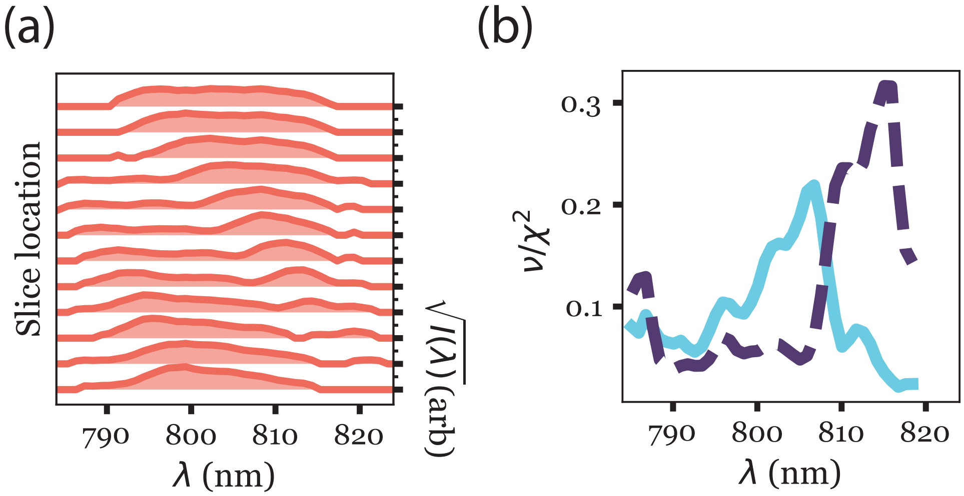

Figure 3.15: Phase matching angle and the DLA tilt. (a) A wire slices a notch out of spectrum (b) Goodness of fit for a linear model $\Delta E \propto \sqrt{I(\lambda)}$. *The peaks indicate the best fit for an aligned DLA (solid, blue) and a tilted DLA (dashed, purple).*

804 nm. The peak actually occurs at 806 nm, indicating that we can only align the DLA with about 2.5 mrad repeatability. We then rotated the DLA around the x-axis by $\theta_i = -2.5\,\text{mrad}$ to speed up the phase velocity, and we rotated the DLA about the y-axis to slow-down the electron velocity along the DLA ($\beta \rightarrow \beta\cos(100\,\text{mrad})$). We would expect the phase matching angle to increase by 6 nm, and we measure (dotted-line) an increase of 8 nm, which is reasonable given the uncertainty in this measurement.

PFT Compared to the flat-pulse the PFT-pulse is significantly more sensitive to θ_i. We can see this in the kinematic approximation, where we found that the dephasing comes out to be $\Delta E \propto \mathrm{sinc}(k_0 L \Delta\theta_i/2)$. The larger we can make L, the tighter the sinc pattern (see § 2.2).

We can see this in the measurements of Fig. 3.16(a) where we show ΔE versus the tilt angle. The 0.5 mm channel (purple) dephases more slowly than the 1.0 mm channel (purple). Looking more closely at the data, however, shows us that neither data set is very well described by a sinc function. This is a sign that we have moved beyond the range of validity for the kinematic approximation.

Next we want to compare energy gain to the energy loss. The energy loss is tricky to measure, however, because it overlaps with the stragglers (the electrons which strike the dielectric). To avoid systematic errors we define ΔE as the max/min energy which exceeds the noise of both the CCD and the stragglers combined (and we use the same threshold for both positive and negative energy changes). Even with all these controls we can clearly see that the energy gain and the energy loss are not symmetric. To emphasize the asymmetry, we plot the sum of the energy gain and energy loss in panel (b).

We can understand the asymmetry by considering the motion in a pendulum-like Hamiltonian (Eq. 2.21) with zero resonant phase. Phase plots this system are shown in panel (c) for beams corresponding to the three lines in panel (b). The origin of the asymmetry is then quite obvious: when the beam is injected below the resonant velocity it gains energy, whereas when the beam is injected above the resonant velocity it loses energy.

The amount of asymmetry can be quantified by second-order perturbation theory. Borrowing the result from small-signal FEL theory [138], we find that the spectrum will be shifted up or down by $F(\zeta) = \frac{1}{4}\frac{\partial}{\partial\zeta}\text{sinc}^2(\zeta/2)$, where $\zeta \equiv \eta k_m z/\gamma_m^2$. Maximum energy gain occurs when $\eta_{\text{max}} = 0.41\gamma_m^2/N \approx 0.12$, since we have $N = L_{\text{int}}/\lambda_g = 875$ (for the 1 mm channel). This suggests that we should find maximum energy gain when $mc^2\gamma_{\text{max}} = mc^2\gamma_m(1+\eta_{\text{max}})$ $=7.3\,\text{MeV}$, just like we see in panel (b).

While the perturbation theory does a good enough job of predicting which detuning will maximize the energy exchange, it does a bad job of predicting the amount of energy exchanged. In particular, it suggests that the maximum energy loss should match the maximum energy gain, while our results show a larger gain than loss. This is because the Kerr effect makes a small quadratic phase change across the structure which acts like an accelerating bucket and keeps the fast electrons better matched than the slow ones. To justify this explanation, we have done full simulations (using the simulation tree from Fig. 2.5) and plotted the results in Panels (a) and (b).

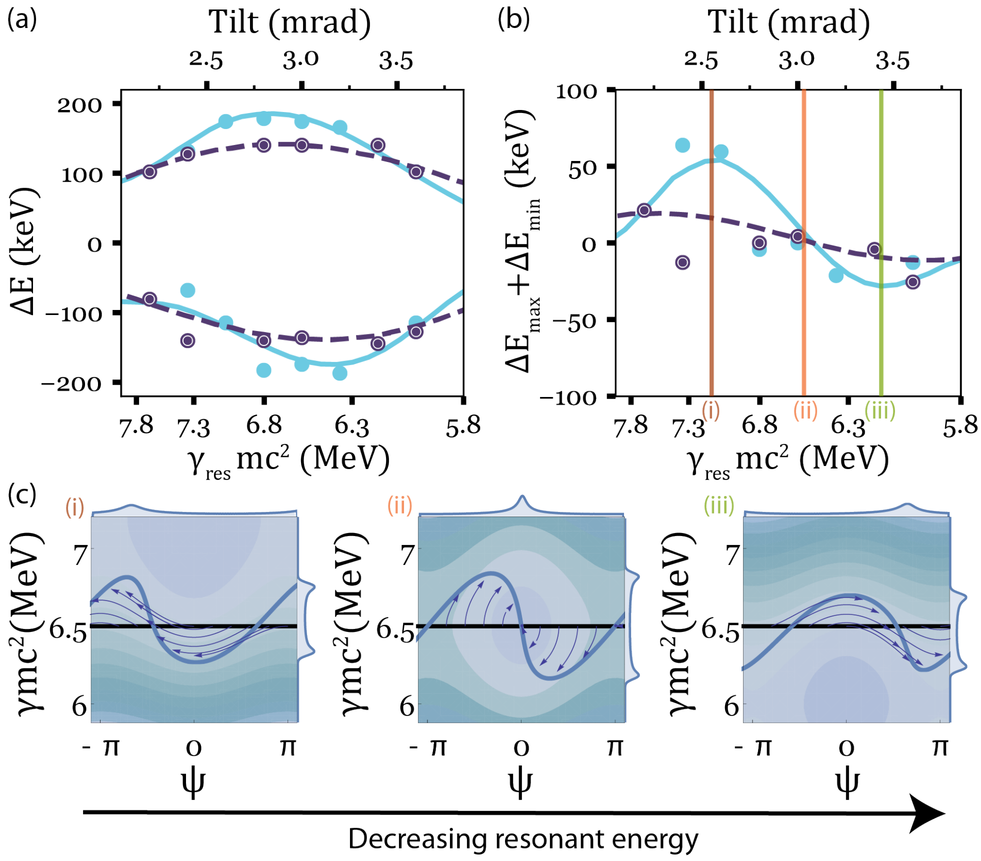

Figure 3.16: Longitudinal dynamics in a DLA. Maximum energy gain and energy loss (at top and bottom of the plot respectively) are shown in (a) and their difference in (b) as a function of θ_i for structures A (purple) and B (blue). Dots are measurements and lines are simulation. Computed Hamiltonian dynamics are shown in (c) for structure B, corresponding to the three resonant energies marked by vertical lines in part (b). Each plot shows the potential energy (contours), the initial (black) and final (blue) beam distributions, individual trajectories (arrows), and histograms of the final distribution (top and right).

3.4.5 DLA records

Our results currently hold the record for the highest peak gradient, highest average gradient, and highest total energy gain from a DLA. To be fair, these results should only be compared to relativistic DLA experiments, meaning that what we have done is beat the results previously achieved at NLCTA using the exact same structures [22, 28]. Nonetheless, the records in table 3.3 represent important progress and set a new bar for future experiments to beat.

Perhaps the most important record is the one for peak gradient, for which we list 1.9 GV/m. Sometimes this result gets cited as an "effective gradient" of 850 MV/m, which is what you would get if you calculated the effective length of the DLA by ignoring the Kerr effect. Obviously this is unphysical, but it is more immediately comparable to the previous results from NLCTA. For the purpose of measuring the fields inside the DLA, however, it is kind of meaningless. Given our understanding of dephasing, and how we can control it, it should be clear that the number which is relevant to future accelerators like [38] is 1.9 GV/m.

Table 3.3: Records for structures A, B, and C in PFT and flat conditions. The parenthesis in the C-flat column refer to the dispersion compensated data (at maximum energy gain).

Parameter	Symbol	Units	A PFT	B PFT	C PFT	A flat	C flat
Structure length	L	mm	0.5	1.0	0.5	0.5	0.5
Vacuum gap	g	nm	800	800	400	800	400
Structure factor (fit)	$\lvert c_1 + c_2\rvert$		0.08	0.08	0.11	0.08	0.2
Incident electric field	E_0	GV/m	7	7	7	9(6)	9
Interaction length	L_{int}	µm	470	700	470	22(88)	22
Max energy gain	$\Delta E_{\max}$	keV	250	315	280	7(15)	18.5
Average gradient	$\overline{G}$	MeV/m	500	315	560	320(170)	850
Peak gradient	G	MeV/m	560	560	770	720(360)	1900

Our other set of records come from setting up a laser with PFT. This allows us to accelerate for a full 1 mm, which previously had only been done with a low-gradient ps laser [22]. This is an important step, because even though PFT might not be used in an on-chip DLA [43], it is currently the only way we have to power a DLA that is long enough to test accelerator stability [38, 37].

The best (one-sided) PFT spectra we measured were shown back in Fig. 3.10. To make the fits shown in those plots we used the simulation chain from Fig. 2.5, only we replaced the Lumerical simulations by a fit parameter for $|c_1 + c_2|$. The results are shown in table 3.3. Notably, the efficiency of the 400 nm channel (C) appears to be less than when measured with a tilted pulse front than with a flat pulse front. This is because the PFT measurement is starved for signal: on the 400 nm channel only a few hundred electrons are accelerated per shot, and they are spread over hundreds of keV on the detector. Moreover, the higher accelerating gradient corresponds to higher de-focusing fields, causing more of the beam to be dumped into the wall of the DLA (simulations suggest 80% of the otherwise transmitted electrons will be kicked into the DLA wall). Thus the spectra from the 400 nm channel are noisier, having many islands of charge disconnected from the main beam (like the one seen in Fig. 3.10(c)). For this reason, we should defer to the flat-pulse measurement of this structure.

Nonetheless we can use the measurements from structures A and B to see the effect of long-distance + high gradient and set records for the largest energy gain and highest average gradient. Of course, these are still spectra from an unbunched beam, so the net energy gain in Fig. 3.16 is only 50 keV. Since our measurements are nearly a quarter synchrotron oscillation they are close to the limit of what can be achieved with a simple DLA. To beat this will we need to move beyond "DLA spectroscopy" into a realm of accelerator physics where we design and test structures which can focus, bunch, and accelerate a moving bucket of electrons.

CHAPTER 4

Relativistic microscope design

Shortly after the Knoll and Ruska invented the electron microscope it was realized that spherical and chromatic aberrations would limit the spatial resolution of the device to the nanometer scale [139]. Researchers then realized that one way to improve resolution was to use increasingly energetic electrons, because increasing the beam energy reduces the geometrical emittance and thus allows the microscope to use more paraxaial rays[1]. This line of thinking led to the development of a 1.2 MeV high energy TEM [141] with 0.1 nm resolution, but the complexity of this device (based on a Cockcroft-Walton generator) seemed to preclude going further, and eventually it was surpassed by lower-energy 200 keV microscopes with multipole aberration correcting units [142]. In our microscope design, we are reviving the the idea of using highly-energetic electrons, but we use them for a different purpose: suppressing space-charge repulsion.

Space charge is important in our microscope because we are trying to reduce the exposure time from milliseconds to picoseconds, causing an enormous increase in the peak current. The short exposure time will allow us to freeze defect motion, phase transitions, or other stochastic phenomena and capture an image of their instantaneous structure. In the introduction we profiled other techniques for ultrafast electron microscopy (UEM), but, for pushing single-shot time-resolution to the picoscecond scale, there is no alternative to high energy. By high energy we mean at least 3-4 MeV, which is beyond the reach of most direct-current (DC) electron sources. This requires a paradigm shift in microscope design in order to adapt concepts from the world of high energy particle accelerators.

---

[1] Originally high energy microscopes were also seen as an attractive way to image through thick samples [140], however the solution to this problem seems to have been learning how to make thinner samples.

One of the biggest changes in switching to an accelerator based source is that the electrons we use are incoherent, meaning that we can treat them as classical point particles. In other words, the number of electrons in an uncertainty-limited (i.e. Fourier-transform limited) phase space area is small ($\frac{\hbar}{2m_ec}\frac{Q}{\epsilon_n} << 1$), such that we are not near the limit set by the Pauli exclusion principle. Another way to interpret incoherence is to define a coherence length $L_c = \frac{\hbar}{m_ec}\frac{\sigma_x}{\epsilon_n}$ which is the spatial extent of a Gaussian wave-packet with angular spread σ_θ (with $\epsilon_n = \beta\gamma\sigma_x\sigma_\theta$)[2]. The coherence length then tells us the separation over which we can expect interference effects. For us it is much smaller than the resolution of the microscope, and so we can safely ignore quantum mechanics.

A classical re-interpretation of the coherence length is that it tells us which scattering angles we can distinguish from the direct beam. To see this, recall that the angle of a plane-wave scattered with wavenumber $k_\perp = 2\pi/L_c$ (e.g. from diffraction with atomic spacing L_c) is $\theta_s \approx \sin(\theta_s) = k_\perp/(2\pi/\lambda)$ for an electron with de Broglie wavelength $\lambda = h/p$. Then we can substitute in definitions until we get $\theta_s = L_c/(h/p) = 2\pi\sigma_\theta$, which says that the scattered electrons will be separated by a little more than the beam angular spread σ_θ (for a Gaussian distribution). Thus, even in the classical picture, the coherence length is a useful metric for judging which objects can be seen by the microscope. In particular, if L_c is less than the atomic-spacing in a material then we will have a difficult time getting diffraction contrast.

To increase L_c we can always collimate the beam by using a bigger spot-size, but this will reduce the electron density forming our image. Thus one of the pillars of MeV UEM design is finding ways to improve the transverse beam brightness, Q/ϵ_n^2 of accelerator-based sources. Conventional TEMs get their high brightness by using tip emitters to confine their electron source size, but we can't do this because tip emitters are incapable of generating the high current density we need in a single-shot UEM. Instead we try to push Q by photo-emitting lots of electrons into a high-gradient RF gun that will accelerate them before they have a

---

[2] In this representation the beam is made of a distribution of wavepackets, each with angular spread σ_θ. This is the standard approach [143], but it is only an approximation. The wavepackets might actually have smaller angular spread, in which case the beam will also have a distribution of mean angles.

chance to repel each other. As we discussed in the introduction (§ 1.3.1), every RF gun has a trade-off between charge, current, and transverse brightness, and so to optimize the injector we will need to know exactly what is required of the UEM. In this chapter we will do just that. We will start by giving a UEM design and describing the accelerator components that we use to make it. Then we will discuss the basic brightness requirements for the machine, and finally we will give a perturbation model for space-charge in the microscope and use it discuss how we might shape the charge distribution to limit space-charge aberrations.

MeV UEM components The basis for a relativistic single-shot transmission electron microscope (TEM) is a 2014 article by Li and Musumeci [144] in which they show the feasibility of the approach and outline a beamline capable of 10 nm-10 ps resolution. Since then, our experience implementing the microscope at Pegasus has lead us to update the design with an increased emphasis on generating contrast from thin samples. A cartoon of our current concept for the TEM is shown in Fig. 4.1. The key components are: an s-band high gradient RF gun, a higher-frequency (x-band) cavity to reduce the beam's energy spread, two compact permanent magnet quadrupole (PMQ) triplets (providing a total 900x magnification), a rectangularly symmetric collimator system for the back focal plane, an electromagnetic quadrupole set to image the back focal plane, and an in-vacuum microscope objective to achieve near diffraction limited imaging of the scintillator. The electromagnetic quadrupoles are the most obvious addition to the design of Li and Musumeci and, though they haven't been tested yet, their utility will become obvious in the next chapter when we discuss operating the microscope.

The first, and perhaps most important, part of the design is the high-gradient RF gun. As discussed in the introduction (§ 1.3), the RF gun is ideal for emitting high current electron bunches, and this has an obvious appeal for time-resolved studies. But the brightness requirements for a microscope are quite strict, and can only be met by pushing the RF gun to its limits. For this reason, the microscope design calls for 1.4 cell R.F. gun (currently under construction [145]) in order to get a higher extraction field. It also explains the 10 ps time-resolution: any shorter, and the gun simply can't produce the required transverse

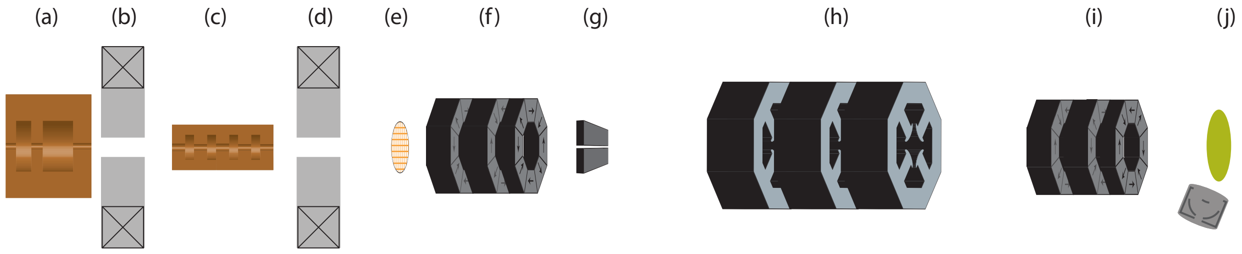

Figure 4.1: Cartoon design of a relativsitic single-shot TEM. (a) High-gradient RF photogun. (b) Gun solenoid. (c) Longitudinal phase space linearizer. (d) Condenser solenoid. (e) Sample plane. (f) Objective PMQ triplet (30x). (g) Rectangular collimator in back focal plane. (h) Conventional quadrupole multiplet. (i) PMQ triplet magnification stage (30x). (j) Thin Yag:Ce scintillator observed by in-vacuum infinity corrected objective.

brightness. The 10 ps bunch length, however, has it own problems: at s-band 10 ps is 10° of phase corresponding to $\Delta\phi^2/2$ relative energy spread. And while achromatic lenses can be made out of combinations of electrostatic and magnetostatic quadrupoles [146], at MeV energies we only have the magnetostatic variety. Thus a higher frequency cavity is needed to reduce the energy spread by making the second-order RF curvature vanish.

The next part of the design is the beam optics. The condenser lens, used to focus the beam onto the sample, is a conventional solenoid lens which we use to make a few micron spot size. The objective and magnification stage are more innovative: in order to make the beamline fit inside the Pegasus bunker we have developed high gradient (600 T/m) permanent magnet quadrupoles (PMQs). Proving that these lenses are suitable for use in a microscope is one of the experimental results for this thesis.

All together, the quadrupole stages are designed to image the sample with 900x magnification. When the aperture stop is closed down to block scattered electrons, then a shadow of the sample will form on a single-crystal Yag:Ce scintillator. The light from the scintillator is then imaged by an infinite-conjugate in-vacuum microscope objective (15x magnification[3],

---

[3] The magnification of a picture will, of course, also depend on the camera lens used to image the virtual image projected by the microscope objective.

0.28 NA) onto an intensified camera[4]. The in-vacuum objective allows us to both have a high collection efficiency and have near diffraction-limited resolution. Since the scintillator peaks in the green part of the spectrum, the resolution after the microscope objective will be, at best, around 0.6 nm. In practice, aberrations and space-charge effects in the electron transport are expected to limit the microscope to 10 nm resolution. Nonetheless, we must be careful to avoid blurring when imaging the scintillator, and thus we are forced to use a very thin scintillator (20 µm) in order to match the depth-of-focus of the detector optics.

4.1 Contrast mechanisms

The role of the microscope is to map each point in the object space to a point in image space, without regard for the initial angle. If we were to do this perfectly, then we would not be able to see the sample, because there would be no difference between the scattered and unscattered electrons. Thus we must introduce some mechanism which maps the unscattered and scattered electrons differently, and we call this mapping the contrast mechanism.

In our microscope we generate contrast by introducing an aperture at the back focal plane of the objective lens. At this plane the scattered and unscattered beams will be separated from each other, like we show in Fig. 4.2(a), and so we can clip one of them with a beam-stop. When the remaining electrons reach the image plane, we will see an intensity difference between locations where there was a scattering object and where there wasn't.

Often the contrast is imperfect, because the aperture-stop (a.k.a. "collimator") can't fully discriminate between the two beams. This might happen because the aperture is in the wrong location (for example, we often use the edge of the PMQs as an aperture, even though they aren't small enough and aren't in the back focal plane), or it might happen because the scattering angle is less than twice the divergence ($\theta_s < 2\sigma_y'$). This situation is shown in panel (b), and it raises the question of where exactly to put the edge of the aperture a. We will put off answering this question until the next section, where we develop a model for the

---

[4] A Princeton instruments PI-MAX III is used to detect the dim signal from a scintillator. The camera can be gated to reduce noise from dark current.

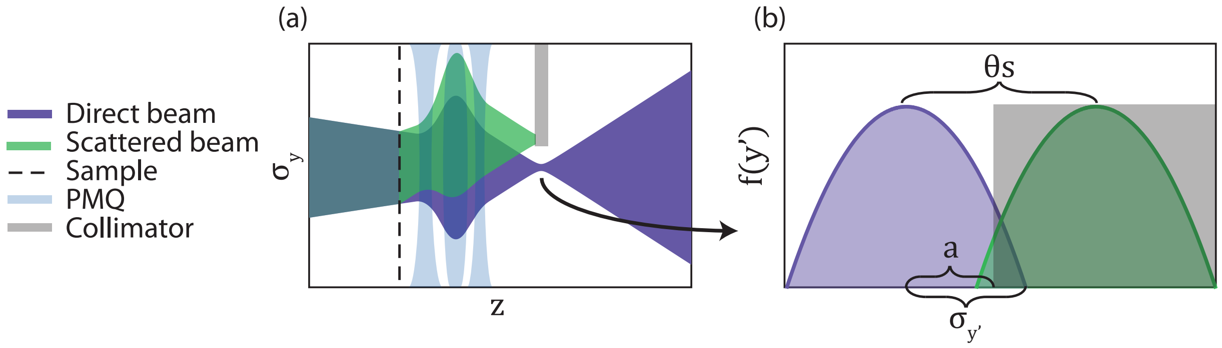

Figure 4.2: Contrast mechanism: (a) Beam envelopes of the direct and scattered beam in the vicinity of the back focal plane. (b) Beam distributions in the back focal plane, showing their geometry relative to the collimating aperture.

signal-to-noise ratio in the microscope.

This technique is called amplitude contrast, and in conventional microscopy it is only one of many techniques for generating contrast [147]. We use it because it works well with incoherent electrons, like the ones we generate at Pegasus. The same is true of Livermore's DTEM, which we profiled in the introduction [65]. The main problem with it is that it requires a strongly scattering sample, and so it places some constraints on what materials our microscope will be able to study.

One way to get a large scattering angle is to look at electrons which undergo Rutherford scattering from high-Z atoms. This is called mass-thickness contrast, and in practice it includes contributions from a combination of elastic and inelastic collisions. The main effect, however, is multiple scattering, for which the scattering angle scales like $\theta_s \propto \gamma^{-1}\sqrt{t/X_0}\ln(t/X_0)$ (where X_0 is the radiation length). This is a very generic form of scattering, and it applies to almost all the materials we have looked at with our microscope. In particular, it is the reason we can see the TEM grids which we use to calibrate the magnification of our machine. The grids are 20μm thick and made of copper or gold, and so we can calculate the multiple scattering and inelastic energy loss as $\theta_s(\mathrm{Cu}) = 0.1, \Delta E(\mathrm{Cu}) = 29\,\mathrm{keV}, \theta_s(\mathrm{Au}) = 0.2, \Delta E(\mathrm{Au}) = 68\,\mathrm{keV}$ [148, 149]. The scattering is so strong that we are able to get contrast (and see the difference between copper and gold), even when the aperture stop is in the wrong location [1].

More interesting for the eventual science applications of the microscope is diffraction contrast. Electrons have a de Broglie wavelength of $\lambda = h/p$ yielding a Bragg diffraction angle of $\theta_s \approx n\lambda/2d$. For relativistic electrons (as $\beta \to 1$) we can approximate $h/p \approx 1.2\,\mathrm{pm}/\gamma$ and $\theta_s = 1.2\,\mathrm{pm}/2\gamma d$. A typical lattice plane spacing for gold is $0.143\,\mathrm{nm}$ leading to a scattering angle of $\gamma\theta_s \approx 5\,\mathrm{mrad}$.

Notice that both diffraction scattering and multiple scattering have angles which scale like $\theta_s \propto 1/\gamma$. This is no accident, it happens because these kicks have a momentum dependence rather than an angular dependence. However, we don't actually pay a penalty for going to high energies, because the angular spread of the beam damps out as the beam gains energy ($\sigma_{y'} \propto 1/\gamma$).

4.2 Beam brightness

Making a high-resolution TEM requires us to put a large number of electrons in a small volume of phase space. On the one hand, we need to put enough electrons onto the sample, within a 10 ps exposure, to illuminate the entire image. But on the other hand, we need to keep the angular spread smaller than the scattering angle, or else we won't get contrast. Additionally, the relative energy spread needs to be below 10^{-4} to avoid chromatic aberrations [144]. Sum it all up and we have constraints on all six dimensions of the beam phase space.

In order to estimate the beam brightness, we need to make some criteria to decide when an image is resolvable. For this model we consider imaging a solitary object in the middle of a large field of view, as shown in Fig. 4.3. We treat the spatial resolution of the microscope as perfect, and calculate what happens when the contrast is not. For a given contrast level and electron density, we want to know if we can distinguish the object from the background.

To judge this we use the Rose criterion [150]. We consider the background of the image to have an average density $\langle n_b \rangle$, which the object modifies by some signal density $\langle \Delta n_s \rangle$.

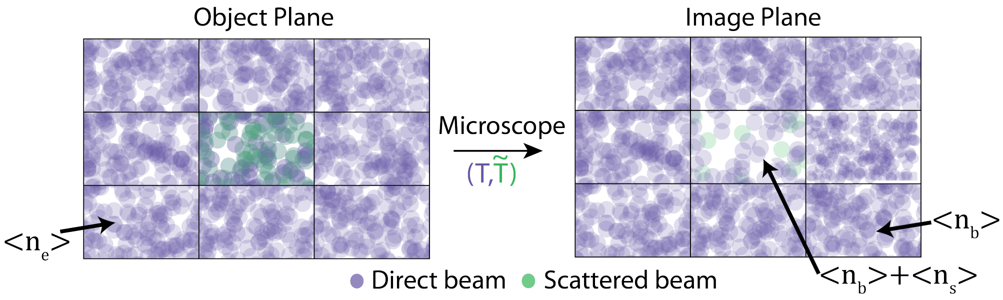

Figure 4.3: Rose criteria for signal detection. We check to see if we can distinguish a box with no scattered electrons to one with some scattered electrons.

We can then define a contrast $C = \langle \Delta n_s \rangle / \langle n_b \rangle$ which varies from 0 (no contrast) to 1 (full contrast). We assume the contrast is small, so that the shot-noise is constant across the image. For counting statistics the noise is Poissonian, so variance of the total number of counts in a box of area A will be equal to the mean number of counts: $\sigma_n^2 = A\langle n_b \rangle$. Then we can define a signal-to-noise ratio as $\mathrm{SNR} = A\langle \Delta n_s \rangle / \sigma_n$, In terms of the contrast, this is $\mathrm{SNR} = C\sqrt{A\langle n_b \rangle}$. Rose, for whom this model is named, found (experimentally) that an SNR of at least 5 is required to detect a signal by eye (with 50% success) [151]. Although the Rose criteria has been replaced by more advanced signal detection models, these results are a useful rule of thumb and match Bayesian inference under certain restrictions [150].

Already we can see that the SNR favors beams with higher contrast, more electrons, and bigger objects. Unfortunately, adding more electrons to the beam usually reduces the emittance and thus makes the contrast worse. In order to optimize on these quantities, we consider the collimation aperture we introduced last section, in Fig. 4.2. Looking at panel (b) of that figure, we see the aperture will pass some fraction of the direct beam and some smaller fraction of the scattered beam. We call these fractions T for the direct beam and $\tilde{T}$ for the scattered beam. Thus, if the incident beam has an average density of $\langle n_e \rangle$, then the background level at the image will have a density of $T\langle n_e \rangle$ electrons. Then, to calculate the signal, we consider what happens to an object which has a scattering probability p. We find $\langle \Delta n_s \rangle = T\langle n_e \rangle - (p\tilde{T}\langle n_e \rangle + (1-p)T\langle n_e \rangle) = p\langle n_e \rangle (T - \tilde{T})$, which is the difference in transmission fractions multiplied by the density of scattered electrons. Finally, the noise level of the Poisson process will be $\sigma_n^2 = A(T\langle n_e\rangle + \langle\sigma_c\rangle)$, where $\langle\sigma_c\rangle$ is the average noise level of the camera. Putting these terms together we find that:

$$\mathrm{SNR} = p\langle n_e A\rangle(T - \tilde{T})/\sqrt{\langle n_e\rangle AT + \langle\sigma_c\rangle A} \tag{4.1}$$

Now we want to optimize the SNR. The first thing we can do, is we can change the aperture gap a, which controls T and $\tilde{T}$. Looking at Fig. 4.2(b), it is obvious how to do this if $2\sigma_{y'} < \theta_s$, since we can just set a to clip the entire scattered beam and get $T = 1$, $\tilde{T} = 0$. Another case where the answer is obvious is when the camera noise is very large. In that case, we are just optimizing $T - \tilde{T}$, and so we set a to the point where the two parabolas intersect. If neither of these are the case, then we can still optimize $\mathrm{SNR}(a)$, but we have to do it numerically.

After optimizing on a we still have some freedom to continue optimizing the SNR. Depending on the value of the camera noise we may think about trading some charge density, N_e/A for smaller or larger σ_y'. For example, we could increase the beam size at the sample, which would trade $\langle n_e\rangle \propto \sigma_{y'}$ (assuming fixed emittance and that the sample is at a waist). This turns out to be a good trade, and for small camera noise we can numerically find that the optimum occurs when the distributions are barely overlapping: $\sigma_{y'} \approx 0.64\theta_s$. In this case, the SNR is $\mathrm{SNR} = \frac{0.9p\langle n_e\rangle A}{\sqrt{0.94\langle n_e\rangle A + \langle\sigma_c\rangle A}}$ such that SNR increases with $\sqrt{A\langle n_e\rangle} = \sqrt{N_e}$, the square root of the number of particles inside our resolution box. This can be further optimized by reducing the spot size on the cathode to get a better emittance. For fixed $\sigma_{y'} \approx 0.64\theta_s$, a better emittance will give us a better spot size, and thus more charge density like $\langle n_e\rangle \propto Q/\sigma_x\sigma_y$. The cigar regime scaling laws then suggest we should shrink the spot size on the cathode as much as possible (i.e. approaching a tip source).

Now we are in a position to plug in some numbers and calculate the beam parameters required for the microscope. We will need to use lower yield (and higher resolution) screens than in [144], so we will assume $A\langle\sigma_c\rangle = 5$[5]. Then, with a scattering probability of $p =$

---

[5] In [66] Li et. al. find that the camera noise level is 1/4 the yield of a single electron from a Lanex fine screen. They note a 200 µm thick YAG:Ce has half the light of a Lanex screen, and our 20 µm thick YAG:Ce will have 1/10 of that.

55%, the rose criteria sets N_e equal to 100 electrons[6]. For $A = 10\,\text{nm}\times 10\,\text{nm}$ we get an electron density of $1/(\text{nm}^2)$ which implies a 1 pC bunch charge for a 2 µm beam radius (and uniform beam distribution). Setting the beam divergence equal to a scattering angle of $\gamma\theta_s = 1\,\text{mrad} = 1.3\sigma_{x',y'}$ implies a normalized transverse emittance of 2.6 nm. Additionally, the energy spread is $(\Delta\gamma/\gamma) = 10^{-2}\,\%$ and the time resolution is $\Delta T = 10\,\text{ps}$ (full width) so that we can get an order of magnitude (normalized) 6D brightness as: $B \approx \frac{100}{A(0.64\gamma\theta_s)^2\Delta T(\Delta\gamma/\gamma)} = 2\cdot 10^{19}\,\text{A}/\text{m}^2/\text{rad}^2/\text{percent-energy-spread}$[7]. This beam has an average current of 100 mA and will need to maintain a current ripple of less than 2 mA (see footnote 14).

This brightness should actually be taken as lower bound for a 10 nm, 10 ps MeV UEM, because we have so far treated the microscope as if it had no aberrations. In fact, the microscope will have a point-spread function on the order of 10 nm, so that the edges of our object will get blurred out and convolved with the background. The combination of the beam shape, aperture stop, and microscope point spread function all combine to influence the resolution. This is often represented by taking the Fourier transform of the mapping from object to image space (including the apertures and the beam size). This "modulation transfer function" thus tells you what size periodic structures can be seen. For a good treatment of the method see chapters 22, 23, and 28 of [147].

4.3 Permanent magnet quadrupole lenses

The workhorse lens of a conventional electron microscope is the round magnetic lens, i.e. a solenoid lens [146]. The magnetic solenoid has several nice properties: it is mechanically simple, it is has an easily adjustable focal length, and it is inherently stigmatic. But it scales poorly to high electron energies since the focal length is proportional to γ^2. The solenoids on our beamline, which already weigh several hundred pounds, have $f > 0.5\,\text{m}$ such that a 30x

---

[6] I chose this p to match the number of electrons in [152], but it is quite large, and the scaling of N_e with p is quite severe. We can Taylor expand the SNR (using the rose criterion that SNR=5) to get $N_e \approx 5 + 30/(p^2)$.

[7] This brightness is in terms of the radii of the uniform distributions we plan to generate. To convert to an rms-based definition we need a factor $\sqrt{1/3}$ in the denominator for each dimension and a total brightness 27 times higher than this estimate).

magnification stage would have a 13.5 m working distance and take up most of the Pegasus bunker. To build a more compact beamline we turn to the magnetic quadrupole lens.

The focal length of a quadrupole with an aperture field B_0, gap radius g, and length d is $f = B\rho(g/B_0)/d$, where $B\rho[Tm] = p_0[MeV/c]/299.8$ is the magnetic rigidity and we have assumed a thin lens $d << f$. Both the residual field of permanent magnets and the saturation field of iron allow aperture fields on the order of 1 T, but permanent magnets can be more easily assembled with a small gap, g, and then placed entirely inside vacuum. Our quads, for example, have a gap of 3.5 mm and generate a record gradient of 600 T/m in order to provide a focal length of 1.5 cm when installed on the Pegasus beamline.

The main drawback of the quadrupole lenses is that they are simultaneously focusing in one plane and defocusing in the other. This means that they must be arranged in sets (called multiplets) in order to achieve net focusing in both planes. The additional magnets add degrees of freedom which make the microscope harder to align. And permanent magnets are all the harder to align because they have to be physically shifted (compared to adjusting the current in an electromagnet). Despite these engineering challenges, the PMQs remain appealing for their compact form.

4.3.1 Halbach quadrupole

Halbach [153] analyzed permanent magnet multipoles built out of an array of wedges like those in Fig. 4.4(a). To make a quadrupole, the wedge centered at an angle θ needs to have a magnetization pointing in the $3\theta \pm \pi/2$ direction (the ± giving the orientation of the quadrupole). And the more wedges the magnet is divided into, the more perfectly quadrupolar the field is. For M identical wedges of residual field B_r having inner gap g and outer gap R, the magnetic field is can be expressed as a sum:

$$B_x - iB_y = B_r \sum_{\nu=0}^{\infty} \left(\frac{y-ix}{g}\right)^{1+\nu M} \frac{2+\nu M}{1+\nu M} \left(1 - \left(\frac{g}{R}\right)^{1+\nu M}\right) \cos^{2+\nu M}(\frac{\pi}{M}) \mathrm{sinc}\left(\frac{(2+\nu M)\pi}{M}\right) \tag{4.2}$$

where we have assumed the quadrupole is oriented to give a normal (rather than a skew) quad. The $\nu = 0$ term gives a pure quadrupole and for large M and large R it has an

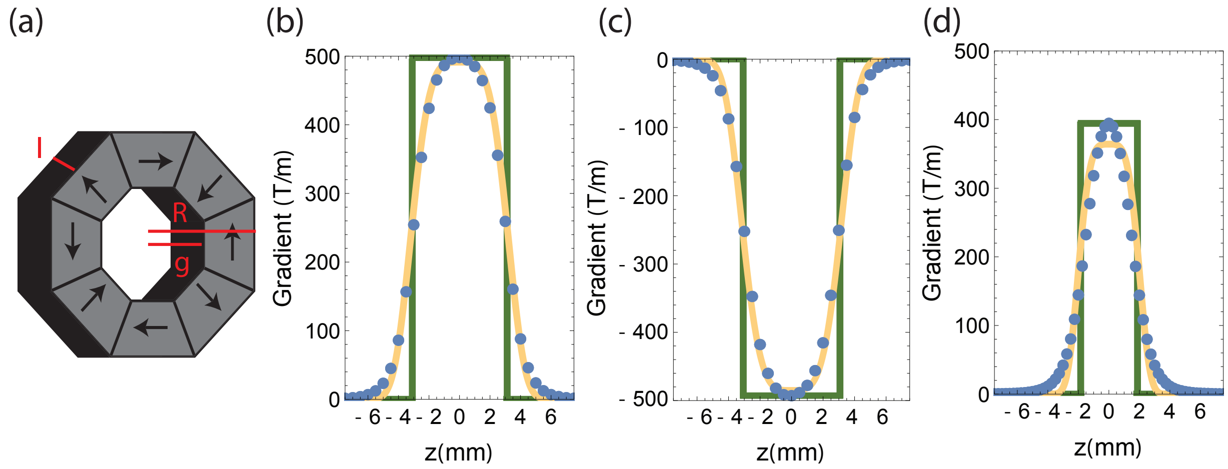

Figure 4.4: PMQ diagram and field maps for triplet 'B'. (a) Diagram of a M=8 Halbach PMQ showing the direction of the magnetization of each wedge. (b-d) Hall probe measurements of gradient for quadrupoles B1, B2, and B3 respectively. Also shown are supergaussian (light yellow) and hard-edge (green) fits.

aperture field that is nearly $2B_r$. The next harmonic is the $2+M^{\text{th}}$ multipole, and its $(x/g)^{M+1}$ dependence means that it will only be noticeable near the edges of the gap.

The permanent magnet quadrupoles developed for our microscope are constructed from wedges of Neodymium Iron Boron (NdFeB), arranged in $M = 16$ ('A' quadrupoles) and $M = 8$ ('B' quadrupoles) Halbach arrays. They have gaps of 3.5 mm (diameter) and physical lengths of either 6 or 3 mm depending on the desired focal length. As a result they have significant fringe fields. The numerical code Radia [154] is used for magnetostatic simulation of the fields. Once built, the integrated gradient through the quads can be measured with a miniature hall probe or via the vibrating wire technique [155] with the results listed in Table 4.1 and illustrated in Fig. 4.4. Note that the Hall probe appears to underestimate the peak field due to it's 0.5 mm active area.

When numerically tracking particles through a quadrupole we often want to replace a large field map with an analytical approximation. In addition to speeding up the simulations, the analytical field is nice because it satisfies Maxwell's equations. The field map, on the other hand, has to interpolate between points and so it can cause numerical emittance growth. General particle tracer (GPT), our particle tracker of choice, now has a built-in model based

Table 4.1: Parameters for the two sets of PMQ's used in the microscope.

	Gradient (Radia)	Gradient(Hall probe)	d_{eff}	Vib. wire Gd_{eff}
A1	597 T/m	523 T/m	6.16 mm	3.23 T*m
A2	-597 T/m	-512 T/m	6.16 mm	3.55 T*m
A3	495 T/m	396 T/m	3.6 mm	1.68 T*m
B1	597 T/m	500 T/m	6.25 mm	-
B2	-597 T/m	-492 T/m	6.25 mm	-
B3	495 T/m	395 T/m	3.9 mm	-

on fitting the on-axis gradient with Enge-functions, but when we started modeling PMQ's it did not, and so we developed our own model.

To start, we find an expression $f(z) = G(z)|_{r=0}/G_0$ to describe the the on axis quadrupole gradient. Then we proceed perturbatively: first we set $\boldsymbol{B} = -\nabla\phi$ (in free space with no sources) and then we assume $\phi^0 = -xyf(z)$. Then we try to solve $\nabla^2(\phi^0 + \phi^1) = 0$ and, keeping the $x - y$ symmetry in mind, find that $\phi^1 = (x^3y + y^3x)\partial_z^2 f(z)/12$. The resulting magnetic field (normalized to gradient G_0) is:

$$\begin{aligned} B_x/G_0 &= y\left(f(z) - \partial_z^2 f(z)(3x^2 + y^2)/12\right) \\ B_y/G_0 &= x\left(f(z) - \partial_z^2 f(z)(3y^2 + x^2)/12\right) \\ B_z/G_0 &= xy\left(f(z) - \partial_z f(z)\right) \end{aligned} \tag{4.3}$$

Our magnets are fit reasonably well by a super-Gaussian $f(z) \propto \exp(-x^4/\sigma^4)$ and so we use the fits in Fig. 4.4 for in-depth particle tracking. When doing simple linear optics we use an even simpler fit based on a flat-top magnetic field. Following convention, we choose a flat-top with the same peak value as the actual field, and then we calculate the effective length which gives the same integrated kick gd_{eff} (i.e. $d_{\text{eff}} = \int dz\, f(z)$).

4.3.2 Quadrupole multiplet

To achieve net focusing in all directions we must combine quadrupole lenses into arrays called multiplets. The smallest combination, a doublet, can focus in both directions, but the smallest combination which can produce stigmatic images is a triplet (stigmatic implies the image planes for the x and y directions coincide and have the same magnification). To see this, consider that we need to set four independent parameters: the image plane in x, the image plane in y, and the magnifications m_x and m_y (If $m_x \neq m_y$ this can be called pseudo-stigmatic). This requires an algebraic system with four tunable parameters, namely the distances between the object, each of the three quadruples, and the image.

In the linear approximation ("Gaussian optics") we can describe the beam dynamics in a multiplet as an algebraic system constructed from the matrices:

$$\begin{aligned} \begin{pmatrix} x \\ x' \end{pmatrix}_f &= \boldsymbol{M} \begin{pmatrix} x \; x' \end{pmatrix}_o \\ \boldsymbol{M}_{\text{quad}} &= \begin{pmatrix} \cos\left(\sqrt{\kappa} d\right) & \frac{1}{\sqrt{\kappa}} \sin\left(\sqrt{\kappa} d\right) \\ -\sqrt{\kappa} \sin\left(\sqrt{\kappa} d\right) & \cos\left(\sqrt{\kappa} d\right) \end{pmatrix} \\ \boldsymbol{M}_{\text{drift}} &= \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \end{aligned} \tag{4.4}$$

Where the quadrupole has oscillator strength $\kappa = G/B\rho$ and length d, and the drift has length L. If the quadrupole is "thin" then we can Taylor expand the trig functions to first order. Also, notice that the gradient, G, switches sign between the x and y directions (leading to cosh's and sinh's). We will refer to elements of the matrix by the notation M_{ij} to refer to the i^{th} row and the j^{th} column. For example, imaging means $M_{12} = 0$ with magnification M_{11} since we would have $x_f = M_{11} x_o$.

The triplet design we use comes from the optimization Lim et. al [156] did to find the optimal focus for an inverse Compton scattering experiment. They build a matrix for an arbitrary triplet made of thin-lenses with focal lengths f_1, f_2, and f_3 which are separated by lengths of L_{12}, L_{23}. They then maximize the term M_{21} term (i.e. making a short focal length triplet) subject to 3 constraints: (a) that M_{21} be the same in both planes, (b) that M_{11} be the same in both planes, and (c) that there is some minimum focal length f_{min} which corresponds to the strength of the middle lens. They find an optimum for $f_1 = 2f_{\text{min}}$ and $L_1 = L_2 = f_{\text{min}}$. In practice we get close to this $f, f, 2f$ configuration by making PMQs of

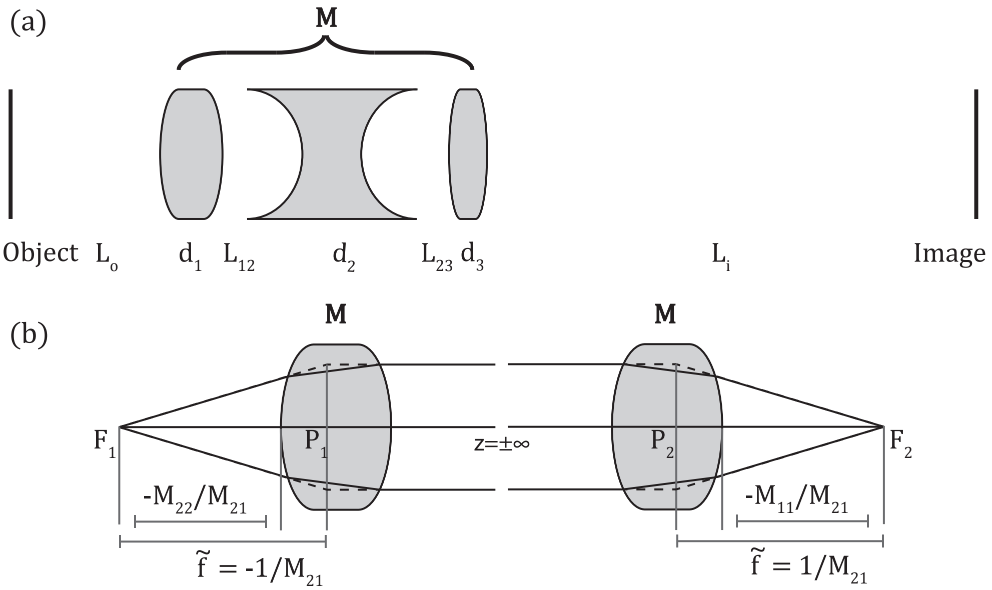

Figure 4.5: Cartoon of a PMQ triplet. (a) Triplet in imaging configuration. (b) Geometry of the effective-focal-length model showing the principal planes.

equal gradient with lengths of $d, d, d/2$.

Intuitively, we can see why the Lim design also works for stigmatic imaging: the beam exits the triplet with the same size (M_{11}) and divergence (M_{21}) in both directions. If the image forms a distance d_i behind the triplet, then the magnification is $M_{11} + M_{21}d_i$ and we can see the advantage of having optimized for a short focal length (large M_{21}). The Lim prescription does not, however, guarantee that both x and y will form at image at the same plane, although in in the limit of large M_{21} it guarantees that the image planes will be close to each other. To achieve simultaneous imaging, we will actually have to adjust the quad-spacing away from Lim's prescription (we also need corrections to account for the thick-lens nature of the real quadrupoles).

It can be useful to reduce the action of the PMQ triplet to that of a single thin lens. Were the triplet just a thin-lens in disguise then we would identify $M_{21} = -1/f$, but in fact it is not, and so we have to introduce the idea of principal planes as shown in Fig. 4.5. The idea is to call $-1/M_{21}$ the effective focal length ($\tilde{f}$) and then construct two planes at which all of the diffraction can be thought to happen. We define the first principal plane

(P_1) as the plane for which rays emanating from the front focal plane ($x(z = F_1) = 0$) would intersect the rays which emerge collimated ($x'(z = \infty) = 0$), and the second principal plane (P_2) as the plane for which initially collimated rays ($x'_0(z = -\infty) = 0$) intersect rays coming from the back focal plane ($x_f(z = F_2) = 0$). Then we can use the simple lens makers equation $s_o^{-1} + s_i^{-1} = \tilde{f}^{-1}$ so long as we define the object distance s_o from the object to the first principal plane and the image distance s_i from the second principal plane to the image. By jumping between P_1 and P_2 we can replace the triplet with a thin lens. For triplet 'B' as measured by the Hall probe (table 4.1) and using a 5 MeV beam energy we find that $\tilde{f}_x = 16\,\text{mm}$, $\tilde{f}_y = 15\,\text{mm}$, and most of the principal planes are located about 1 cm inside the edge of the triplet (P_2 in the y plane is the exception, it is located 3 cm after the triplet, and it is enough different that we end up with stigmatic imaging for the nominal working distance of 41 cm).

The asymmetry between the x and y effective focal lengths hints at one of the main challenges of the triplet setup: For a given triplet (and beam energy) we can only achieve imaging from a single object plane to a single image plane. If we were using a spherical lens, then we would be able to achieve imaging at any working distance by adjusting the lens location L_o (or, for a conventional lens, the focal length via the current). Instead, we have to tweak the three parameters L_o, L_{12}, and L_{23}. This requires very precise motion control of the PMQs along the optical axis.

An alternative would be to replace the triplet with an anti-symmetric quadruplet. Such an arrangement guarantees stigmatic imaging such that, like a round lens, there is only one condition required to be made imaging (and the image is tuned by adjusting just the lens position) [157, 146]. This nice feature requires perfectly anti-symmetric PMQs, which, given the variability of the residual magnetization, is difficult to achieve with PMQs. And without perfect anti-symmetry the quadruplet offers no simplification over the triplet. Quintuplets, however, have the opportunity to reduce the spherical aberrations, and thus may eventually be of interest.

4.3.3 Quadrupole aberrations

Even ideal lenses have unavoidable aberrations. Famously, Scherzer showed that the spherical (and chromatic) aberration coefficients for round lenses can never change sign [139]. This has prompted decades of research into producing and aligning complicated arrangements of multipoles which can cancel the round lens aberration [158]. The potential of quadrupoles for use in microscopes has a long history, and their ability to make a line focus has made them a key component in many aberration correction schemes [146].

A complete treatment of the geometrical aberrations in ideal quadrupoles can be found in [157]. Hawkes ultimately finds that the aberrations of a quadrupole system can be manageable, especially when a combination of electrostatic and magnetostatic quadrupoles are used to make an achromatic lens. At relativistic energies, and with a lower beam quality commensurate with bunched electron beams, we have to re-asses the importance of the various quadrupole aberrations.

In [144] Li and Musumci use start-to-end particle tracking simulations to assess the microscope resolution. They find that the aberrations due to ideal quadrupole fields are significantly less important than the scattering due to Coulomb forces[8], at least provided the relative beam energy spread is kept below 10^{-4}. In the next section we will show how smooth “space-charge” aberrations can dominate the resolution if the electron beam distribution is not suitably uniform.

That is not to say that we cannot see any of the ideal quadrupole aberrations. Quadrupoles produce significant distortion ($x_f \propto x_i^3$) and so when the field-of-view is large (compared to the resolution), then the image can appear stretched or bent. The distortion does not, however, blur the image, and so even with a large field-of-view it is possible to remove the distortion in post-processing (given, for example, a calibration image of a grid).

---

[8] Parasitic aberrations, that is the aberrations due to imperfections in manufacturing and aligning the magnets, has not been thoroughly explored for PMQs, but is expected to also be small.

4.4 Coulomb interactions

All along the microscope the electron beam will repel itself and distort the trajectories of individual electrons. For low enough charge density this distortion always becomes negligible, but when pushing limits of single-shot time resolved microscopy, it can be significant. In fact, we expect the resolution limits of our microscope to be determined more strongly by collective effects than by geometric or chromatic aberrations.

The bare Coulomb potential is particularly onerous since it diverges as electrons get close to each other. But, when the inter-particle separation is small compared to the Debye length ($\lambda_D = \sqrt{\epsilon_0 kT/n_0 e^2}$), we can expect that the bare Coulomb interactions will be strongly screened and thus that we can describe the arrangement of charges by a continuous charge distribution. The electrostatic potential of this continuous distribution is called the space-charge force and it is the primary component of the Coulomb interactions for our beam.

When the charge distribution takes the form of a uniformly filled ellipsoid (in x-y-z) the space-charge force is linear in each of the coordinates. Thus we like to think of the Coulomb interaction as a defocusing lens distributed along the beamline. The strength of the Coulomb defocusing is complicated: it depends both on the initial charge distribution and on the all of the optics along the beamline; but, provided the space-charge force is everywhere linear, it can be perfectly compensated by tuning the magnetic lenses. Unfortunately it is difficult to make a beam which has linear space-charge along the entire column, and thus the microscope will have some unavoidable "space-charge aberrations".

The main trouble with "space-charge aberrations" is that in order to calculate them, one has to solve a fully coupled Vlasov equation. With magnetic lenses, the optical properties (including aberrations) are properties of the static-field alone and so can be calculated without consideration of the collective beam dynamics. But space-charge depends explicitly on the current profile along the microscope, and the current profile, besides depending on the initial conditions at the electron gun, also changes in response to any movement of the magnetic lenses. This makes it difficult to understand what will happen when combining

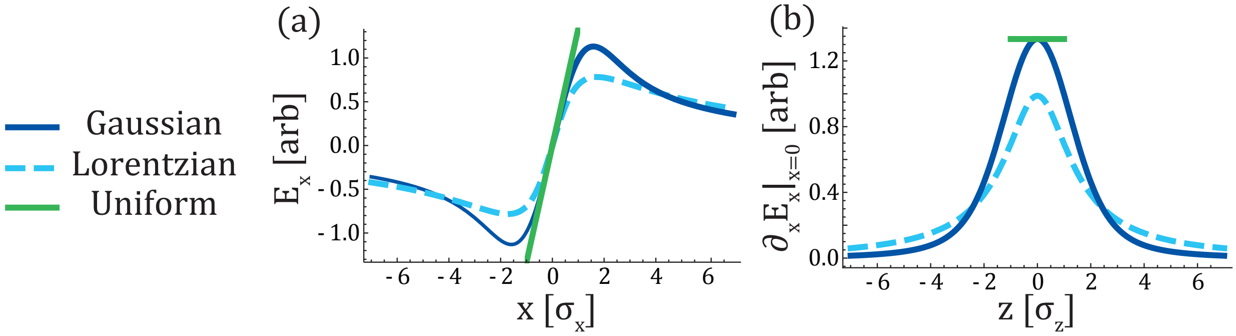

Figure 4.6: Space charge potential for several beam distributions. (a) E_x-kick (b) Defocusing strength (slope from a) as a function of z. The plotted distributions are "round" ($\sigma_x = \sigma_y = \sigma_z$) and have a peak charge density of 1.

multiple elements and thus makes aberration compensation a complicated proposition. We try to reduce the impact of space-charge by shaping the input the beam distribution. In this section we will discuss the space-charge force as it applies to several beam distributions and then solve approximate models to understand the impact of linear and non-linear space charge kicks.

4.4.1 Space charge potential

Since we use quadrupole elements we will calculate the potential of an elliptically symmetric charge distribution. By elliptically symmetric we mean, explicitly, charge distributions which only depend only on the generalized radius $s = \left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2 + \left(\frac{z}{\sigma_z}\right)^2$ via some function $\rho = f(s)$. The Laplace equation can then be solved in elliptical coordinates [159] to give the potential:

$$\phi(x,y,z) = -\frac{\sigma_x\sigma_y\sigma_z}{4\epsilon_0}\int_0^\infty \frac{dt}{\sqrt{(\sigma_x^2+t)(\sigma_y^2+t)(\sigma_z^2+t)}}\int_0^{S(t)} ds f(s) \tag{4.5}$$

where $S(t) = \left(\frac{x^2}{\sigma_x^2+t}\right) + \left(\frac{y^2}{\sigma_y^2+t}\right) + \left(\frac{z^2}{\sigma_z^2+t}\right)$.

In Fig. 4.6 we have plotted this integral for several beam distirbutions. Actually, we have plotted the derivative of the potential (to get the force) and the second derivative (to get the amplitude of the linear part of the defocusing kick). This gives a sense of how the space-charge force changes along the beam.

This formulation for ϕ is most useful when $f(s)$ is integrable such that potential reduces to a single integral. For example, a Gaussian distribution $f(s) = \frac{Q}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z}e^{-s/2}$ has the potential:

$$\phi(x, y, z) = \frac{Q}{(2\pi)^{3/2}4\epsilon_0}\int_0^\infty \frac{1 - \exp(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2+t}\right) - \frac{1}{2}\left(\frac{y^2}{\sigma_y^2+t}\right) - \frac{1}{2}\left(\frac{z^2}{\sigma_z^2+t}\right))}{\sqrt{(\sigma_x^2 + t)(\sigma_y^2 + t)(\sigma_z^2 + t)}} \tag{4.6}$$

The multivariate Gaussian is a popular distribution for low current machines because it provides a simple description of uncorrelated (or linearly correlated) phase spaces. It also has the nice property that after linear transport the beam remains Gaussian. Unfortunately, the space charge force is clearly non-linear and is also coupled between x and y such that the self-fields eventually cause the beam to deviate from a Gaussian.

Over short distances (or at low currents) we can calculate the space charge by assuming the beam remains Gaussian. We can get some intuition by considering the 2D limit ($\sigma_z \to \infty$) of a round beam (σ_x=σ_y). In this case the transverse field along, for example, the x-axis is: $E_x = \frac{2\lambda}{(2\pi)\sigma_x^2}(1 - e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}})/x$. For small x the Gaussian beam has a linear defocusing force ($E_x \approx \frac{2Q}{\pi^{3/2}\sigma_x\sigma_y\sigma_z}x$), but for $x > 1.5\sigma_x$, the field returns to zero (see Fig. 4.6). For elliptical beams the relations are more complicated but the general intuition holds: Near the axis the field is linear, but in the tails of the Gaussian it decays towards zero.

We can do away with nonlinearities if we have a uniformly filled ellipsoid ($f(s) = \frac{3Q}{4\pi\sigma_x\sigma_y\sigma_z}$ for $s < 1$ and $f(s) = 0$ otherwise). This has a potential inside the beam which is quadratic in the coordinates:

$$\phi(x, y, z) = -\frac{3Q}{16\pi\epsilon_0}\int_0^\infty dt\frac{\left(\frac{x^2}{\sigma_x^2+t}\right) + \left(\frac{y^2}{\sigma_y^2+t}\right) + \left(\frac{z^2}{\sigma_z^2+t}\right)}{\sqrt{(\sigma_x^2 + t)(\sigma_y^2 + t)(\sigma_z^2 + t)}} \tag{4.7}$$

In the 2D limit ($\sigma_z \to \infty$) this integral can also be evaluated, in which case we get the simple result analytical result $\phi = -\frac{\lambda}{2\pi\epsilon_0(\sigma_x+\sigma_y)}(\frac{x^2}{\sigma_x} + \frac{y^2}{\sigma_y})$.

4.4.2 Beam evolution

It is not in the scope of this thesis to review the entire kinetic theory of beams (for that refer to Riser [160]), but we do wish to briefly consider a distribution f, which evolves according to the collisionless Vlasov equation: $\partial_t f + \{f, H\} = \sum_i \partial_{qi} \dot{q}_i + q_e(\boldsymbol{E} = \boldsymbol{v} \times \boldsymbol{B})_i \partial_{Pi} f = 0$. If f can be written as a function of a conserved Hamiltonian H, then $\partial_t f = 0$ and we call the distribution "stationary". This would be nice, because in this case we would only have to calculate the space-charge field of a single distribution. If f were a uniform ellipsoid, then the space charge force would be linear and we would have a fully self-consistent stationary distribution.

In the microscope $H = \kappa_x(z)x^2 + \kappa_y(z)y^2$ is explicitly a function of z and so it is not conserved, but we can still find an invariant via a generalized canonical transformation [161]. This invariant is variously called the Courant-Snyder parameter, the Ermakov-Lewis invariant, or, most commonly, the beam emittance. It can be used to create a stationary distribution by starting particles in the so-called Kapchinsky-Vladimirsky (K-V) distribution lying on the surface of a hyper-ellipsoid of constant emittance. Ignoring any z dependence (i.e. a long beam of constant cross section) then the projection of the ellipsoidal shell into the $x-y$ sub-space is a uniformly filled ellipse with linear space charge. The K-V distribution is thus a self-consistent paraxial theory [160].

There is no generalization of the K-V distribution to 6D. Particles uniformly distributed on the 6D shell (a 5-sphere) will stay on the shell in the presence of linear forces, however the projection down to (x, y, z) (a 3-ball) does not create a uniform distribution (Riser problem 5.12 [160]). Most generally, this can be seen as a consequence of the Archimedean Projection Property [162] which allows a measure to be projected from an n sphere onto an $n-1$ dimensional ball. A solution does exist if we can imagine our beam has zero energy spread, but in this case we are entirely ignoring longitudinal dynamics. Other self-consistent 6D solutions are derived in [163] and a zero-emittance version is used for the simple theory of the photoinjector blowout regime [164]. In the end there the only difference between these distributions and a 2D K-V distribution is the definition of the generalized perveance[9].

In the K-V model the focusing force becomes $\kappa_x(z) = \kappa_{x0}(z) - \frac{2K}{\sigma_x(\sigma_x+\sigma_y)}$ and vice-versa for y. The generalized perveance, K, can be calculated from Eq. 4.7 as $K = \frac{2I}{I_0\beta^3\gamma^3}$ where $I_0 = 4\pi\epsilon_0 m_e c^3/q_e \approx 17\,\mathrm{kA}$ is the Alfven current. This is a standard linear transport problem and the envelope equation associated with the emittances ϵ_x and ϵ_y is:

$$\begin{aligned} \sigma_x'' + \kappa_{x0}\sigma_x - \frac{2K}{(\sigma_x+\sigma_y)} - \frac{\epsilon_x^2}{\sigma_x^3} = 0 \\ \sigma_y'' + \kappa_{y0}\sigma_y - \frac{2K}{(\sigma_x+\sigma_y)} - \frac{\epsilon_y^2}{\sigma_y^3} = 0 \end{aligned} \tag{4.8}$$

Once we have solved this (coupled) second order differential equation for known quadrupole strengths $\kappa_0(z)$ then we have solved the whole problem. We will have imaging ($M_{12} = M_{34} = 0$) when the betatron phase advance is π, i.e. $\int_0^{z_I} \frac{\epsilon_{\text{geo}}}{\sigma_{x,y}^2} = \pi$. And if $\sigma_x(z_I) = \sigma_y(z_I)$ then the imaging is stigmatic [10].

This provides an easy method to check if the transport between two planes works like a microscope, but it does not show us how to construct the solution. Without space-charge we were able find a solution because we could reduce the transport to the multiplication of pre-computed matrices (see Eq. 4.4) and then solve for the drift lengths which make $M_{21} = M_{3,4} = 0, M_{11} = M_{33}$. But now the transport matrix depends on $w(z)$ so that we have to re-solve the envelope equation every time we move a lens. Since M_{12} is smoothly varying and it is easy to show that a solution will exist (at least for sufficiently strong lenses), but in practice the easiest way to find the answer is to simply "guess and check" (i.e. numerically optimize).

---

[9] Alternatively see Riser 5.494 [160] for coupled transverse-longitudinal envelope equations (which do not, however, apply to a self-consistent beam distribution.

[10] The reader's intuition may be aided by remembering that the envelope and betatron phase come from assuming individual particles have the solutions $x = \sigma_x(s)e^{\pm i\psi_x(s)}$. Proper choice of the (necessarily constant) Wronskian gives the relation $\psi_x'(s) = 1/\sigma_x^2$. In the K-V picture all the particles lie on the same ellipse and are indexed by their phase (ψ_x and ψ_y). When the phases advance by π the particles are 'imaging' with a (negative) magnification given by the growth of σ during the transport.

4.4.3 Perturbation

Another way to approach the space-charge problem is to treat it as a small perturbation from the zero current case. In this case, we can calculate the first order space charge kick in x (and, by permutation of the subscripts, in y) as:

$$\delta_x''(s) = \frac{q_e}{\gamma m_e c^2} E_x(x(s), y(s), z(s), \sigma_x(s), \sigma_y(s), \sigma_z(s)) \tag{4.9}$$

Here E_x is the space charge force coming from any of the potentials in section 4.4.1 and it is evaluated assuming the coordinates (x, y, z) and the beam dimensions $(\sigma_x, \sigma_y, \sigma_z)$ are known from the zero-current matrix based transport. We choose boundary conditions $\delta_x(0) = \delta_x'(0) = 0$. Then the approximate position at the sample plane is $x(z_I) + \delta_x(z_I)$ (and the angle is $x'(z_I) + \delta_x'(z_I)$).

To the extent that the space charge force is linear, these perturbations can be incorporated in an error matrix [104] and then propagated piecewise. The way this works is that we take a transfer matrix $\boldsymbol{M}^{(0)}$ and use it calculate all the un-perturbed beam moments. Then we integrate Eq. 4.9 for two different linearly independent initial conditions, so that we can calculate the error matrix:

$$\begin{pmatrix} \delta_x \\ \delta_x' \end{pmatrix}_f = \boldsymbol{E} \begin{pmatrix} \delta_x & \delta_x' \end{pmatrix}_o \tag{4.10}$$

Then we update the transfer matrix as $\boldsymbol{M} = \boldsymbol{M}^{(0)} + \boldsymbol{E}$ to include the error terms. Usually there is some E_{12}, so that if $\boldsymbol{M}^{(0)}$ was imaging, then $\boldsymbol{M}$ is not, but we can make it zero by adding or subtracting some drifts to the end of the beamline. Of course, this isn't as easy as it sounds, because the perturbation will change every time we adjust the drifts. However, we can get close by ignoring this, and just multiplying $\boldsymbol{M}$ by a drift matrix (Eq. 4.4) and solving for the distance which eliminates the M_{12} element.

To make this concrete, let us put some numbers to the problem. Let us consider a realistic case, where triplet B (table 4.1) is imaging over a working distance of 82 cm. From the earlier discussion of principal planes, we know the effective focal length of the triplet is about 1.6 cm, so we can calculate that the magnification is about -50 and so (in units of cm) $\boldsymbol{M}^{(0)} \approx \left(\begin{smallmatrix} -50 & 0 \\ -1/1.6 & -1/50 \end{smallmatrix}\right)$. Plugging this into the equation for δ_x, we can find that (in

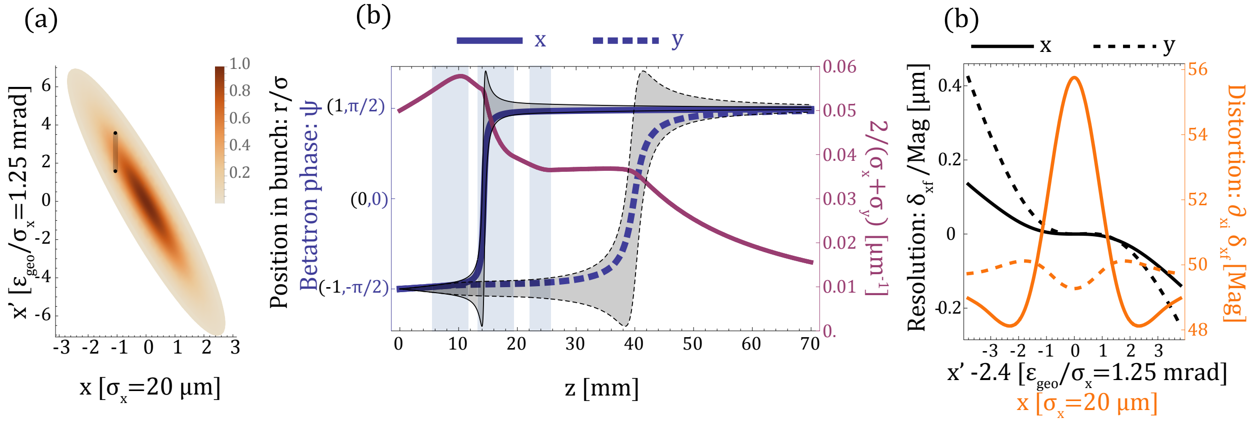

Figure 4.7: Space charge as an aberration. (a) The initial phase space of the (round) Gaussian beam being studied. (b) Rays from the black region in a are propagated through the first 70 mm of the 82 cm beamline. They receive a space charge kick which is proportional to their position in the bunch (black) and the average bunch size (purple). Their propagation closely tracks the Betatron phase advance (blue), but some rays cross the optical axes ahead or behind the main bunch. (c) Meridonal ray fan showing the "space-charge" aberrations. The black lines show the spherical aberration along the black line a, while the orange lines show the distortion along the major-axis of a.

units of cm) $\boldsymbol{E} \approx \left(\begin{smallmatrix} -6 & 6 \\ 1/10 & 1/12 \end{smallmatrix}\right)$. As might be expected, we can see that the error matrix is slightly de-focusing, but also that it actually increases the magnification, because most of the defocusing comes *after* the crossover. We can also see that it has some E_{12} that we could correct by increasing the object distance by roughly 1.2 mm. Of course moving the object will change E_{12}, but if we repeat this procedure a couple times then the corrections become very negligible, and we can solve for the imaging condition. In reality we also have to adjust the individual quadrupoles in the triplet as well, because the x and y corrections are slightly different.

The perturbation method really shines when considering the effect of nonlinear space-charge forces. In this case we can't characterize δ_x by a 2D matrix and so even an optimized triplet will have "space-charge aberrations". Nonetheless, it is still helpful to distinguish "cos-like" trajectories (which depend on the initial position) from "sin-like" trajectories (which depend on the initial angle). When "cos-like" trajectories are deformed the final image will appear distorted, but when "sin-like" trajectories are deformed the final image will be blurred. In Fig. 4.7 we illustrate the origin of theses aberrations by tracking a set of rays along the same 0.2 A beam discussed in the previous paragraph, but now with $\delta_{x,y}$ calculated assuming a (2D) Gaussian cross-section.

Aside from the generalized perveance, K, the space-charge kick of the elliptically Gaussian beam depends on three parameters: the particle's position relative to the bunch (x/σ_x), (y/σ_y), and the bunch eccentricity (σ_x/σ_y). The force can be tabulated for a reasonable set of these parameters and then applied to quickly integrate the perturbation along the beamline. For explanatory purposes, the three parameters we plot in Fig. 4.7(b) are the particle positions (the gray shading shows "sin-like" rays originating from the gray region of panel (a)) and the average beam size (the purple line shows $2/(\sigma_x + \sigma_y)$). The product of these two curves (multiplied by K) gives the kick on a K-V distribution, and also the linear portion of the Gaussian kick (for particles on the x-axis). Notice that the purple line starts increasing, because the beam is initially converging, but then falls immediately upon entering the first quadrupole despite the crossover in x. This happens because emittance pressure keeps σ_x from getting as small as σ_y gets large. After the whole triplet, both σ_x and σ_y are diverging and the kick decays inverse-linearly.

In the region near the quadrupoles, where the space-charge kick is highest, the particles undergo an important shift where they cross the optical axis and space-charge kick flips sign. But for most of the beamline, including the 75 cm from the end of Fig. 4.7(b) to the image plane, the flow is laminar and the "sin-like" trajectories (gray) are all bunched together. Thus all of the blurring comes from the region near the crossover where some of the particles have a faster phase advance than others. Those rays which cross the axis early end up displaced further from the origin, while those which cross late end up closer to the origin.

The net effect of the aberrations is summarized in Fig. 4.7 where the "sin-like" trajectories we have been following are colored black and a set of "cos-like" trajectories are colored orange (for rays taken from the major-axis of the ellipse in panel (a)). For the "sin-like" trajectories we see a cubic blurring, similar to spherical aberration, such that particles near the axis are-well imaged, but those far from the axis have significant blurring[11]. The blurring is worse for the y direction because the slower focus creates extended non-laminar flow. We can also see that the y curve is noticeably asymmetric because the particles which crossover first (the bottom ray in panel (b)) see a larger space-charge kick than the particles which crossover last (the top ray in panel (b)). The opposite is true for the distortion, where the x direction experiences more defocusing. This is the result of a balancing act: the distortion primarily comes from the laminar regions, but it faces opposite ways before and after the focus. In the y direction the kicks cancel, so there is little remaining distortion, but for the x direction the defocusing kick dominates.

Analyzing a 0.2 A beam like this is starting to push the perturbation technique to its limit. The kick is still small a very small percentage of x/σ_x (except in the immediate vicinity of $x = 0$), but it is large enough to begin to deform the phase space. This raises a bigger problem, which is that can't actually transport a 0.2 A Gaussian beam from the gun to the sample without distorting the phase space. Since the Gaussian is not a self-consistent distribution it will evolve and by the time it reaches the sample the initial phase space will already have these same "aberrations". Such a beam must be studied numerically, as we show later when discussing experimental results.

4.4.4 Simulation and Optimizations

In order to model space-charge in non-stationary distributions we use the particle tracking code general particle tracer (GPT)[165]. With GPT, we can initialize a large number of particles (5000 in our example) and then follow them through the PMQ fields from § 4.3. At each time-step, GPT meshes the charge distribution and then solves Poisson's equation (in the beam rest frame) in order to get the space-charge fields. Each simulation takes 1 min on a typical desktop (depending on the desired accuracy) and so it is possible to run large optimization routines on the input beam. Most importantly, we can shift the quadrupole positions in order to find the sharpest image.

---

[11] We have scaled the x-axis to units of $\epsilon_{\mathrm{geo}}/\sigma_x$ so that 1 unit corresponds to 1 standard deviation of $x'(x_0)$.

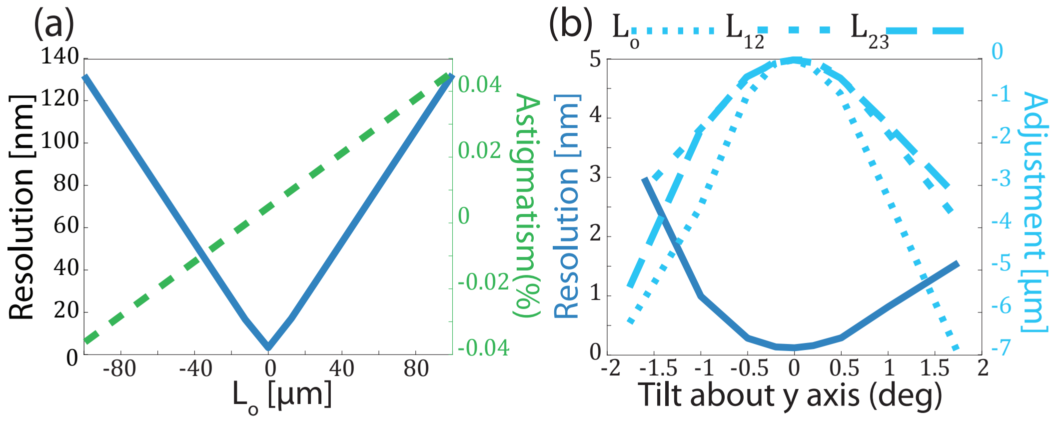

Figure 4.8: Detuning a quadrupole triplet (no space charge). (a) Moving the whole triplet. (b) Tilting the whole triplet (and re-optimizing the individual PMQ locations).

A typical optimizer works by trying to minimize a scalar cost function, f_c. To calculate the cost function, we output the particle positions at the sample plane and the image plane. We then define a basic cost function as the average error in the final particle position $f_c = (1/N)\sum_i \sqrt{(x_i(z_I)/m - x_i(z_o)^2 + (y_i(z_I)/m - y_i(z_o))^2}$. Since the magnification, M, is not known we fit a new M for each iteration by minimizing $f_c(m)$. The optimizer then varies the quad positions to look for the lowest f_c[12]. This cost function will work if the initial guess is good, but, if not, it is very inefficient. In practice, it is better to allow different magnifications m_x and m_y and then to add a penalty for large astigmatism. It is also good practice to limit the sum to only include particles near the middle of the bunch (since we care more about blurring than distortion), and, if the there are misaligned optics, to account for the centroid movement.

In the absence of space-charge this procedure will quickly align the quads to sub-nm (rms) resolution, demonstrating that the spherical and chromatic aberration of the quads should be tolerable for our input beam ($\epsilon_n = 20\,\text{nm}$, $\Delta\gamma/\gamma = 1 \times 10^{-4}$). Obtaining this kind of resolution, however, requires μm level positioning of the magnets, as we show in the detuning curves for Fig. 4.8. For example in panel (a) we show that moving the object distance L_o by 1 unit alters the resolution by a factor of $\sigma'_x \approx 1\,\text{mrad}$ (as expected from the

---

[12] Note that we allow the fringe fields to overlap without checking for demagnetization.

thin-lens approximation). We also show the astigmatism, which is related to a ratio of the magnifications by $(1 - m_x/m_y)$. With GPT, we can also analyze off-axis misalignment: for example, in panel (b) we tilt the triplet about the y-axis and then translate the triplet in x to remove the steering and then re-tune the triplet positions to optimize the resolution (the asymmetry is unphysical and is caused by the optimizer getting stuck in a local minimum).

With space charge the optimization is slower and more complicated since it strongly depends on the initial phase space of the beam. In Fig. 4.9 we illustrate this by comparing the imaging-error of two idealized distributions. The cubic shape is represents distortion and the thickness of the curves represents the blurring. Both beams have approximately uniform cross-sections in $x - y$ and $x' - y'$ [13], but the beam on the left has a uniform current distribution (20 mA 5 ps), while the one on the right is actually a uniform ellipsoid (i.e. football-shaped). Both beams have the same total charge, but the uniform ellipsoid has significantly better imaging. This shows the importance of controlling for slice-based space-charge effects.

The optimal phase-space distribution, and how to produce it, is still an open question. We know that the blurring is most sensitive to the region of the crossover, and so it is here that we are most concerned having linear forces. But near the crossover the bunch is rapidly switching between the near and far fields, so that short of actualizing a self-consistent distribution there is no perfect answer. What we do know is that an (un-truncated) Gaussian bunch causes intolerable aberration, while a uniform-uniform distribution can get down to the 10 nm level. We also know that the current profile needs to be constant to better than 2 mA in order to avoid slice-based aberration[14]. In addition to these concerns, going below

---

[13] These could be created either by operating the photo-injector in the blowout regime or by clipping a Gaussian beam to roughly 0.5σ. To clip in both position and angle requires two irises separated by a matrix proportional to $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

[14] Here is the short argument: Recall that for the 0.2 A beam the perturbation (correctly) predicted that we would need to move the triplet by about 1 mm (at 50x magnification). The kick is proportional the current, so if the current decreased by 2 mA we would have to shift the triplet by 10 µm or else suffer an $M_{12} = 500\,\mu\text{m}$ (which would cause 500 nm of blurring or 10 nm of resolution for a 1 mrad beam). If, instead of the whole beam decreasing its current by 2 mA, the beam gained a distribution of currents with rms strength 2 mA then the blurring of the total beam will be roughly the same. Thus both the shot-to-shot and slice-based current should be kept steady at better than 2 mA.

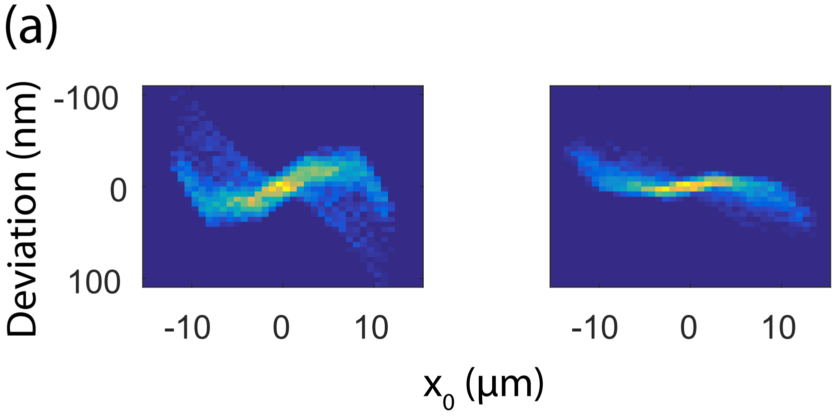

Figure 4.9: Ray fans for two different initial phase space distributions. Each plot shows the deviation of a ray from ideal imaging as a function of the object distance. (a) A truncated Gaussian distribution (at 0.5σ*) and (b) A uniformly filled ellipsoid of equivalent rms size. Both cases have* 3^{rd} *order aberration off-axis, but notably the uniformly filled ellipsoid has lower spread overall (in large part due to a lack of* z *correlation in the beam).*

10 nm in resolution will require grappling with stochastic (pairwise) Coulomb interactions, as studied by Li and Musumeci [144].

CHAPTER 5

Microscope operation

For the work in this thesis we tested some critical components of the microscope outlined in Fig. 4.1. Namely we measured beam brightness and we built a microscope with two PMQ triplets (but no electromagnetic quads and no x-band linearizer). The experience taught us about microscope design and let us study space charge effects, but due to shortcomings such excessive RF jitter and beamline misalignment, it did not help us find the resolution limits of either the microscope or of the PMQ triplets.

The resolution of our images was ultimately limited by two effects: when using one magnification stage (30x) we were limited by magnification/optical-resolution of the scintillator; and when using two stages (900x) we were limited by our ability to align the beamline, leading to poor contrast. Nonetheless, we review our techniques and measurements in order demonstrate the potential of PMQs to form a compact imaging stage.

5.1 Accelerator working point

Satisfying the transverse beam brightness requirements for the microscope is only possible if we stretch the uv pulse on the cathode to reduce the effect of space charge. In principle we have birefringent crystals we can use for "pulse stacking" [166, 74] to generate a parabolic 10 ps long pulse [152], but without an x-band cavity to reduce the energy spread this will not be particularly helpful, since the chromatic aberration would limit the resolution. Instead we use a prism stretcher to make the uv laser 1 ps (rms)[1]. Then we reduce the (rms) source

---

[1] The shortest electron beam we used was measured on an x-band transverse deflecting cavity to be 0.9 ± 0.15 ps

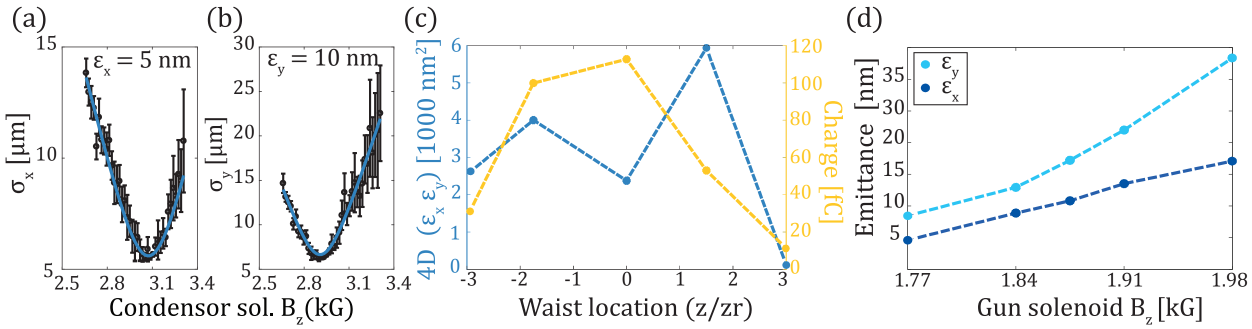

Figure 5.1: Emittance measurement at oblique incidence. (a,b) Solenoid scan at 20 fC showing record low emittance. (c) Normalized (rms) emittance and charge versus location of the uv focus (in units of rayleigh length). (d) Normalized (rms) emittance (at 20 fC) versus the gun solenoid strength.

size to 8 × 18 μm by focusing the uv with a high NA lens (f = 175 mm). The ellipticity is a consequence of the fact that the shortest path through the gun to the cathode is on a 72° oblique incidence port. Comparing the transverse and longitudinal beam sizes immediately after the cathode, we see that this beam has an aspect ratio of roughly 1. That means it is in neither the pancake nor cigar blowout regime and does not have a clear route to emittance preservation. Nonetheless, by controlling the strength of the gun solenoid we can transport the beam through a linac to the condenser solenoid and then measure a Pegasus-best (normalized, rms) emittance of 5 × 10 nm at 20 fC bunch charge.

The emittance measurements are made by scanning the current of the condenser solenoid and looking at the beam on an in-vacuum microscope objective (in order to resolve the few-μm waist). Fig. 5.1(a,b) shows a single scan and the fit based on a solenoid transfer map [167]. The noise in the data is correlated with RF amplitude jitter (and thus beam energy) and so the precision of the fit is somewhat greater than might initially be assumed. Fig. 5.1 also shows how we can align the position of the high NA lens (i.e. find the zero point of the x-axis of panel (c)) by monitoring the emittance (which depends on source size and charge), or the total emitted charge. The total charge increases at high intensity due to multi-photon photoemission (provided the current is not space-charge limited), while the emittance decreases with source size, but also with charge. Finally, in Fig. 5.1(d) we show

Q (fC)	ϵ_n (nm)	σ_r (μm)	$\sigma_{\beta r}$ (mrad)	α_{twiss}	σ_t (ps)	$\gamma m_e c^2$ (MeV)	$\sigma_\gamma m_e c^2$ (keV)
< 100	< 20	20	0.3	-0.02*	1	5(4)	10*

Table 5.1: Working point for UEM measurements as known at the sample plane. Numbers with an asterisk are known only through simulation. The beam energy was 4 MeV in [1] and 5 MeV (boosted by the linac) for the two-stage and space-charge measurements. The nominal beam parameters are listed for the 5 MeV case.

how the emittance can be reduced by tuning the current in the gun solenoid. At such low charge, simulations suggest there is little need for emittance compensation, so we attribute the decrease in emittance to either a skew quadrupole associated with the solenoid [168] or, more likely, to the fact that at 1.77 kG the beam focuses inside the linac (and thus avoids transverse fields from the linac). In practice, this constrains us to run experiments with the gun solenoid as low as feasible and thus reduces some of the flexibility implied by a two lens condenser.

Even 5×10 nm emittance, however, is not quite good enough for a microscope, and it was produced with only 7.5 mA peak current. At 36 mA, still shy of the 100 mA needed for the microscope, we measure an increased emittance of 12×18 nm. Hence the desire for a novel 1.4 cell gun: the short first cell allows for a launch phase close to the peak voltage, such that we can take advantage of the cigar regime (with a 10 ps bunch) scaling law $I \propto E_0^{3/2}$[2]. It is also likely that we will need to aperture the beam in order to reduce the emittance. In fact, we could avoid the oblique incidence configuration entirely and just aperture the beam after the gun as in the original proposal of Li and Musumeci [144] (provided we still launch a beam in the blowout regime to make a uniform-ellipsoid).

In practice, the brightness requirements are not so strict since we have been unable to test the microscope at 10 nm resolution. At higher spatial resolutions, between 100 nm and 1 μm, the brightness requirements are significantly relaxed and we have no need for special considerations. For this reason, all of the data shown in this thesis was gathered in the

---

[2] The current emittance measurements are only at 50 MV/m. The 1.4 cell gun is expected to produce upwards of 120 MV/m for a factor of 4 increase in peak current.

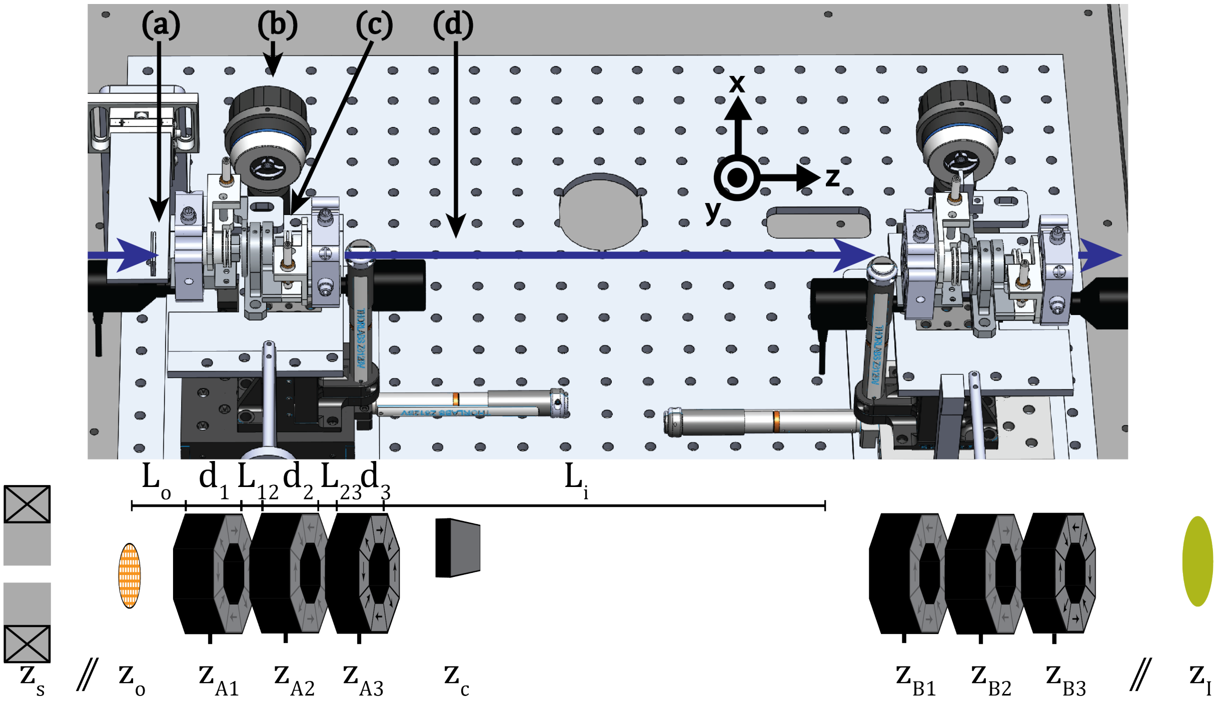

Figure 5.2: CAD drawing of the UEM installed at Pegasus. (a) The sample is mounted on a two axis stage (x-y). (b) An infinite conjugate microscope objective imaging a scintillator. (c) A PMQ triplet (reached by translating the whole stage in x relative to the scintillator). (d) The back focal plane (a collimator, not pictured, comes down from the roof of the chamber). The purple line shows the electron path through the two identical magnification stages. At bottom is cartoon labeling the location of all the elements (solenoid, object, triplet A, collimator, triplet B, image-plane).

simple oblique-incidence configuration listed in Table 5.1 (except where otherwise listed).

5.2 Mechanical design and alignment

The objective and projector stages we used are shown in the technical drawing of Fig. 5.2. The electron beam travels from left to right, first through the object then the objective triplet, the back focal plane (knife edge collimator not pictured), and finally the projector triplet. After leaving the vacuum box the beam reaches a final screen where it is imaged with an intensified camera (Princeton instruments PI-MAX 3).

Between two triplets, the sample, and the collimator there are 48 degrees of freedom available, of which 13 are motorized and the rest must be pre-aligned by hand. The 13 controllable axes are: the sample position (x and y), the z of each quadrupole, the triplet center (x and y) and the knife edge position (y). Importantly, the magnification stages can be moved roughly 2 inches in x to reach either the PMQ triplet, a low-resolution screen (for alignment)[3], a hole (to pass the beam through) or a thin (20µm) YAG:Ce scintillator. The thin scintillator is imaged by a nearly diffraction-limited microscope objective in order to characterize the beam (at the sample plane and at the intermediate image plane). At the final screen we use a thicker (100 µm) YAG to improve yield (at the cost of some optical resolution).

To align the non-motorized degrees of freedom we make extensive use of helium neon (HeNe) alignment lasers co-propagating alongside the electron beam[4]. For the best results we use both a forward propagating (red) HeNe and a backwards propagating (green) HeNe. Having two HeNes helps with the angular alignment of high aspect-ratio, opaque, elements (like the triplet) and for aligning the back reflection off of the sample and the collimator. In order for the HeNe lasers to be a precise reference we have to overlap the electrons onto the HeNe at two screens separated by only a drift (we use the screens on the first and second triplet stages). The condenser solenoid must lie exactly on this trajectory or else the electron beam will steer every time we adjust the illumination. The solenoid can be aligned to 0.1 mm, 1 mrad level by electron beam-based alignment (i.e. fitting the centroid position vs solenoid current using the same field maps from the solenoid scan emittance measurement), which is good enough to vary the illumination between between 7 µm and 500 µm with a centroid movement of only a few microns.

The knife-edge collimator is the weakest point in the microscope design. It is designed to descend form the roof of the box to clip particles which have a large y angle at the sample.

---

[3] For alignment it was also helpful to attach a screen around the aperture of the PMQs.

[4] Most of the 48 degrees of freedom come from the individual quadrupoles in each the triplet. These DOF were pre-aligned by Radiabeam using a miniature Hall probe with a 0.5 mm active area. Collaborators from ASU then slammed a sample into the triplet and, in a rush, we had to re-align the quadrupoles by eye.

z_s(mm)	z_o	L_o	z_{A2}	z_{A3}	z_c	z_{i1}	L_o	z_{A2}	z_{A3}	z_{i2}
-320	0	9	18	24	41	410	419	427	434	820

Table 5.2: Nominal position of the lenses in the two-stage measurements. All distances in mm. Note that this does not yield imaging in both planes due to confusion installing the quadrupoles (the L_{23} *distances could not be made long enough). Most distances are known to the* 500 µm *level, but* z_s *and* z_{i2} *are only known to within a few centimeters.*

In order to stop (or rather, sufficiently scatter), the knife edge needs to be thick (mm scale) and thus aligned perfectly normal to the beamline. An angular tilt of just 1 mrad will smear the collimator edge by 1 µm, which is significant since the entire separation at the back focal plane is only 14 µm (recall $f = 14\,\text{mm}$ and $\theta_s = 1\,\text{mrad}$). Thus we choose the edge of a square mirror to be our knife edge and carefully align the angle based on a back-reflection from the HeNe. Unfortunately, the collimator could not be easily moved in z and so we ended up placing it outside of the back focal plane (see Fig. 5.7), making it much less effective. For improving contrast it would also be much more appropriate to have a slit in place of a knife edge.

With all the ancillary alignment out of the way, we are left the most important alignment: the quadrupole locations along the optical axis. Nominal locations are listed in Table 5.2 and then motors allow us to move the whole triplet can be moved by ±4 mm and the interquad spacings by ±1.5 mm. The polarity of the triplets is such that in the x direction the first quadrupole focusing, defocusing, focusing (FDF). Due to miss-communication with the quadrupole manufacturer (Radiabeam Technologies), the second quadrupole is also FDF, although aberrations would be reduced if it were rotated to be DFD. Both triplets were also installed such that the third (short) magnet cannot be moved far enough away from the other two to make stigmatic images. Nonetheless, we can still study the optics by examining the x and y planes separately.

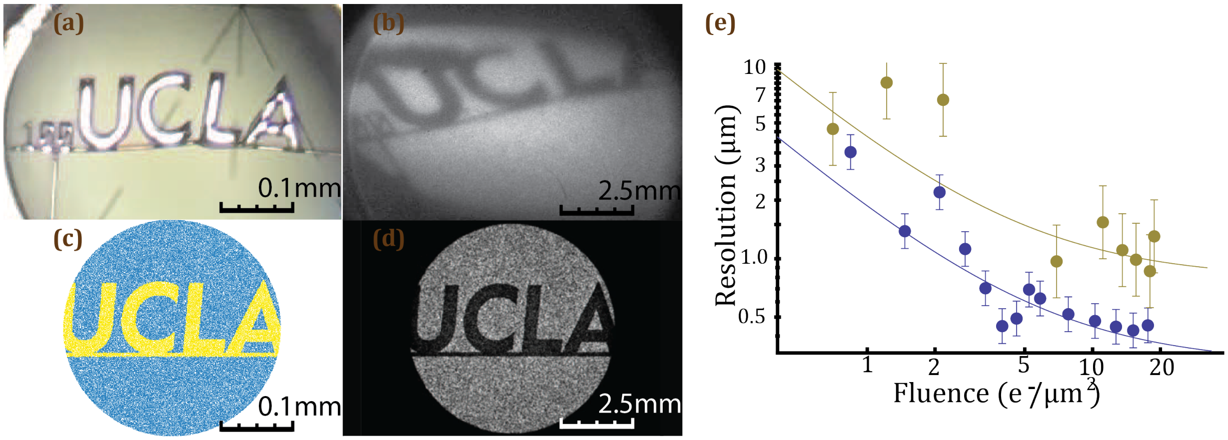

Figure 5.3: -30x imaging with a PMQ magnification stage (a) Nanofabricated sample. (b) Single-shot (1 ps) image of the sample. (c) Simulation (initial distribution). (d) Simulation (final distribution). (e) Resolution versus incident electron fluence for simulated (blue) and measured (gold) data. The error bars are due to shot-noise and the solid lines show the $1/N$ *scaling.*

5.3 Single stage results

In [1] we use a PMQ triplet as an objective lens to image a nanofabricated ‘UCLA’ mask (gold, 20µm thick), shown in Fig. 5.3(a-d) along with a piece of dirt which attached itself to the sample during installation. The magnification in this image is -32×-25 (in $x \times y$) with a astigmatism caused by the incorrect position of the last quadrupole (as mentioned in the previous section). Our results are the first to use a PMQ as a high gradient lens for MeV electron microscopy.

The spatial resolution of the ‘UCLA’ image in panel (b) is 1 µm (rms), hardly better than the optical-image in part (a). This has nothing to do with the electron optics, it’s just that the resolution of this image is limited (in order) by the magnification, the contrast, the shot-noise, and lastly the electron imaging. The magnification is most important because our detector system has a large point-spread function which blurs any edges which were otherwise in the beam at the detector. We can force the contrast and shot-noise to play a bigger role by lowering the illumination on the sample until the the “UCLA” begins to disappear. We can see this effect in Fig. 5.3(e) where we have defined the resolution as our ability to locate the edge of the "L" in the UCLA sample (to be precise, we fit a lineout near the L with an erf and define the resolution as the standard deviation of an ensemble of fits). As the illumination is lowered, the "L" becomes broader and the apparent resolution is worse. We also simulate this process (error bars indicate shot-to-shot variations), and in both simulations and experiment we see the expected $1/N$ behavior (when increasing the area A from § 4.2). The simulations assume an optical point-spread function of 20 µm but the measured data suggests a 50 µm psf. 50 µm is far from diffraction limited and needs to be improved in the final microscope design (primarily by switching from a 100 µm thick YAG:Ce scintillator to a thinner, 20 µm thick scintillator).

Given the difficulty of measuring the electron-contribution to the resolution, we have devised a new technique for measuring the amount of defocus. Normally we would imagine aligning the lens by moving the quadrupoles and looking for the sharpest image, but, as we have seen, the resolution of single-stage images is often limited by factors other than the M_{12} of the transport (e.g. contrast, shot-noise, and magnification). Thus we use another trick: we vary the condenser strength to sequentially illuminate a sample with electron beams which are converging ($\alpha > 0$), focused ($\alpha = 0$), and then diverging ($\alpha < 0$). When the major axis of the phase space (an eigenvector the beam matrix) falls along the line $x'_i = ax_i$ then the apparent magnification of a point object starting at x_i will be $x_f = (M_{11} + M_{12}a)x_i$. Only when $M_{12} = 0$ (imaging) will the apparent magnification be independent of the condenser solenoid.

This alignment of the objective is shown in Fig. 5.4(a). To quantify the magnification we image TEM2000 grids (12.5 µm pitch) and measure the spacing of grid-points at the detector plane. In contrast to the 'UCLA' images the magnification here is -60×-33 (in $x \times y$) since our working distance is 82 cm instead of 41 cm (our beam energy is also 5 MeV instead of 4 MeV). The imaging is again astigmatic, so we look only at grid-points along the x direction and we only vary the whole triplet location (L_o). In this plot the diverging beam slopes upwards and the converging beam points downwards and the three lines cross near an 8 ± 1 mm object distance. This is in agreement with particle tracking simulations (which suggest a roughly 9.5 mm object distance) given that the zero-point of the x-axis (i.e. the

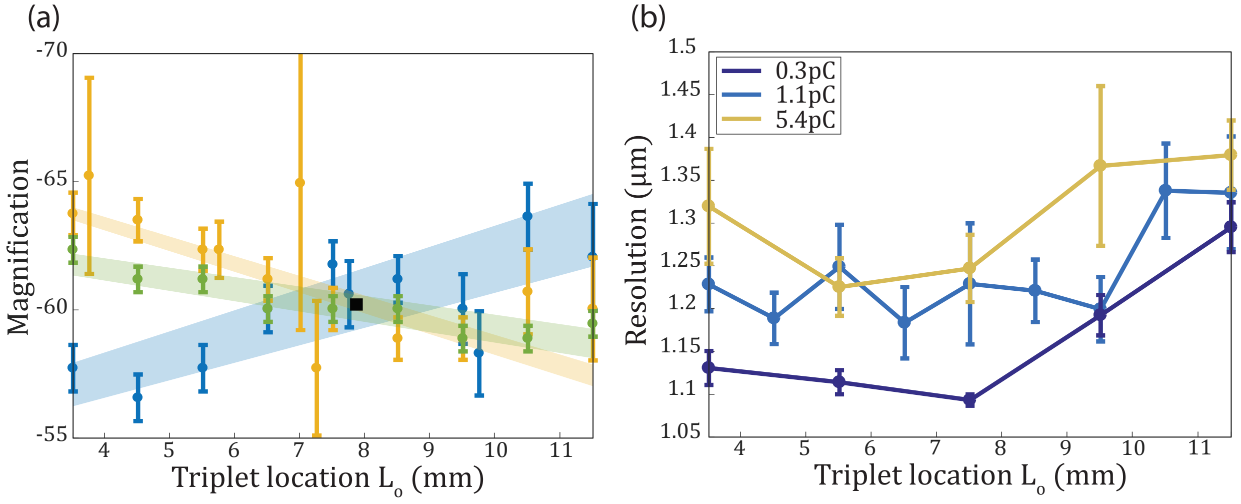

Figure 5.4: Tuning the imaging condition in x. (a) The apparent magnification as a function of triplet position for overfocused (blue) and underfocused (yellow,green) initial phase spaces (using 0.3pC bunch charge). (b) The measured resolution (rms) for 3 different bunch charges.

closest the PMQs can get to the sample) has to be measured by hand. If the quadrupole were stigmatic the same measurement could be made along the y-axis until all 6 lines matched. However, the scope this technique is limited to low charge. When space-charge is included the 3 beams no-longer have the same image plane (in fact, their image planes can diverge by a millimeters or more). In this case we could align the quadrupole at low current and then adjust it for higher current based on a numerical model.

We can also try to measure the resolution of the beam directly, as in Fig. 5.4(b). Here we fit erf functions to the edges of the grid-points and then quantify the resolution as the rms of the (assumed) Gaussian point spread function. At low charge this measurement agrees with panel (a) and has a few mm resolution. At higher charge, the resolution measurement is much noisier and the image plane is hard to identify. This happens in part because of space-charge within the column (see § 4.4.3), but mainly because at high current the phase space of the beam becomes highly distorted[5].

---

[5] The current corresponding to the three charges in Fig. 5.4(b) is not precisely known. At low charge the bunch length is 1.5 ps rms, but for 5 pC bunches we should expect significant longitudinal broadening in the Gun. For the 1 pC case, the GPT simulation for Fig. 5.5 suggests the expansion is not so large however, so that the peak current is 200 mA.

The non-linear phase-space of a high current (initially Gaussian) beam leads to the barrel and pincushion distortions seen in Fig. 5.5. The barrel converts to a pincushion as the triplet is moved further away from the object and the M_{12} of the transport changes sign from negative to positive. This is essentially the same effect studied at low charge in Fig. 5.4(a), but now the phase space has a nonlinear correlation developed during propagation from the gun to the sample. If the transport matrix were well-characterized then we could use these images to reconstruct the beam phase space (the so-called "grid-lens" technique of emittance reconstruction [169]), but since it is not we merely look for consistency with simulation.

To capture the whole effect we need a start-to-end simulation from the gun through to the end of the microscope. To match experimental conditions we use 1 pC, 1.5 ps bunch with a Gaussian 65 µm spot size[6] and 0.4 eV thermal kinetic energy. The gun accelerates the electrons to $\gamma = 6.3$ and then linac boosts them to $\gamma = 9.97$ (the gun solenoid focuses into the linac). By the time the beam approaches the focus of the condenser solenoid, its phase space (Fig. 5.5(a)) hardly resembles a Gaussian. However, if we look at the colored particles (which comes from the central row) we can understand what has happened: particles at the tail of the bunch (yellow) focus faster than particles at the head (blue) and so the grid has a $x \to z \to x'$ correlation. The hourglass shape of the entire phase space is a vestige of the initially Gaussian distribution (on axis the forces are linear), and this reappears in panel (b) where we have reconstructed the $x - x'$ phase space from a pepper-pot image (For comparison purposes the pepper-pot image has by propagated through a drift to have the same Twiss α as the simulation).

Although the initial phase space is perfectly round, the grids appear more strongly distorted in x than in y. This is merely a product of the 2x ratio in their magnifications: the M_{12} element of a defocused triplet is approximately $M_{11}\Delta L_o$, where ΔL_o is the displacement of objective lens from imaging. Thus the x direction gets twice as defocused as the y direction. We can also only see the barrel and pincushion distortions when ΔL_o is large, which is not surprising considering that the maximum x' (from simulation) is still only 1.8 mrad.

---

[6] To get run Pegasus at high current we used a large spot size on the cathode (at normal incidence).

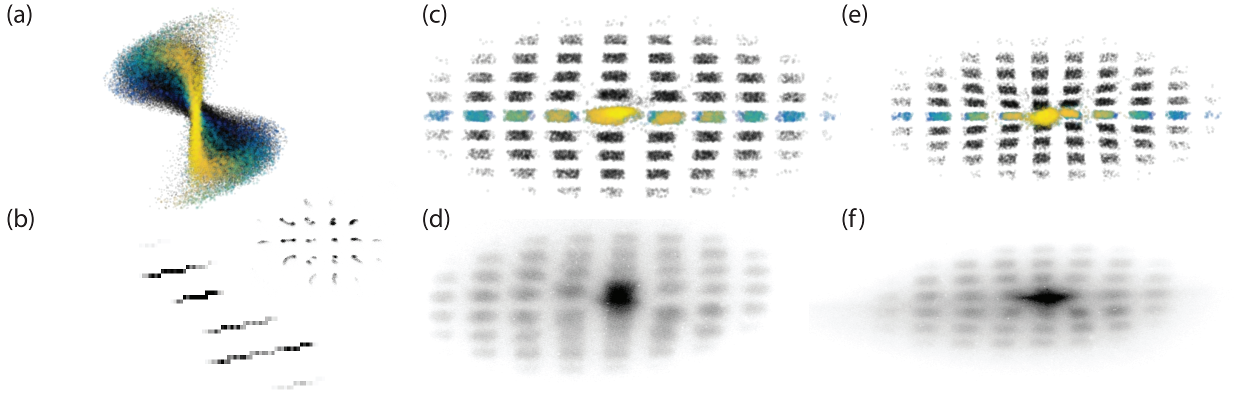

Figure 5.5: Distortion of images due to nonlinear phase-space correlations. The top row is simulation and the bottom row is measured. (a,b) Transverse phase space (as usual, x is on the abscissa and x' is on the ordinate). The data in b is reconstructed from a pepper pot (shown in the inset). (c,d) Images of a TEM2000 grid with L_o=4.5 mm. (e,f) Images of a TEM2000 grid with L_o=11 mm. In the simulations the central row is color-coded by z (yellow is the tail of the bunch and blue is the head).

Nonetheless, like the technique in Fig. 5.4(a), we can use the grid curvature to quickly identify the image plane (by looking for the L_o in which there is minimal distortion). At high charge we expect some additional barrel-type distortion in x (but not y, see § 4.4.3), although at 0.2 A the effect would still be only 1/5 the distortion seen here. The slice (z-dependent) kick we see here illustrates the importance of generating a constant current beam.

5.4 Two stage experiments

Adding a second, identical, triplet to serve as a projector lens should square the magnification. And in Fig. 5.6 we can see a copper TEM2000 grid (12.5µm pitch) imaged through a single stage and then through two stages for a total of 900 × 420 magnification. For this comparison, both images come from the same detector, so in (a) the objective is set to image over an 82 cm working distance and has -60×-33 magnification, while in (b) the objective and projector each image over 41 cm and each have roughly -30×-20 magnification. As before, the astigmatism comes from the distance to the third quadrupole in each triplet.

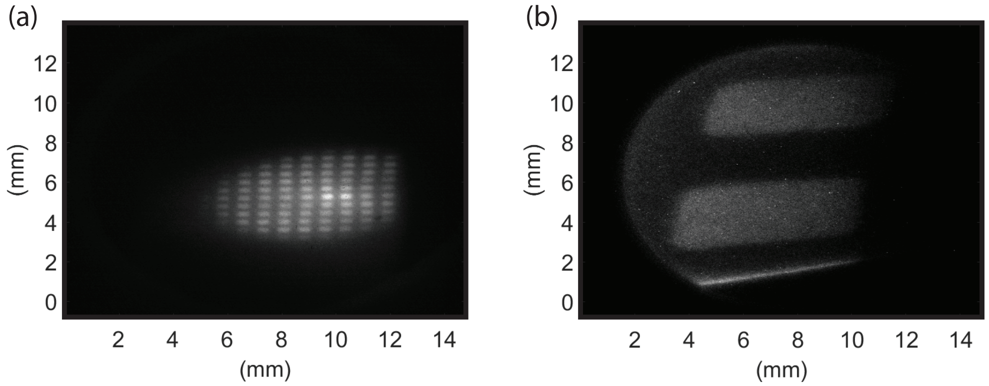

Figure 5.6: Two triplets increases the magnification. (a) Image of a TEM 2000 grid (12.5 µm pitch) through one triplet (the right-most part of the beam is clipped by the second triplet stage). (b) The same TEM grid imaged through both triplets with total magnification of 420×922×. At the bottom of the image is a caustic (i.e. not clipping).

With higher magnification, our resolution is no longer limited by the point spread function of the detector, and yet the best resolution we could measure was 400 nm fwhm (in either x or y, but not at the same time). This is clearly an improvement from the single stage resolution, but it is far from being useful. One culprit of the poor resolution is a misalignment of the individual quadrupoles inside each triplet. We know that the electron beam does not follow the HeNe through the triplets, but rather it gets a large dipole kick (especially in y for triplet 'A'), such that we have to offset the beam by more than half a gap in order to steer the beam back to our detector. As we offset triplet 'A' in y the beam actually steers up (off screen) and then reverses direction and comes back down (as if $y_f(y_i)$ were parabolic instead of linear). This nonlinear behavior leads to a caustic, like the one seen at the bottom of panel (b), and can fold parts of the image back on top of itself.

The blurring caused by excessive misalignment is then exacerbated by poor contrast. When imaging the TEM grids, the contrast we do get comes primarily comes from multiple scattering (at the 100 mrad level) with the PMQs serving as an aperture stop (see § 4.1). This is more effective with two stages than with one, but in both panels (a) and (b) there is still a noticeable (though uniform) background between the grid-points. As in Fig. 5.3(e), the poor contrast hurts our ability to resolve the edge. Moreover, the contrast gets worse at

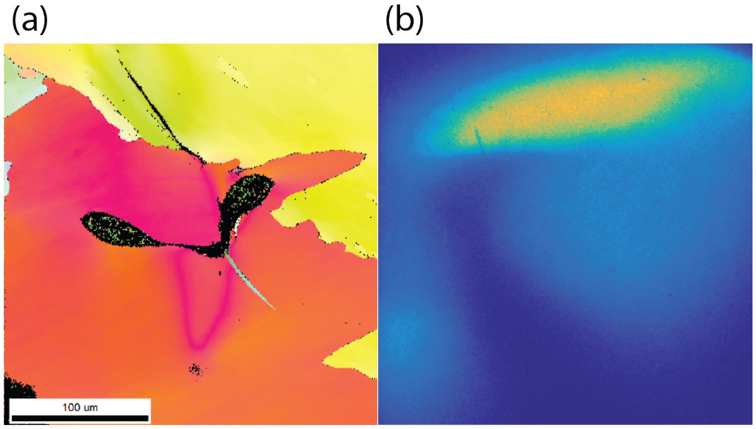

Figure 5.7: Diffraction contrast. (a) EBSD map of a Ti:Al sample (colors indicate crystallographic domains, black is an alignment hole). UEM image of the alignment hole showing simultaneous bright and dark-field imaging.

the edge of a TEM grid: a 1° rotation of the grid spreads a 20μm thick edge over 350 nm, and since the grids are not very reflective it is difficult to pre-align them better than this. This rotation likely contributes to the lack of resolution.

In principle, we should be able to improve contrast by using the knife-edge collimator to shrink the aperture stop. Unfortunately, we failed to place the collimator in the back focal plane. A quick proof of this is shown in Fig. 5.7 where we show images of a Ti:Al sample provided by Lawrence Berkeley labs. The left half of the image shows an electron backscatter diffraction (EBSD) map of the sample with color coding indicating the crystal lattice of various grains. The black part of the sample is a hole, meant for alignment, which we image with our UEM in panel (b) (using just the objective lens). It is immediately obvious that this image is simultaneously in the bright and dark field which demonstrates that the aperture stop is located at a plane where y and y' are coupled. Despite this, we could still get some small amount of contrast and make images of the grain boundaries.

To improve the contrast (and hence resolution) in these images we need to add new capabilities to the microscope. Most importantly, we need to be able to align the sample and then to place collimator slits exactly at the back focal plane. In order make scientific use of the collimator we also need to be able to image the back focal plane and see which part of the diffraction pattern the collimator is selecting. For this we imagine adding an electromagnetic triplet (Fig. 4.1) which can be turned on to project the diffraction pattern onto the detector. When imaging the sample the electromagnetic quad would be turned off, except for small currents which could be used to help fine-tune the optics and thus reduce the astigmatism.

APPENDIX A

Symmetries of a grating-style DLA

Here we aim to make a more rigorous derivation of the TM modes in a DLA. These DLA modes have been derived many times before (e.g. [99, 4, 170] among others), but our analysis is based on the diffraction grating analogy. Following the treatment Joanoppolus developed for photonic crystals [171], we are able to identify symmetry properties of solutions to the (Fourier transformed) wave equation. To do this, it is useful to write the wave equation as an eigenvalue problem:

$$\hat{\Theta}\boldsymbol{H}(\boldsymbol{r}) = \nabla \times \left(\frac{1}{\epsilon(r)}\nabla \times \boldsymbol{H}(\boldsymbol{r})\right) = \left(\frac{\omega}{c}\right)^2 \boldsymbol{H}(\boldsymbol{r}) \tag{A.1}$$

It is numerically convenient to solve directly for the magnetic field and then calculate $\boldsymbol{E}(\boldsymbol{r}) = (i/\omega\epsilon_0\epsilon(\boldsymbol{r}))\nabla \times \boldsymbol{H}(\boldsymbol{r})$ since this both guarantees $\nabla \cdot \boldsymbol{E}(\boldsymbol{r}) = 0$ and states the wave equation as a Hermitian eigenvalue problem with eigenvalue $(\omega/c)^2$ (provided the fields are either periodic or go to zero at large distances).

A.1 Bloch's theorem

The most important symmetry in Fig. 2.1 is the z periodicity of the dielectric. For propagating waves this leads to the grating equation, and our goal is to show that it does the same thing for evanescent waves. To do this we will derive Bloch's theorem and show that it allows us to decompose the fields into a set of plane waves separated by multiples of the grating wavenumber, k_g.

Since we expect the grating periodicity will somehow appear in the field we start by examining the translation operator $\hat{T}(\boldsymbol{d})$ which takes $\boldsymbol{r}$ to $\boldsymbol{r} + \boldsymbol{d}$. If we translate one grating period then $\hat{T}(d_g\hat{z})\epsilon(\boldsymbol{r}) = \epsilon(\boldsymbol{r})$ and thus $\hat{T}$ commutes with the wave operator $\hat{\Theta}$: $[\hat{T}(d_g\hat{z}), \hat{\Theta}] =$

0. Thus any field $\boldsymbol{H}(\boldsymbol{r})$ satisfying Eq. A.1 is simultaneously an eigenfunction of $\hat{T}(d_g\hat{z})$. And we can show that eigenvalues λ of $\hat{T}$ must have modulus one or else translating n grating periods would cause $\boldsymbol{H}$ to blow up: $\hat{T}(nd_g\hat{z}) = \hat{T}(d_g\hat{z})^n = \lambda^n\hat{T}(d_g\hat{z})$. Thus we can write $\lambda = e^{i2\pi\alpha}$.

Now we are in a position to show Bloch's theorem, namely that if $\boldsymbol{H}$ is an eigenfunction of $\hat{T}(d_g\hat{z})$ then it can be written in the form:

$$\boldsymbol{H}(\boldsymbol{r}) = e^{ik_{\mathrm{B}}z}\boldsymbol{u}(\boldsymbol{r}) \tag{A.2}$$

where k_{B} is called the Bloch wavenumber and $\boldsymbol{u}(\boldsymbol{r})$ is a periodic function ($\boldsymbol{u}(\boldsymbol{r}+d_g\hat{z}) = \boldsymbol{u}(\boldsymbol{r})$). To prove Bloch's theorem consider the equality $\hat{T}(d_g\hat{z})\boldsymbol{H}(\boldsymbol{r}) = \lambda\boldsymbol{H}(\boldsymbol{r}) = \boldsymbol{H}(\boldsymbol{r}+d_g\hat{z})$. If we plug the Bloch form into this equation then we get $\lambda e^{ik_{\mathrm{B}}\boldsymbol{z}}\boldsymbol{u}(\boldsymbol{r}) = e^{ik_{\mathrm{B}}(z+d_g)}\boldsymbol{u}(\boldsymbol{r}+\boldsymbol{d_g})$. For this to be true we just need to identify that $\lambda = e^{ik_{\mathrm{B}}d_g}$ and that u is periodic.

The importance of the Bloch theorem becomes more clear if we expand u in a Fourier series: $\boldsymbol{H}(\boldsymbol{r}) = e^{i(k_{\mathrm{B}})z}\sum_m c_j e^{imk_g z}$, where k_{B} is the Bloch wavenumber. We can then interpret each term in the sum as one of the diffraction orders in the grating equation Eq. 2.1. Moreover, we know that k_{B} should be equal to $k_z^{(i)} = k_0\sin(\theta_i)$ if this wave is to match a plane-wave at the boundary of the grating region. Thus Bloch's theorem shows us how the grating converts an incoming plane wave into a set of grating modes.

A.2 Reflection symmetries and the TE/TM distinction

As with translations, we can describe reflections by an operator $\hat{M}_r$ which takes $\boldsymbol{r} \to -\boldsymbol{r}$, that is an operator which inverts the coordinates $(x, y, z) \to (-x, -y, -z)$. Applying $\hat{M}$ twice returns us to the original system so that $\hat{M}$ has eigenvalues of either +1 or -1. We call these eigenvalues even and odd respectively. Often we are interested in a reflection about only one direction ($\hat{M}_x$), which can be described as by a spatial inversion of all coordinates ($\hat{M}_r$) followed by an 180° rotation about the x-axis. The spatial inversion, also known as the parity operation, deserves special consideration since Maxwell's equations are invariant under parity (provided the sources are reflected by the parity operation as well, c.f. Jackson

§6.10 [103]).

The transformation of electromagnetic quantities under parity is somewhat subtle and deserves elaboration. To start, consider the transformation of mechanical quantities under parity: a displacement vector $\boldsymbol{r}$ is odd, and so is a velocity $\boldsymbol{v} = \partial_t \boldsymbol{r}$, but angular momentum is even because $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ is a product of two odd quantities. The odd quantities (under parity) are known as vectors and the even quantities as pseudo-vectors (or axial-vectors). In electromagnetism $\boldsymbol{E}$ (and $\boldsymbol{D}$) are vectors while $\boldsymbol{B}$ (and $\boldsymbol{H}$) are pseudo-vectors. To see intuitively why $\boldsymbol{E}$ and $\boldsymbol{B}$ might be different, consider that $\boldsymbol{B}$ is formed from currents which switch both position and direction under parity, while the electric field is formed form charges which only switch positions. Or consider that a force has to transform like a displacement and so the Lorentz force law $\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$ suggests that $\boldsymbol{E}$ is odd and $\boldsymbol{B}$ is even.

By itself parity is not especially restrictive, but when combined with a symmetry in the dielectric ($\epsilon(\boldsymbol{r})$) it becomes more powerful because then it commutes with the wavefunction operator. In particular, if the structure (and drive laser) both look the same after the reflection then the eigenmodes can be labeled as either even or odd.

2D structures, such as our DLA, have reflection symmetry about the invariant axis (i.e. $\hat{x}$ for our planar DLA) and so the eigenmodes have to be either even or odd about that axis. Explicitly, we can consider $\hat{M}_x$ acting on an eigenmode $\boldsymbol{H}$: $\hat{M}_x$ consists of a parity operation, under which $\boldsymbol{H}$ is odd, and a 180° rotation about x which flips the sign of H_y and $H_z$back to positive. Thus if $\boldsymbol{H}$ is odd under $\hat{M}_x$ then it can only have H_x and if it is even under $\hat{M}_x$ then it can only have H_y and H_z. A similar logic applies to $\boldsymbol{E}$ and so the eigenmodes of our structure have either an $\hat{M}_x$ odd TM mode with (E_z, E_y, H_x) or an $\hat{M}_x$ even TE mode (H_z, H_y, E_x) (TM and TE stand for transverse magnetic and transverse electric respectively, where in accelerator physics the longitudinal coordinate is the z direction).

A.3 Free-space accelerating mode

The accelerating region of a DLA lays in free-space in the absence of sources (barring the negligible current of the electron beam). This greatly simplifies Maxwell's equations, and when combined with the fact that we are looking for a longitudinally polarized wave is enough to determine the analytic structure of the fields in a DLA.

Let us build our solution from plane-waves indexed by m as: $\boldsymbol{E_m} \cdot \hat{z} = E_0 E_m(x,y) \exp(i(k_m z - \omega t))$. The wave equation now takes on the simplified form $(\nabla_\perp^2 + (\omega^2/c^2 - k_m^2))\boldsymbol{E_m} = 0$. This is easily solved by restricting ourselves to a Cartesian representation ($\nabla_\perp^2 = \partial_x^2 + \partial_y^2$) and then summarily ignoring ∂_x because of the planar symmetry. That leaves us with a harmonic oscillator having y wavenumber $k_y^2 = (\omega^2/c^2 - k_z^2)$. The electric field becomes $E_m = (c_1 \exp(i(k_y y)) + c_2 \exp(-i(k_y y)))$ [1]. The two complex amplitudes c_1 and c_2 account for modes decaying down from the top teeth and up from the bottom grating teeth respectively, although it is simpler to write them centered at $y = 0$ as we have done.

Now since we started with $E_z \neq 0$ we have a TM mode with only (E_z, E_y, H_x). Then Gauss's law ($\nabla \cdot \boldsymbol{E} = 0$) tells us E_y and Faraday's law ($\nabla \times \boldsymbol{E} = i\omega \boldsymbol{B}$) gives us B_x. Since $\boldsymbol{B}$ only has an x-component and is independent of x by planar symmetry it is clearly divergenceless, which leaves as a last step to show that $\boldsymbol{B}$ calculated via Faraday's law is consistent with Ampere's law. Leaving this to the reader we list the resulting solution to Maxwell's equations for a TM (transverse magnetic) wave in planar Cartesian symmetry:

$$\begin{aligned}
\boldsymbol{E_m} \cdot \hat{z} &= iE_0 \left(c_{(1,m)} e^{-\Gamma_m y} + c_{(2,m)} e^{\Gamma_m y}\right) e^{i(k_m z - \omega t)} \\
\boldsymbol{E_m} \cdot \hat{y} &= E_0 \gamma_m \left(c_{(1,m)} e^{-\Gamma_m y} - c_{(2,m)} e^{\Gamma_m y}\right) e^{i(k_m z - \omega t)} \\
\boldsymbol{B_m} \cdot \hat{x} &= -\frac{E_0}{c} \beta_m \gamma_m \left(c_{(1,m)} e^{-\Gamma_m y} - c_{(2,m)} e^{\Gamma_m y}\right) e^{i(k_m z - \omega t)}
\end{aligned} \tag{A.3}$$

where the mode index m tells us the wave-number $k_m = k_0/\beta_m$ and damping $\Gamma_m = \sqrt{(k_m^2 - \omega^2/c^2)} = k_0/(\beta_m \gamma_m)$. The apparent increase of E_y and B_x with γ_m is offset by the smallness of the exponential since $c_1 - c_2 \approx 0$ for a symmetric mode. A similar expression can be obtained for the TE mode by starting from $B_z \neq 0$, yielding a mode with large

---

[1] If we had a structure with cylindrical symmetry then we would get $I_0(k_m r)$ in place of $\exp(-i(k_y y)$ [95].

E_x:

$$\begin{aligned}
\boldsymbol{B_m}\cdot\hat{z} &= iB_0\left(c_{(1,m)}e^{-\Gamma_m y}+c_{(2,m)}e^{\Gamma_m y}\right)e^{i(k_m z-\omega t)}\\
\boldsymbol{B_m}\cdot\hat{y} &= B_0\gamma_m\left(c_{(1,m)}e^{-\Gamma_m y}-c_{(2,m)}e^{\Gamma_m y}\right)e^{i(k_m z-\omega t)}\\
\boldsymbol{E_m}\cdot\hat{x} &= -cB_0\beta_m\gamma_m\left(c_{(1,m)}e^{-\Gamma_m y}-c_{(2,m)}e^{\Gamma_m y}\right)e^{i(k_m z-\omega t)}
\end{aligned} \tag{A.4}$$

It is worth pausing to summarize what we have done so far. By asserting translational and reflectional symmetries we were able to deduce that eigenmodes of Eq. A.1 can be split up and labeled as TE or TM for each k_z value described by the grating equation. Armed with this we were able to solve to solve Maxwell's equations in free space and list the solutions above. Actually that isn't entirely true, we needed one more assumption: namely that $\partial_x = 0$. This only applies if both the structure *and* the drive laser are invariant in x, otherwise there will be some set of k_x accompanying k_y and k_z. We can ignore the x dependence provided the laser profile in x changes slowly compared to the decay length $((\partial_x\mathcal{E}_z)/\mathcal{E}_z << k_0/(\beta\gamma))$.

A.4 Additional symmetries in DLA

In addition to the symmetries discussed in the proceeding sections, the DLA has a more complicated relationship to symmetry in 'y' (about the z-axis). As drawn in Fig. 2.1 the top and bottom teeth are aligned with each other, but this is not generally the case: the DLA is produced by bonding two separate gratings together to form the top and bottom halves and there is typically some misalignment of the teeth. Additionally we often have drive laser on only one side of the DLA and we only bring the electron beam from one side.

In cases where there is symmetry in 'y' (including drive lasers from top and bottom) we do get a nice simplification to the fields. Similar to how we found TE/TM modes we find that E_z, E_x, and H_y can be even under $\hat{M}_y$ while H_z, H_x, and E_y must be odd. In practice, this means that for the TM mode in Eq. 2.3 we must have $c_1 = c_2$ and we get the so-called cosh accelerating mode. If the reflection was anti-symmetric (the phase of the drive laser is different at top and bottom) then the structure would be in the sinh accelerating mode $(c_1 = -c_2)$.

More often we have no explicit 'y' symmetry, but we can still come up with some relationships between the field in various configurations. Consider two gratings offset from each other as in Fig. 2.8. Then the following operations are useful:

1. Reflections: Reflecting the whole system about z-axis ($\hat{M}_y$) or the about the y-axis ($\hat{M}_z$) means exchanging the sign of that $\boldsymbol{E}$-component (i.e. for $\hat{M}_y$ $\langle E_x(x, y, z)$, $E_y(x, y, z)$, $E_z(x, y, z)\rangle_{\text{reflected}} = \langle E_x(x, -y, z)$, -$E_y(x, -y, z)$, $E_z(x, -y, z)\rangle$).
2. To switch the direction of the drive laser: Reflect about the z-axis ($\hat{M}_y$) and then switch the direction of the tooth offset.
3. To switch the direction of the electron beam: Reflect about the y-axis ($\hat{M}_z$) and then switch the direction of the tooth offset.

Incidentally by looking at condition 2 we can easily see that using dual drive does not necessarily make a symmetric accelerating mode unless the teeth aligned. This means we have to be able to control the teeth alignment in a dual grating DLA. One way around this is to replace the second grating with a Bragg mirror which reflects the mode [101].

APPENDIX B

Nonlinear pulse propagation

While coupling a drive laser to our dual-grating DLA the laser passes through a large area of dielectric where we must take into account nonlinear effects including self-phase modulation, self-focusing and induced free carrier density [111, 112, 172]. In the main text we have discussed how self-phase modulation causes electrons to dephase from the drive laser and limit acceleration. In this appendix we will look in more detail at our analysis of nonlinear optical phenomena relevant to the DLA experiment.

Firstly we will show measurements of the nonlinear refractive of index of a DLA sample. Secondly, we will derive the NLSE in a rotated coordinate system in order to compactly describe a laser with a tilted pulse front. And finally we will show approximate calculations of photo-ionization based on the Keldysh model.

B.1 Z-scan measurement of n_2

Models of high-intensity pulse propagation in a dielectric are extremely sensitive to the value of n_2 used to calculate the nonlinear phase. Reliable values for n_2 have to be determined experimentally, but the reported for n_2 [107] have significant uncertainty and, moreover, are expected to be time-dependent since the underlying non-linearity can be caused by both Raman and electronic process [108]. To empirically ground our simulations we used our drive laser to independently measure the n_2 of an un-bonded DLA wafer (note that we observed no difference in n_2 when aligning the laser through a sub-wavelength grating or through bare wafer).

Measurements of n_2 using a high-intensity laser can be performed by using the well-

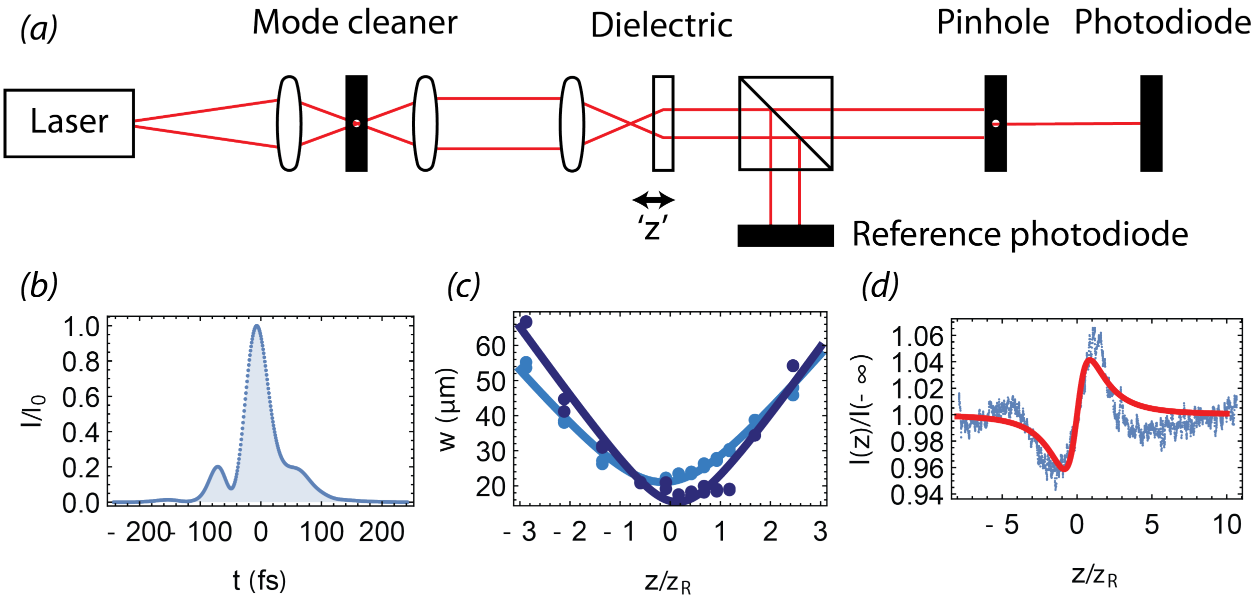

Figure B.1: (a) Layout for a Z-scan measurement of n_2: The power transmitted through the pinhole is measured as a function of the dielectric position, z, relative to a laser waist, for a well-characterized laser mode. (b) Temporal profile of the laser used to determine n_2 from $\langle n_2 \rangle$. (c) Waist scan with the dielectric sample removed showing that the mode is not perfectly TEM$_{00}$ (w_x is light blue, w_y is dark purple). (d) The power transmitted through the pinhole for I_0 =0.4 TW/cm^2 and L = 500 μm thick fused silica sample (blue dots are measured, red line is a fit using Eq. (B.1)).

characterized 'Z-scan' technique [173], which takes advantage of the lensing properties of a nonlinear dielectric to alter the far-field diameter of the laser. A layout of the scheme in Fig. B.1(a) shows the basic process: the TEM$_{00}$ component of a high-power laser is focused down near a dielectric sample; the sample then refracts the beam and thus, depending on the location of the sample relative to the laser waist, either increases or decreases the transmission through a pinhole. It can be intuitively understood that if the dielectric causes the beam to focus (positive n_2) then the transmission will be increased for positive z (when the dielectric is placed after the waist) since the dielectric helps to collimate the beam; while the opposite is true if n_2 is negative or if the dielectric is located before the waist.

In order to calculate the transmission through the pinhole as a function of z, I_0, n_2, k_0, and L (the sample thickness) it is necessary to know the spatio-temporal profile of the laser.

The role of the transverse profile is especially complicated since it effects both the nonlinear lens properties (through the intensity) and the forward propagation (through the wavefront). The conventional approach, then, is to simplify the problem and calculate the on-axis power transmission for a Gaussian (TEM_{00}) slice of a beam in the far field (many Rayleigh lengths past the dielectric) for small $\Delta\psi_{nl}$ [173]:

$$T(z, \Delta\psi_{nl}(t)) \approx 1 - \frac{4\Delta\psi_{nl}(t)\left(\frac{z}{z_R}\right)}{\left(\left(\frac{z}{z_R}\right)^2 + 9\right)\left(\left(\frac{z}{z_R}\right)^2 + 1\right)} \tag{B.1}$$

where z is the location of the dielectric relative to the waist and z_R is the Rayleigh range of the incident beam. Since our photodiode will measure the transmission integrated over the time profile of the beam, we need to average the transmission over time—an oft-forgotten step in the analysis of Z-scan measurements (e.g. [108]). In the small-perturbation limit the only time-dependent term is $\Delta\psi_{nl}(t)$ which can be replaced with an average weighted by the energy in each slice: $\langle\Delta\psi_{nl}\rangle = k_0 L \frac{\int_{-\infty}^{\infty} dt\, n_2 I^2}{\int_{-\infty}^{\infty} dt\, I}$. We assume that the nonlinear response is instantaneous so that n_2 is independent of time and we can pull it out of the integral. For a Gaussian this prescription yields $\langle\Delta\psi_{nl}\rangle = \Delta\psi_{nl}/\sqrt{2}$, while for the reconstructed FROG trace Fig. B.1(b) we find that $\langle\Delta\psi_{nl}\rangle = 0.56\Delta\psi_{nl}$.

Fitting Eq. B.1 to measured data shown in Fig. B.1(d) yields $\langle\Delta\Phi\rangle \approx 0.2$ from which we can estimate $n_2 \approx 2.3 \cdot 10^{-16}\,\mathrm{cm}^2/\mathrm{W}$. Estimating the precision of this measurement is difficult because the transmission is sensitive to the transverse beam quality, the surface roughness of the sample, and other nonlinear processes. We ruled out surface roughness and nonlinear absorption by recording the total transmission as a function of z at both low and high intensities (a so-called open iris Z-scan), and we showed that the B integral ($B = k_0 \int dz\, n_2 I$) from air is relatively small (compared to the dielectric). Despite that, the Z-scan cleraly isn't a perfect fit. The likely culprit is an asymmetric waist and non-Gaussian mode. The mode-cleaner is supposed to give us a TM_{00}, but from the waist scan in Fig. B.1(c) we can be pretty sure that this is not the case. The idea behind the mode-cleaner is to place a pinhole at a sharp waist where the diameter of each mode is $\approx \sqrt{n} w_0$ ([104] Eq. (17.5.44)), so that the higher order modes are clipped. Unfortunately for this measurement, we burned a hole in the pinholes and so additional mode content made it through. Nonetheless, all things considered, our measurement is consistent with reported values of n_2 for multi-cycle pulses and rules out any large (e.g. factor of 2) discrepancies in n_2.

B.2 Derivation of nonlinear pulse propagation

Pulse propagation in nonlinear media is a well-studied problem, drawing interest from the fiber community [111], the defense community (for directed energy weapons) [174], and in general the ultrafast laser and strong-field physics communities [175]. We provide a derivation in order to highlight the approximations that are made, and in particular how they apply in pulse front tilt coordinates.

This derivation closely follows that of [174]. We start from the scalar wave equation with a nonlinear polarization density:

$$\left(\nabla^2 - \frac{1}{c^2}\partial_t^2\right) E\left(\boldsymbol{r}, t\right) = \frac{1}{\epsilon_0 c^2}\partial_{t^2} P\left(\boldsymbol{r}, t\right) = \frac{1}{\epsilon_0 c^2}\partial_{t^2}\left(P_{\mathrm{l}}\left(\boldsymbol{r}, t\right) + P_{\mathrm{nl}}\left(\boldsymbol{r}, t\right)\right) \tag{B.2}$$

The scalar wave equation assumes that the wave is transverse and so we can treat $\boldsymbol{E}$ as a scalar, which is valid so long as laser field remains primarily transverse. The linear part of the polarization density is given by:

$$P_{\mathrm{l}}\left(\boldsymbol{r}, t\right) = \epsilon_0 \int_{-\infty}^{t} \chi(t - t') E\left(\boldsymbol{r}, t\right) \mathrm{d}t' \tag{B.3}$$

This convolution integral is easy to do in the Fourier domain, and by identifying $\epsilon(\omega) = 1 + \chi(\omega)$ we can simplify Eq.(B.2) as:

$$\left(\nabla^2 + \epsilon(\omega)\frac{\omega^2}{c^2}\right) E\left(\boldsymbol{r}, \omega\right) = \frac{\omega^2}{\epsilon_0 c^2} P_{\mathrm{nl}}\left(\boldsymbol{r}, \omega\right) \tag{B.4}$$

We can prepare to further simplify the problem by explicating writing the fields as an envelope times a plane wave:

$$E\left(\boldsymbol{r}, \omega\right) = \mathcal{E}\left(\boldsymbol{r}, \omega - \omega_0\right) e^{i(k_0 z)} + \mathrm{c.c} \tag{B.5}$$

$$P_{\mathrm{nl}}\left(\boldsymbol{r}, \omega\right) = \mathcal{P}\left(\boldsymbol{r}, \omega - \omega_0\right) e^{i(k_0 z)} + \mathrm{c.c} \tag{B.6}$$

where c.c. stands for complex conjugate, and which I will not continue to write explicitly (though it is implied). Next we identify the dispersion through the series expansion $\sqrt{\epsilon(\omega)}\frac{\omega}{c} = k(\omega) = \sum_{n=0}^{\infty}\frac{k^{(n)}(\omega_0)}{n!}\left(\omega-\omega_0\right)^n$, with k complex to account for dispersion. Finally splitting ∇ into longitudinal (z) and transverse compenents, we can re-write Eq.(B.2) in the ω and t domains as:

$$\left({\nabla_\perp}^2 + \left(\partial_z + ik_0\right)^2 + \left(\sum_{n=0}^{\infty}\frac{k^{(n)}\left(\omega_0\right)}{n!}\left(\omega-\omega_0\right)^n\right)^2\right)\mathcal{E}\left(\boldsymbol{r},\omega\right) = \frac{\omega^2}{\epsilon_0 c^2}\mathcal{P}_{\text{nl}}\left(\boldsymbol{r},\omega\right) \tag{B.7}$$

$$\left({\nabla_\perp}^2 + \left(\partial_z + ik_0\right)^2 + \left(\sum_{n=0}^{\infty}\frac{k^{(n)}\left(\omega_0\right)}{n!}\left(i\partial_t\right)^n\right)^2\right)\mathcal{E}\left(\boldsymbol{r},t\right) = \frac{\left(i\partial_t+\omega_0\right)^2}{\epsilon_0 c^2}\mathcal{P}_{\text{nl}}\left(\boldsymbol{r},t\right) \tag{B.8}$$

Let us single out the real numbers $k_0 = k(\omega_0)$ and $k_1 = k^(1)(\omega_0) = \partial_\omega(k(\omega))|_{\omega=\omega_0}$ (since these are related to the phase and group velocities respectively), while grouping the rest of the k terms into a higher-order dispersion operator:

$$i\hat{D}_s\left(\omega_0\right) = \text{Im}\left(k_0\right) + \text{Im}\left(k_1\right)i\partial_t + \sum_{n=2}^{\infty}\frac{k^{(n)}\left(\omega_0\right)}{n!} = i\frac{\alpha}{2} - \frac{\alpha_1}{2}\partial_t + \sum_{n=2}^{\infty}\frac{k^{(n)}\left(\omega_0\right)}{n!}\left(i\partial_t\right)^n \tag{B.9}$$

At the same time notice that ∂_z and ∂_t commute so that we can expand and factor these derivative like polynomials to re-write Eq.(B.8) as:

$$\left({\nabla_\perp}^2 + \left(\partial_z + 2ik_0 - k_1\partial_t - \hat{D}_s\right)\left(\partial_z + k_1\partial_t + \hat{D}_s\right)\right)\mathcal{E}\left(\boldsymbol{r},t\right) = \frac{\left(i\partial_t+\omega_0\right)^2}{\epsilon_0 c^2}\mathcal{P}_{\text{nl}}\left(\boldsymbol{r},t\right) \tag{B.10}$$

At this point we will boost to a frame of reference moving at the group velocity of the laser pulse. At the same time we will introduce a rotation (and scaling) in (x,τ) which allows us to remove any $x-\tau$ correlation from the description–a step which will be helpful when numerically gridding a tilted pulse front. Formally we introduce new coordinates as[1] [2]

$$\begin{pmatrix} p \\ y \\ \xi \\ \tau \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \frac{c}{\beta} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{1}{\beta} & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -k_1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \tag{B.11}$$

We use the inverse of this transform to substitute for all the derivatives (recalling that the partial derivatives commute and so work like polynomials):

$$\left(\left({\partial_y}^2 + \left(\partial_p - \frac{1}{\beta c}\partial_\tau\right)^2\right) + \left(\partial_\xi - 2k_1\left(\frac{c}{\beta}\partial_p + \partial_\tau\right) + 2ik_0 - \hat{D}_s\right)\left(\partial_\xi + \hat{D}_s\right)\right)\mathcal{E}\left(\boldsymbol{r}, t\right)$$

$$= \frac{{\omega_0}^2\left(\frac{i}{\omega_0}\left(\frac{c}{\beta}\partial_p + \partial_\tau\right) + 1\right)^2}{\epsilon_0 c^2}\mathcal{P}_{\mathrm{nl}}\left(\boldsymbol{r}, t\right) \tag{B.12}$$

Up until this point our only point of approximation is the original choice of the scalar wave equation. Now we will use a slowly varying envelope approximation, in which we assume the envelope functions $(\mathcal{E}, \mathcal{P}_{\mathrm{nl}})$ vary slowly on the optical (plane-wave) scale. In particular we will use the parabolic slowly varying envelope approximation:

$$|\partial_\xi^2| << |k_0 \partial_\xi| \tag{B.13}$$

$$|\left(\frac{c}{\beta}\partial_p + \partial_\tau\right)^2| << |\omega_0\left(\frac{c}{\beta}\partial_p + \partial_\tau\right)| \tag{B.14}$$

These conditions says that line-outs along the envelope must be large compared to the optical period.

In addition to the envelope approximation we will assume the dispersion is sufficiently small that the phase and group velocities are similar: $k_1 - k_0/\omega_0 \approx 0$ (for fused silica $k_1 - k_0/\omega_0 = 0.15/c$). Combining these approximations we are able to re-write Eq.(B.12) as:

---

[1] A more compact numerical representation could be obtained if the rotation in Eq.(B.11) were replaced with a skew-transform (exchanging the $\frac{c}{\beta}$ on the top row for a 1). The current formulation has the lesser advantage of remaining orthogonal (in the moving coordinate system and with units of $c = 1$), and at some point I overvalued orthogonality and so implemented this formulation into my numerical solver.

[2] Note also that in the numerical code I wrote these equations slightly differently so when using GAFFE one should enter $\beta_{\text{GAFFE}} = -1/\beta$.

$$\left(\left(\left(\frac{i}{\omega_0}\left(\frac{c}{\beta}\partial_p+\partial_\tau\right)+1\right)\right)^{-1}\left({\partial_y}^{2}+\left(\partial_p-\frac{1}{\beta \mathrm{c}}\partial_\tau\right)^{2}\right)+(2ik_0)\left(\partial_\xi+\hat{D}_s\right)\right)\mathcal{E}\left(\boldsymbol{r},t\right) \tag{B.15}$$

$$=\frac{{\omega_0}^2\left(\frac{i}{\omega_0}\left(\frac{c}{\beta}\partial_p+\partial_\tau\right)+1\right)}{\epsilon_0 c^2}\mathcal{P}_{\mathrm{nl}}\left(\vec{r},t\right) \tag{B.16}$$

Our final approximation will be:

$$\left(\left(\frac{i}{\omega_0}\left(\frac{c}{\beta}\partial_p+\partial_\tau\right)+1\right)\right)^{-1}\approx 1-\frac{i}{\omega_0}\left(\frac{c}{\beta}\partial_p+\partial_\tau\right) \tag{B.17}$$

which is valid so long as $\omega_0 >> \left(\frac{c}{\beta}\partial_p+\partial_\tau\right)$—a small extension to the slowly varying envelope approximation. Using this approximation and and grouping the transverse derivatives into a diffraction operator, we can solve for the propagator:

$$\hat{D}_f=\frac{1}{2k_0}\left(1-\frac{i}{\omega_0}\left(\frac{c}{\beta}\partial_p+\partial_\tau\right)\right)\left({\partial_y}^{2}+\left(\partial_p-\frac{1}{\beta \mathrm{c}}\partial_\tau\right)^{2}\right) \tag{B.18}$$

$$\left(\partial_\xi+\hat{D}_s+\hat{D}_f\right)\mathcal{E}\left(\boldsymbol{r},t\right)=\frac{{\omega_0}^2}{2ik_0\epsilon_0 c^2}\left(\frac{i}{\omega_0}\left(\frac{c}{\beta}\partial_p+\partial_\tau\right)+1\right)\mathcal{P}_{\mathrm{nl}}\left(\boldsymbol{r},t\right) \tag{B.19}$$

To finish the derivation we need to decide what physics to include in the nonlinear term. The simplest possible nonlinear term includes just the (instantaneous, local) Kerr phase:

$$\mathcal{P}_{\mathrm{nl}}=2n_0 n_2\epsilon_0|\mathcal{E}^2|\mathcal{E} \tag{B.20}$$

A more thorough approach takes into account the time dependence (still assuming locality, i.e. no r convolution):

$$P_{\mathrm{nl}}\left(\boldsymbol{r},t\right)=\epsilon_0\int_{-\infty}^{t}dt_1dt_2dt_3\chi^{(3)}(t-t_1,t-t_2,t-t_3)\vdots\mathcal{E}\left(\boldsymbol{r},t_1\right)\mathcal{E}\left(\boldsymbol{r},t_2\right)\mathcal{E}\left(\boldsymbol{r},t_3\right) \tag{B.21}$$

where the $\vdots$ is a way to indicate the tensor sum with i^{th} component $\sum\limits_{jkl}\chi^{(3)}_{ijkl}\mathcal{E}_j\mathcal{E}_k\mathcal{E}_l$. This sum simplifies to a single term since we only have one polarization to be concerned with (nor do we need to keep track which polarization since fused silica is amorphous). To continue it is necessary to find (empirically or otherwise) some form for $\chi^{(3)}$. A common choice to account for a delayed Raman response (molecular motion) is to choose:

$$\chi^{(3)}_{iiii}(t-t_1, t-t_2, t-t_3) = \chi^{(3)}_{iiii}\sum_j R(t-t_j)\delta(t-t_{j+1}) \tag{B.22}$$

and then to approximate $R(t-t_j) \approx 1 - T_R\partial_t$ since we care about limit of instantaneous response (the value of T_R is known from literature [111]). When we substitute the plane-wave expansion (Eq.(B.6)) we will find terms going like $e^{in\omega t}$ with $n \in (-3, -1, 1, 3)$, however we will only care about the term with $n = 1$ since this one will propagate with the wave and so have a chance to grow ($n = 3$ also propagates forward, but will dephase due to dispersion). Thus we get:

$$\mathcal{P}_{\text{nl}} = 2n_0n_2\epsilon_0\mathcal{E}\left(|\mathcal{E}^2| - T_R\partial_t|\mathcal{E}^2|\right) \tag{B.23}$$

We should note that these derivatives are taken with respect to t and should be transformed to the correct coordinate system before being used. As usual this will slightly modify the assumption: namely we have assumed a nearly instantaneous response, which in the pulse front tilted coordinates means that the laser pulse must be shorter than the Raman time scale ($> 60\,\text{fs}$) along a lineout in t. It is also worth pointing out that there is some discrepancy in the definitions of T_R and χ^3 between various sources [174, 111, 112], however these numbers are empirically determined and so long as the final result matches measurements of the form $n \approx n_0 + n_2I$ then the exact form is immaterial.

Since $\mathcal{P}_{\text{nl}}$ is proportional to $\mathcal{E}$ we can group these terms into a nonlinear operate $\hat{N}$. For compactness we rewrite the propagator Eq.(B.19) here in terms of the 3 operators $\hat{D}_s, \hat{D}_f, \hat{N}$ (dispersion, diffraction and nonlinear).

$$\frac{\partial\mathcal{E}}{\partial y} = \left[\hat{D}_f + \hat{D}_s + \hat{N}\right]\mathcal{E} \tag{B.24}$$

For compactness we write the operators here in flat-pulse coordinates (as we have demonstrated one can convert to p and τ directly by the coordinate transform from Eq.(B.11), with appropriate changes to the assumptions). Note that here λ refers to the *vacuum* wavelength, while k_0 and ω_0 were the plane wave expansion $\omega_0 = (c/n)k_0$:

$$\hat{D}_f = \frac{i\lambda}{4\pi n_0}\left(1 - \frac{i\lambda}{2\pi c}\partial_t\right)\left(\partial_x^2 + \partial_z^2\right) \tag{B.25}$$

$$\hat{D}_s = -\frac{i}{2}k^{(2)}\partial_t^2 - \frac{1}{2}\alpha + \dots \tag{B.26}$$

$$\begin{aligned}\hat{N} =& \frac{i2\pi n_2}{\lambda}\left|\mathcal{E}^2\right| - \frac{1}{\mathcal{E}}\left(\frac{n_2}{c}\right)\partial_t\left(\mathcal{E}\left|\mathcal{E}^2\right|\right)\\ &- \frac{i2\pi n_2}{\lambda}T_r\partial_t\left|\mathcal{E}^2\right| - \frac{\beta_6}{2}\left|\mathcal{E}^{10}\right|\end{aligned} \tag{B.27}$$

B.3 Numerical beam propagation

We can solve Eq.(B.24) numerically using a finite difference approach to $\partial_y\mathcal{E}$ in order to propagate the beam envelope. We start by defining the beam envelope at $y = 0$ as a matrix of values representing the complex-valued $\mathcal{E}(x, z, t)$ along a linear grid. We then use Eq.(B.24) to calculate the $\partial_y\mathcal{E}(x, z, t)$ and make a discrete jump to a subsequent plane $y = \Delta y$. The linear operators (diffraction and dispersion) can be easily solved in the Fourier domain [104], where derivatives become multiplication, while the nonlinear term is most easily represented in the time domain. Consequently we peruse a split-step method [111, 112] where we apply a half-step linear kick in the Fourier domain, followed by a full step non-linear kick in the time domain, and end with another half-step linear kick. We adjust the step-size to keep the error (calculated relative to a higher-order propagator) to a small value.

The main numerical challenge is controlling the mesh $\mathcal{E}(x, z, t)$ so that the beam is well-represented in both Fourier planes without occupying too much memory. Equivalently, we need a large enough grid to encompass the whole beam, but also enough grid points to resolve the envelope. To handle the meshing we use a Matlab framework called GAFFE (Generalised Adaptive Fast-Fourier Evolver) [113] which allows us to adjust the mesh as the beam propagates. For example, if the beam starts to focus we will want to magnify the grid or else the beam will overfill the mesh in the Fourier domain. Thus we try to propagate a step and if the beam overfills the grid in either domain then we try re-sizing the grid and propagating again. This allows to always use the smallest possible grid to describe the beam, although for self-focusing in three-dimensions this can still be many gigabytes.

The parameters we use when solving Eq.B.24 are in table B.1.

Table B.1: Table of model inputs

Parameter	Value	Notes
Material	SiO_2	
n_0	1.45	
n_2	$2.45{\times}10^{16}$ cm^2/W	[107]
$k^{(2)}$	36.163 fs^{2}m^{-1}m	
κ	$10{\times}10^{-6}$ 1/cm	
β_6	$5{\times}10^{-83}$ m^9/W^6	[112]
τ_r	3 fs	[111]
λ	800 nm	
w_x	45 µm	Gaussian
w_z	500 µm	Gaussian
τ	45 fs	FROG trace
Pulse Energy	$2 - 200$ µJ	

B.4 Photoionization

The generation of free carriers is neglected in Eq. (2.15) because post-hoc estimates suggest that the change in index of refraction n due to free carriers is a small fraction of the contribution due to self phase modulation. But near the damage threshold this is no longer true. We can see this by using a Keldysh model and then applying the Drude model to calculate the free carrier contribution to n.

The Keldysh model [176] describes the ionization of electrons from an s-orbital to free space in the presence of a strong (below work-function) electric field. At low frequencies this reduces to tunneling and at high frequencies it gives the multi-photon absorption. In order to apply the model to fused silica we us the parameters from [172] which have been chosen to match the propagation of 800 nm light through fused silica. For our case, the

Keldysh parameter is $\gamma \approx 14/E_0$ for E_0 in GV/m. And for all our experiments $\gamma > 1$ and often $\gamma >> 1$ such that we can mostly think of the ionization as coming from multi-photon absorption, albeit with a modified cross-section.

The Keldysh ionization is responsible for the initial generation of free electrons (i.e. conduction band electrons), but damage to the sample happens when the avalanche effect takes over. Again using parameters from [172] we can calculate the total free electron density as a function of time including both avalanche and recombination terms. Because the Keldysh effect doesn't seed the avalanche until the peak of the laser intensity, we find that the free carriers lag the main the pulse by about 20 fs.

Once we know the free carrier density we can calculate the change to the refractive index by including a Drude term: $n = \sqrt{(n_0 + n_2 I)^2 - \left(\frac{\omega_p}{\omega_0}\right)^2 \frac{2}{1+i/(\omega_0 \tau_c)}}$, where ω_p is the plasma frequency of the free electrons, ω_0 is the laser frequency, and τ_c is the collision time. Even small changes in n can add up since, like the Kerr effect, the phase change builds across the 500μm of the fused silica wafer.

The importance of the photonionization term can be characterized by comparing it to the Kerr term. At an incident electric field of 5 GV/m, where the Kerr effect causes a π phase change, the free electron term causes a negligible phase change of 0.7 mrad. But the multi-photon-photoionization is highly nonlinear, and by the time $E_0 = 10\,\text{GV/m}$ the free electron term is a full 20% of the Kerr term. For such high intensities the photionization should be included in the numerical beam propagation, however doing so would be computationally expensive and would only change the results near the damage threshold.

REFERENCES

[1] D. Cesar, J. Maxson, P. Musumeci, Y. Sun, J. Harrison, P. Frigola, F. O'Shea, H. To, D. Alesini, and R. Li, "Demonstration of Single-Shot Picosecond Time-Resolved MeV Electron Imaging Using a Compact Permanent Magnet Quadrupole Based Lens," *Physical Review Letters*, vol. 117, no. 2, p. 024801, Jul. 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.024801

[2] P. J. Bryant, "A Brief history and review of accelerators," pp. 1–16, 1992.

[3] A. Robinson, "History of Synchrotron Radiation Sources," in *X-Ray Data Booklet*, 3rd ed., ser. PUB. LBNL. [Online]. Available: http://xdb.lbl.gov/Section2/Sec_2-2.html

[4] R. J. England, R. J. Noble, K. Bane, D. H. Dowell, C.-K. Ng, J. E. Spencer, S. Tantawi, Z. Wu, R. L. Byer, E. Peralta, K. Soong, C.-M. Chang, B. Montazeri, S. J. Wolf, B. Cowan, J. Dawson, W. Gai, P. Hommelhoff, Y.-C. Huang, C. Jing, C. McGuinness, R. B. Palmer, B. Naranjo, J. Rosenzweig, G. Travish, A. Mizrahi, L. Schachter, C. Sears, G. R. Werner, and R. B. Yoder, "Dielectric laser accelerators," *Reviews of Modern Physics*, vol. 86, no. 4, pp. 1337–1389, Dec. 2014. [Online]. Available: http://link.aps.org/doi/10.1103/RevModPhys.86.1337

[5] T. P. Wangler, *RF Linear Accelerators*. John Wiley & Sons, Nov. 2008.

[6] J. D. Lawson, "Laser Accelerators?" Rutherford Laboratory, Chilton, Oxon, United Kingdom, Tech. Rep. RL-75-043, Mar. 1975. [Online]. Available: http://purl.org/net/epubs/work/21162

[7] E. Esarey, P. Sprangle, and J. Krall, "Laser acceleration of electrons in vacuum," *Physical Review E*, vol. 52, no. 5, pp. 5443–5453, Nov. 1995. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.52.5443

[8] M. Kozák, J. McNeur, K. J. Leedle, H. Deng, N. Schönenberger, A. Ruehl, I. Hartl, J. S. Harris, R. L. Byer, and P. Hommelhoff, "Optical gating and streaking of free electrons with sub-optical cycle precision," *Nat. Commun.*, vol. 8, p. ncomms14342, Jan. 2017. [Online]. Available: https://www.nature.com/articles/ncomms14342

[9] E. Hemsing, G. Stupakov, D. Xiang, and A. Zholents, "Beam by design: Laser manipulation of electrons in modern accelerators," *Reviews of Modern Physics*, vol. 86, no. 3, pp. 897–941, Jul. 2014. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.86.897

[10] N. Sudar, P. Musumeci, I. Gadjev, Y. Sakai, S. Fabbri, M. Polyanskiy, I. Pogorelsky, M. Fedurin, C. Swinson, K. Kusche, M. Babzien, and M. Palmer, "Demonstration of Cascaded Modulator-Chicane Microbunching of a Relativistic Electron Beam," *Physical Review Letters*, vol. 120, no. 11, p. 114802, Mar. 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.114802

[11] E. Curry, S. Fabbri, J. Maxson, P. Musumeci, and A. Gover, "Meter-Scale Terahertz-Driven Acceleration of a Relativistic Beam," *Physical Review Letters*, vol. 120, no. 9, p. 094801, Feb. 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.094801

[12] I. Gadjev, N. Sudar, M. Babzien, J. Duris, P. Hoang, M. Fedurin, K. Kusche, R. Malone, P. Musumeci, M. Palmer, I. Pogorelsky, M. Polyanskiy, Y. Sakai, C. Swinson, O. Williams, and J. B. Rosenzweig, "An inverse free electron laser acceleration-driven Compton scattering X-ray source," *arXiv:1711.00974 [physics]*, Nov. 2017, arXiv: 1711.00974. [Online]. Available: http://arxiv.org/abs/1711.00974

[13] K. Shimoda, "Proposal for an Electron Accelerator Using an Optical Maser," *Applied Optics*, vol. 1, no. 1, pp. 33–35, Jan. 1962. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-1-1-33

[14] A. W. Lohmann, "Particle accelerator utilizing coherent light," US Patent US3 267 383A, Aug., 1966. [Online]. Available: https://patents.google.com/patent/US3267383A/de

[15] Y. Takeda and I. Matsui, "Laser linac with grating," *Nuclear Instruments and Methods*, vol. 62, no. 3, pp. 306–310, Jul. 1968. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0029554X68903789

[16] "ACHIP | Accelerator on a Chip International Program." [Online]. Available: https://sites.stanford.edu/achip/

[17] T. Khabiboulline, V. Puntus, R. M. Dohlus, N. Holtkamp, G. Kreps, S. Ivanov, and K. Jin, "A new tuning method for traveling wave structures," in *Proceedings Particle Accelerator Conference*, vol. 3, May 1995, pp. 1666–1668 vol.3.

[18] D. Alesini, A. Citterio, G. Campogiani, L. Ficcadenti, M. Migliorati, A. Mostacci, L. Palumbo, S. Persichelli, and R. Zennaro, "Tuning procedure for traveling wave structures and its application to the C-Band cavities for SPARC photo injector energy upgrade," *Journal of Instrumentation*, vol. 8, no. 10, pp. P10 010–P10 010, Oct. 2013. [Online]. Available: https://doi.org/10.1088%2F1748-0221%2F8%2F10%2Fp10010

[19] K. E. Echternkamp, A. Feist, S. Schäfer, and C. Ropers, "Ramsey-type phase control of free-electron beams," *Nature Physics*, vol. 12, no. 11, pp. 1000–1004, Nov. 2016. [Online]. Available: https://www.nature.com/articles/nphys3844

[20] A. Ryabov and P. Baum, "Electron microscopy of electromagnetic waveforms," *Science*, vol. 353, no. 6297, pp. 374–377, Jul. 2016. [Online]. Available: http://science.sciencemag.org/content/353/6297/374

[21] G. M. Vanacore, I. Madan, G. Berruto, K. Wang, E. Pomarico, R. J. Lamb, D. McGrouther, I. Kaminer, B. Barwick, F. J. G. d. Abajo, and F. Carbone, "Attosecond coherent control of free-electron wave functions using semi-infinite light fields," *Nature Communications*, vol. 9, no. 1, p. 2694, Jul. 2018. [Online]. Available: https://www.nature.com/articles/s41467-018-05021-x

[22] E. A. Peralta, K. Soong, R. J. England, E. R. Colby, Z. Wu, B. Montazeri, C. McGuinness, J. McNeur, K. J. Leedle, D. Walz, E. B. Sozer, B. Cowan, B. Schwartz, G. Travish, and R. L. Byer, "Demonstration of electron acceleration in a laser-driven dielectric microstructure," *Nature*, vol. 503, no. 7474, pp. 91–94, Nov. 2013. [Online]. Available: http://www.nature.com/nature/journal/v503/n7474/full/nature12664.html

[23] J. Breuer and P. Hommelhoff, "Laser-Based Acceleration of Nonrelativistic Electrons at a Dielectric Structure," *Physical Review Letters*, vol. 111, no. 13, p. 134803, Sep. 2013. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.111.134803

[24] K. J. Leedle, R. Fabian Pease, R. L. Byer, and J. S. Harris, "Laser acceleration and deflection of 963 keV electrons with a silicon dielectric structure," *Optica*, vol. 2, no. 2, p. 158, Feb. 2015. [Online]. Available: https://www.osapublishing.org/abstract.cfm?URI=optica-2-2-158

[25] M. Kozák, M. Förster, J. McNeur, N. Schönenberger, K. Leedle, H. Deng, J. S. Harris, R. L. Byer, and P. Hommelhoff, "Dielectric laser acceleration of sub-relativistic electrons by few-cycle laser pulses," *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 865, no. Supplement C, pp. 84–86, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900216313195

[26] P. Yousefi, N. Schönenberger, J. Mcneur, M. Kozák, U. Niedermayer, and P. Hommelhoff, "Dielectric-laser electron acceleration in a dual pillar grating with a distributed Bragg reflector," p. 5, Feb. 2019. [Online]. Available: http://dx.doi.org/10.1364/OL.99.099999

[27] T. Plettner, P. P. Lu, and R. L. Byer, "Proposed few-optical cycle laser-driven particle accelerator structure," *Phys. Rev. ST Accel. Beams*, vol. 9, no. 11, p. 111301, Nov. 2006. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevSTAB.9.111301

[28] K. P. Wootton, Z. Wu, B. M. Cowan, A. Hanuka, I. V. Makasyuk, E. A. Peralta, K. Soong, R. L. Byer, and R. J. England, "Demonstration of acceleration of relativistic electrons at a dielectric microstructure using femtosecond laser pulses," *Optics Letters*, vol. 41, no. 12, p. 2696, Jun. 2016. [Online]. Available: https://www.osapublishing.org/abstract.cfm?URI=ol-41-12-2696

[29] D. Cesar, S. Custodio, J. Maxson, P. Musumeci, X. Shen, E. Threlkeld, R. J. England, A. Hanuka, I. V. Makasyuk, E. A. Peralta, K. P. Wootton, and Z. Wu, "High-field nonlinear optical response and phase control in a dielectric laser accelerator," *Communications Physics*, vol. 1, no. 1, p. 46, Aug. 2018. [Online]. Available: https://www.nature.com/articles/s42005-018-0047-y

[30] D. Cesar, J. Maxson, X. Shen, K. P. Wootton, S. Tan, R. J. England, and P. Musumeci, "Enhanced energy gain in a dielectric laser accelerator using a tilted pulse front laser," *Optics Express*, vol. 26, no. 22, pp. 29 216–29 224, Oct. 2018. [Online]. Available: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-26-22-29216

[31] M. Kozák, J. McNeur, N. Schönenberger, J. Illmer, A. Li, A. Tafel, P. Yousefi, T. Eckstein, and P. Hommelhoff, "Ultrafast scanning electron microscope applied for studying the interaction between free electrons and optical near-fields of periodic nanostructures," *arXiv:1805.07402 [physics]*, May 2018, arXiv: 1805.07402. [Online]. Available: http://arxiv.org/abs/1805.07402

[32] J. McNeur, M. Kozák, N. Schönenberger, K. J. Leedle, H. Deng, A. Ceballos, H. Hoogland, A. Ruehl, I. Hartl, R. Holzwarth, O. Solgaard, J. S. Harris, R. L. Byer, and P. Hommelhoff, "Elements of a dielectric laser accelerator," *Optica*, vol. 5, no. 6, pp. 687–690, Jun. 2018. [Online]. Available: https://www.osapublishing.org/optica/abstract.cfm?uri=optica-5-6-687

[33] B. Naranjo, A. Valloni, S. Putterman, and J. B. Rosenzweig, "Stable Charged-Particle Acceleration and Focusing in a Laser Accelerator Using Spatial Harmonics," *Physical Review Letters*, vol. 109, no. 16, p. 164803, Oct. 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.109.164803

[34] Helmut Wiedemann, *Particle Accelerator Physics*, 2007. [Online]. Available: http://www.springer.com/us/book/9783540490456

[35] D. S. Black, K. J. Leedle, Y. Miao, U. Niedermayer, R. L. Byer, and O. Solgaard, "Laser-Driven Electron Lensing in Silicon Microstructures," *arXiv:1902.00170 [physics]*, Jan. 2019, arXiv: 1902.00170. [Online]. Available: http://arxiv.org/abs/1902.00170

[36] B. Naranjo, "GALAXIE: A Compact X-ray FEL," San Juan, Puerto Rico, Mar. 2013. [Online]. Available: http://pbpl.physics.ucla.edu/HBEB2013/Program.html

[37] U. Niedermayer, T. Egenolf, O. Boine-Frankenheim, and P. Hommelhoff, "Alternating-Phase Focusing for Dielectric-Laser Acceleration," *Physical Review Letters*, vol. 121, no. 21, p. 214801, Nov. 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.121.214801

[38] D. Cesar, P. Musumeci, and J. England, "All Optical Control of Beam Dynamics in a DLA," in *2018 IEEE Advanced Accelerator Concepts Workshop (AAC)*, Breckenridge, CO, Aug. 2018, pp. 1–5.

[39] Y. C. N. Na, R. H. Siemann, and R. L. Byer, "Energy efficiency of an intracavity coupled, laser-driven linear accelerator pumped by an external laser," *Physical Review Special Topics - Accelerators and Beams*, vol. 8, no. 3, p. 031301, Mar. 2005. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.8.031301

[40] D. Ehberger, J. Hammer, M. Eisele, M. Krüger, J. Noe, A. Högele, and P. Hommelhoff, "Highly Coherent Electron Beam from a Laser-Triggered Tungsten Needle Tip," *Physical Review Letters*, vol. 114, no. 22, p. 227601, Jun. 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.114.227601

[41] N. Bell and L. Phillips, "Generation of Flat Ultra-Low Emittance Beams." JACOW Publishing, Geneva, Switzerland, Jun. 2018, pp. 3398–3400. [Online]. Available: http://accelconf.web.cern.ch/AccelConf/ipac2018/doi/JACoW-IPAC2018-THPAK072.html

[42] T. W. Hughes, S. Tan, Z. Zhao, N. V. Sapra, K. J. Leedle, H. Deng, Y. Miao, D. S. Black, O. Solgaard, J. S. Harris, J. Vuckovic, R. L. Byer, S. Fan, R. J. England, Y. J. Lee, and M. Qi, "On-Chip Laser-Power Delivery System for Dielectric Laser Accelerators," *Physical Review Applied*, vol. 9, no. 5, p. 054017, May 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.9.054017

[43] T. W. Hughes, R. J. England, and S. Fan, "Reconfigurable Photonic Circuit for Controlled Power Delivery to Laser-Driven Accelerators on a Chip," *arXiv:1902.09731 [physics]*, Feb. 2019, arXiv: 1902.09731. [Online]. Available: http://arxiv.org/abs/1902.09731

[44] S. Tan, Z. Zhao, K. Urbanek, T. Hughes, Y. J. Lee, S. Fan, J. S. Harris, and R. L. Byer, "Silicon nitride waveguide as a power delivery component for on-chip dielectric laser accelerators," *Optics Letters*, vol. 44, no. 2, pp. 335–338, Jan. 2019. [Online]. Available: https://www.osapublishing.org/ol/abstract.cfm?uri=ol-44-2-335

[45] Z. Zhao, T. W. Hughes, S. Tan, H. Deng, N. Sapra, R. J. England, J. Vuckovic, J. S. Harris, R. L. Byer, and S. Fan, "Design of a tapered slot waveguide dielectric laser accelerator for sub-relativistic electrons," *Optics Express*, vol. 26, no. 18, pp. 22 801–22 815, Sep. 2018. [Online]. Available: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-26-18-22801

[46] E. Esarey, C. B. Schroeder, and W. P. Leemans, "Physics of laser-driven plasma-based electron accelerators," *Reviews of Modern Physics*, vol. 81, no. 3, pp. 1229–1285, Aug. 2009. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.81.1229

[47] A. Gonsalves, K. Nakamura, J. Daniels, C. Benedetti, C. Pieronek, T. de Raadt, S. Steinke, J. Bin, S. Bulanov, J. van Tilborg, C. Geddes, C. Schroeder, C. Tóth, E. Esarey, K. Swanson, L. Fan-Chiang, G. Bagdasarov, N. Bobrova, V. Gasilov, G. Korn, P. Sasorov, and W. Leemans, "Petawatt Laser Guiding and Electron Beam Acceleration to 8 GeV in a Laser-Heated Capillary Discharge Waveguide," *Physical Review Letters*, vol. 122, no. 8, p. 084801, Feb. 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.122.084801

[48] I. Blumenfeld, C. E. Clayton, F.-J. Decker, M. J. Hogan, C. Huang, R. Ischebeck, R. Iverson, C. Joshi, T. Katsouleas, N. Kirby, W. Lu, K. A. Marsh, W. B. Mori, P. Muggli, E. Oz, R. H. Siemann, D. Walz, and M. Zhou, "Energy doubling of 42 GeV electrons in a metre-scale plasma wakefield accelerator," *Nature*, vol. 445, no. 7129, pp. 741–744, Feb. 2007. [Online]. Available: https://www.nature.com/articles/nature05538

[49] B. D. O'Shea, G. Andonian, S. K. Barber, K. L. Fitzmorris, S. Hakimi, J. Harrison, P. D. Hoang, M. J. Hogan, B. Naranjo, O. B. Williams, V. Yakimenko, and J. B.

Rosenzweig, “Observation of acceleration and deceleration in gigaelectron-volt-per-metre gradient dielectric wakefield accelerators,” *Nature Communications*, vol. 7, p. 12763, Sep. 2016. [Online]. Available: https://www.nature.com/articles/ncomms12763

[50] A. Ody, R. J. England, and Z. Huang, “Simulation of a dielectric deflecting structure for short-wavelength radiation,” in *Proceedings of the IEEE*. Breckenridge, CO: IEEE, 2018, p. 4. [Online]. Available: http://slac.stanford.edu/pubs/slacpubs/17250/slac-pub-17379.pdf

[51] K. Koyama, S. Otsuki, M. Uesaka, M. Yoshida, and A. Aimidula, “Parameter study of a laser-driven dielectric accelerator for radiobiology research,” *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 47, no. 23, p. 234005, 2014. [Online]. Available: http://stacks.iop.org/0953-4075/47/i=23/a=234005

[52] G. H. Campbell, J. T. McKeown, and M. K. Santala, “Time resolved electron microscopy for in situ experiments,” *Applied Physics Reviews*, vol. 1, no. 4, p. 041101, Nov. 2014. [Online]. Available: https://aip.scitation.org/doi/10.1063/1.4900509

[53] W. E. King, G. H. Campbell, A. Frank, B. Reed, J. F. Schmerge, B. J. Siwick, B. C. Stuart, and P. M. Weber, “Ultrafast electron microscopy in materials science, biology, and chemistry,” *Journal of Applied Physics*, vol. 97, no. 11, p. 111101, Jun. 2005. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.1927699

[54] V. A. Lobastov, R. Srinivasan, and A. H. Zewail, “Four-dimensional ultrafast electron microscopy,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 20, pp. 7069–7073, May 2005. [Online]. Available: http://www.pnas.org/content/102/20/7069

[55] F. Carbone, P. Musumeci, O. J. Luiten, and C. Hebert, “A perspective on novel sources of ultrashort electron and X-ray pulses,” *Chemical Physics*, vol. 392, no. 1, pp. 1–9, Jan. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301010411004319

[56] B. Barwick, D. J. Flannigan, and A. H. Zewail, “Photon-induced near-field electron microscopy,” *Nature*, vol. 462, no. 7275, pp. 902–906, Dec. 2009. [Online]. Available: https://www.nature.com/articles/nature08662

[57] O. Bostanjoglo, R. P. Tornow, and W. Tornow, “Nanosecond-exposure electron microscopy of laser-induced phase transformations,” *Ultramicroscopy*, vol. 21, no. 4, pp. 367–372, Jan. 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0304399187900349

[58] A. Lassise, P. H. A. Mutsaers, and O. J. Luiten, “Compact, low power radio frequency cavity for femtosecond electron microscopy,” *Review of Scientific Instruments*, vol. 83, no. 4, p. 043705, Apr. 2012. [Online]. Available: https://aip.scitation.org/doi/full/10.1063/1.3703314

[59] G. Mourou and S. Williamson, “Picosecond electron diffraction,” *Applied Physics Letters*, vol. 41, no. 1, pp. 44–45, Jul. 1982. [Online]. Available: http://scitation.aip.org/content/aip/journal/apl/41/1/10.1063/1.93316

[60] W. E. Spicer and A. Herrera-Gomez, "Modern theory and applications of photocathodes," in *Photodetectors and Power Meters*, vol. 2022. International Society for Optics and Photonics, Oct. 1993, pp. 18–36. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/2022/0000/Modern-theory-and-applications-of-photocathodes/10.1117/12.158575.short

[61] H. S. Park, J. S. Baskin, O.-H. Kwon, and A. H. Zewail, "Atomic-Scale Imaging in Real and Energy Space Developed in Ultrafast Electron Microscopy," *Nano Letters*, vol. 7, no. 9, pp. 2545–2551, Sep. 2007. [Online]. Available: https://doi.org/10.1021/nl071369q

[62] A. Feist, N. Bach, N. Rubiano da Silva, T. Danz, M. Möller, K. E. Priebe, T. Domröse, J. G. Gatzmann, S. Rost, J. Schauss, S. Strauch, R. Bormann, M. Sivis, S. Schäfer, and C. Ropers, "Ultrafast transmission electron microscopy using a laser-driven field emitter: Femtosecond resolution with a high coherence electron beam," *Ultramicroscopy*, vol. 176, pp. 63–73, May 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399116303709

[63] T. LaGrange, M. R. Armstrong, K. Boyden, C. G. Brown, G. H. Campbell, J. D. Colvin, W. J. DeHope, A. M. Frank, D. J. Gibson, F. V. Hartemann, J. S. Kim, W. E. King, B. J. Pyke, B. W. Reed, M. D. Shirk, R. M. Shuttlesworth, B. C. Stuart, B. R. Torralva, and N. D. Browning, "Single-shot dynamic transmission electron microscopy," *Applied Physics Letters*, vol. 89, no. 4, p. 044105, Jul. 2006. [Online]. Available: https://aip.scitation.org/doi/full/10.1063/1.2236263

[64] M. R. Armstrong, K. Boyden, N. D. Browning, G. H. Campbell, J. D. Colvin, W. J. DeHope, A. M. Frank, D. J. Gibson, F. Hartemann, J. S. Kim, W. E. King, T. B. LaGrange, B. J. Pyke, B. W. Reed, R. M. Shuttlesworth, B. C. Stuart, and B. R. Torralva, "Practical considerations for high spatial and temporal resolution dynamic transmission electron microscopy," *Ultramicroscopy*, vol. 107, no. 4, pp. 356–367, Apr. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399106001872

[65] N. D. Browning, M. A. Bonds, G. H. Campbell, J. E. Evans, T. LaGrange, K. L. Jungjohann, D. J. Masiel, J. McKeown, S. Mehraeen, B. W. Reed, and M. Santala, "Recent developments in dynamic transmission electron microscopy," *Current Opinion in Solid State and Materials Science*, vol. 16, no. 1, pp. 23–30, Feb. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1359028611000544

[66] R. K. Li, P. Musumeci, H. A. Bender, N. S. Wilcox, and M. Wu, "Imaging single electrons to enable the generation of ultrashort beams for single-shot femtosecond relativistic electron diffraction," *Journal of Applied Physics*, vol. 110, no. 7, p. 074512, Oct. 2011. [Online]. Available: http://scitation.aip.org/content/aip/journal/jap/110/7/10.1063/1.3646465

[67] J. R. Dwyer, C. T. Hebeisen, R. Ernstorfer, M. Harb, V. B. Deyirmenjian, R. E. Jordan, and R. J. D. Miller, "Femtosecond electron diffraction: 'making the molecular

movie',” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 364, no. 1840, pp. 741–778, Mar. 2006. [Online]. Available: http://rsta.royalsocietypublishing.org/content/364/1840/741

[68] X. Wang, D. Xiang, T. Kim, and H. Ihee, “Potential of Femtosecond Electron Diffraction Using Near-Relativistic Electrons from a Photocathode RF Electron Gun,” *Journal of the Korean Physical Society*, vol. 48, no. 3, pp. 390–396, 2006. [Online]. Available: http://inis.iaea.org/Search/search.aspx?orig_q=RN:40025959

[69] P. Musumeci, J. T. Moody, C. M. Scoby, M. S. Gutierrez, and M. Westfall, “Laser-induced melting of a single crystal gold sample by time-resolved ultrafast relativistic electron diffraction,” *Applied Physics Letters*, vol. 97, no. 6, p. 063502, Aug. 2010. [Online]. Available: http://scitation.aip.org/content/aip/journal/apl/97/6/10.1063/1.3478005

[70] J. Yang, K. Kan, N. Naruse, Y. Yoshida, K. Tanimura, and J. Urakawa, “100-femtosecond MeV electron source for ultrafast electron diffraction,” *Radiation Physics and Chemistry*, vol. 78, no. 12, pp. 1106–1111, Dec. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0969806X0900190X

[71] S. Manz, A. Casandruc, D. Zhang, Y. Zhong, R. A. Loch, A. Marx, T. Hasegawa, L. C. Liu, S. Bayesteh, H. Delsim-Hashemi, M. Hoffmann, M. Felber, M. Hachmann, F. Mayet, J. Hirscht, S. Keskin, M. Hada, S. W. Epp, K. Flöttmann, and R. J. D. Miller, “Mapping atomic motions with ultrabright electrons: towards fundamental limits in space-time resolution,” *Faraday Discussions*, vol. 177, no. 0, pp. 467–491, Apr. 2015. [Online]. Available: http://pubs.rsc.org/en/content/articlelanding/2015/fd/c4fd00204k

[72] S. P. Weathersby, G. Brown, M. Centurion, T. F. Chase, R. Coffee, J. Corbett, J. P. Eichner, J. C. Frisch, A. R. Fry, M. Gühr, N. Hartmann, C. Hast, R. Hettel, R. K. Jobe, E. N. Jongewaard, J. R. Lewandowski, R. K. Li, A. M. Lindenberg, I. Makasyuk, J. E. May, D. McCormick, M. N. Nguyen, A. H. Reid, X. Shen, K. Sokolowski-Tinten, T. Vecchione, S. L. Vetter, J. Wu, J. Yang, H. A. Dürr, and X. J. Wang, “Mega-electron-volt ultrafast electron diffraction at SLAC National Accelerator Laboratory,” *Review of Scientific Instruments*, vol. 86, no. 7, p. 073702, Jul. 2015. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.4926994

[73] F. Fu, S. Liu, P. Zhu, D. Xiang, J. Zhang, and J. Cao, “High quality single shot ultrafast MeV electron diffraction from a photocathode radio-frequency gun,” *Review of Scientific Instruments*, vol. 85, no. 8, p. 083701, Aug. 2014. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.4892135

[74] C. Lu, T. Jiang, S. Liu, R. Wang, L. Zhao, P. Zhu, Y. Liu, J. Xu, D. Yu, W. Wan, Y. Zhu, D. Xiang, and J. Zhang, “Imaging nanoscale spatial modulation of a relativistic electron beam with a MeV ultrafast electron microscope,” *Applied Physics Letters*, vol. 112, no. 11, p. 113102, Mar. 2018. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.5023179

[75] F. Aghamir, W. A. Barletta, D. B. Cline, J. W. Dodd, S. C. Hartman, T. C. Katsouleas, J. Kolonko, S. Park, C. Pellegrini, J. C. Terrien, J. G. Davis, C. J. Joshi, N. C. Luhmann, D. B. McDermott, S. N. Ivanchenkov, Y. Lachin, and A. A. Varfolomeev, "Saturnus: the UCLA high-gain infrared FEL project," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 304, no. 1, pp. 155–158, Jul. 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016890029190839I

[76] S. C. Hartman, N. Barov, C. Pellegrini, S. Park, J. Rosenzweig, G. Travish, R. Zhang, C. Clayton, P. Davis, M. Everett, C. Joshi, and G. Hairapetian, "Initial measurements of the UCLA rf photoinjector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 340, no. 1, pp. 219–230, Feb. 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0168900294913056

[77] M. Hogan, C. Pellegrini, J. Rosenzweig, G. Travish, A. Varfolomeev, S. Anderson, K. Bishofberger, P. Frigola, A. Murokh, N. Osmanov, S. Reiche, and A. Tremaine, "Measurements of High Gain and Intensity Fluctuations in a Self-Amplified, Spontaneous-Emission Free-Electron Laser," *Physical Review Letters*, vol. 80, no. 2, pp. 289–292, Jan. 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.80.289

[78] S. Telfer, S. Reiche, J. B. Rosenzweig, P. Frigola, and G. Andonian, "Commissioning of the UCLA PEGASUS Photoinjector laboratory," in *PACS2001. Proceedings of the 2001 Particle Accelerator Conference (Cat. No.01CH37268)*, vol. 3, Jun. 2001, pp. 2263–2265 vol.3.

[79] G. Andonian, P. Frigola, S. Reiche, J. B. Rosenzweig, S. Telfer, and G. Travish, "Status of the UCLA PEGASUS laboratory," in *Proceedings of the 2003 Particle Accelerator Conference*, vol. 3, May 2003, pp. 2110–2111 vol.3.

[80] S. Reiche, G. Andonian, P. Frigola, J. B. Rosenzweig, S. Telfer, and G. Travish, "The FEL program at the PEGASUS injector," in *Proceedings of the 2003 Particle Accelerator Conference*, vol. 2, May 2003, pp. 947–949 Vol.2.

[81] P. Musumeci, J. Moody, and G. Gatti, "Ultrafast beam research at the PEGASUS laboratory," in *2007 IEEE Particle Accelerator Conference (PAC)*, Jun. 2007, pp. 2751–2753.

[82] J. Maxson, D. Cesar, G. Calmasini, A. Ody, P. Musumeci, and D. Alesini, "Direct Measurement of Sub-10 fs Relativistic Electron Beams with Ultralow Emittance," *Phys. Rev. Lett.*, vol. 118, no. 15, p. 154802, Apr. 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.118.154802

[83] D. Alesini, A. Battisti, M. Ferrario, L. Foggetta, V. Lollo, L. Ficcadenti, V. Pettinacci, S. Custodio, E. Pirez, P. Musumeci, and L. Palumbo, "New technology based on clamping for high gradient radio frequency photogun," *Physical Review Special Topics*

- *Accelerators and Beams*, vol. 18, no. 9, p. 092001, Sep. 2015. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevSTAB.18.092001

[84] N. Barov, R. H. Miller, and D. J. Newsham, "Development of the Dual-Slot Resonance Linac," *Conf.Proc.*, vol. C110328, pp. 1897–1899, 2011.

[85] R. J. England, "Longitudinal shaping of relativistic bunches of electrons generated by an RF photoinjector," Ph.D., University of California, Los Angeles, United States – California, 2007. [Online]. Available: https://search.proquest.com/docview/304875646/abstract/6692F06F804244D0PQ/1

[86] J. E. Clendenin, "RF photoinjectors," 1996. [Online]. Available: http://inis.iaea.org/Search/search.aspx?orig_q=RN:39075023

[87] M. Hachmann and K. Flöttmann, "Measurement of ultra low transverse emittance at REGAE," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 829, pp. 318–320, Sep. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900216000930

[88] D. Filippetto and H. Qian, "Design of a high-flux instrument for ultrafast electron diffraction and microscopy," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 49, no. 10, p. 104003, 2016. [Online]. Available: http://stacks.iop.org/0953-4075/49/i=10/a=104003

[89] J. B. Rosenzweig, A. Cahill, V. Dolgashev, C. Emma, A. Fukusawa, R. Li, C. Limborg, J. Maxson, P. Musumeci, A. Nause, R. Pakter, R. Pompili, R. Roussel, B. Spataro, and S. Tantawi, "Next Generation High Brightness Electron Beams From Ultra-High Field Cryogenic Radiofrequency Photocathode Sources," *arXiv:1603.01657 [physics]*, Mar. 2016, arXiv: 1603.01657. [Online]. Available: http://arxiv.org/abs/1603.01657

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

[91] D. B. Cesar, P. Musumeci, and D. Alesini, "Ultrafast gating of a mid-infrared laser pulse by a sub-pC relativistic electron beam," *Journal of Applied Physics*, vol. 118, no. 23, p. 234506, Dec. 2015. [Online]. Available: http://aip.scitation.org/doi/full/10.1063/1.4937401

[92] I. V. Bazarov, B. M. Dunham, and C. K. Sinclair, "Maximum Achievable Beam Brightness from Photoinjectors," *Physical Review Letters*, vol. 102, no. 10, p. 104801, Mar. 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.102.104801

[93] D. Filippetto, P. Musumeci, M. Zolotorev, and G. Stupakov, "Maximum current density and beam brightness achievable by laser-driven electron sources," *Physical*

Review Special Topics - Accelerators and Beams, vol. 17, no. 2, p. 024201, Feb. 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.17.024201

[94] D. Cesar, J. Maxson, P. Musumeci, X. Shen, R. J. England, and K. P. Wootton, "Optical design for increased interaction length in a high gradient dielectric laser accelerator," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Feb. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900218300263

[95] T. P. Wangler, "Standard Linac Structures," in *RF Linear Accelerators*, 2nd ed. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2008, pp. 83–134. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/9783527623426.ch4/summary

[96] K. Soong, R. L. Byer, E. R. Colby, R. J. England, and E. A. Peralta, "Laser damage threshold measurements of optical materials for direct laser accelerators," *AIP Conference Proceedings*, Dec. 2012. [Online]. Available: http://aip.scitation.org/doi/abs/10.1063/1.4773749

[97] A.-C. Tien, S. Backus, H. Kapteyn, M. Murnane, and G. Mourou, "Short-Pulse Laser Damage in Transparent Materials as a Function of Pulse Duration," *Physical Review Letters*, vol. 82, no. 19, pp. 3883–3886, May 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.82.3883

[98] M. Kozák, P. Beck, H. Deng, J. McNeur, N. Schönenberger, C. Gaida, F. Stutzki, M. Gebhardt, J. Limpert, A. Ruehl, I. Hartl, O. Solgaard, J. S. Harris, R. L. Byer, and P. Hommelhoff, "Acceleration of sub-relativistic electrons with an evanescent optical wave at a planar interface," *Opt. Express*, vol. 25, no. 16, pp. 19 195–19 204, Aug. 2017. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=oe-25-16-19195

[99] T. Plettner, R. L. Byer, and B. Montazeri, "Electromagnetic forces in the vacuum region of laser-driven layered grating structures," *Journal of Modern Optics*, vol. 58, no. 17, pp. 1518–1528, Oct. 2011. [Online]. Available: http://dx.doi.org/10.1080/09500340.2011.611914

[100] K. J. Leedle, A. Ceballos, H. Deng, O. Solgaard, R. F. Pease, R. L. Byer, and J. S. Harris, "Dielectric laser acceleration of sub-100 keV electrons with silicon dual-pillar grating structures," *Opt. Lett.*, vol. 40, no. 18, p. 4344, Sep. 2015. [Online]. Available: http://dx.doi.org/10.1364/OL.40.004344

[101] P. Yousefi, J. McNeur, M. Kozák, F. Gannott, O. Lohse, and P. Hommelhoff, "Silicon dual pillar structure with a distributed Bragg reflector for dielectric laser accelerators: Design and fabrication," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Jan. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900218300822

[102] E. A. Peralta, "Accelerator on a chip design, fabrication, and demonstration of grating-based dielectric microstructures for laser-driven acceleration of electrons /," Ph.D. dissertation, Stanford University, 2015. [Online]. Available: http://purl.stanford.edu/ht547xt5560

[103] J. D. Jackson, *Classical Electrodynamics, 3rd Edition*, Jul. 1998. [Online]. Available: http://adsabs.harvard.edu/abs/1998clel.book.....J

[104] A. E. Siegman, *Lasers.* University Science Books, 1986, google-Books-ID: 1BZVwU-ZLTkAC.

[105] J.-C. Diels and W. Rudolph, *Ultrashort Laser Pulse Phenomena.* Elsevier, Sep. 2006, google-Books-ID: rDQe81K0d3kC.

[106] J. Hebling, "Derivation of the pulse front tilt caused by angular dispersion," *Optical and Quantum Electronics*, vol. 28, no. 12, pp. 1759–1763, Dec. 1996. [Online]. Available: http://link.springer.com/article/10.1007/BF00698541

[107] D. Milam, "Review and assessment of measured values of the nonlinear refractive-index coefficient of fused silica," *Applied Optics*, vol. 37, no. 3, pp. 546–550, Jan. 1998. [Online]. Available: http://www.osapublishing.org/abstract.cfm?uri=ao-37-3-546

[108] A. M. Sommer, *Ultrafast Strong Field Dynamics in Dielectrics.* Springer, Jul. 2016, google-Books-ID: Ixu1DAAAQBAJ.

[109] E. Treacy, "Optical pulse compression with diffraction gratings," *IEEE Journal of Quantum Electronics*, vol. 5, no. 9, pp. 454–458, Sep. 1969.

[110] "FDTD Solutions | Lumerical's Nanophotonic FDTD Simulation Software." [Online]. Available: http://www.lumerical.com/tcad-products/fdtd/

[111] G. Agrawal, "Chapter 2 - Pulse Propagation in Fibers," in *Nonlinear Fiber Optics (Fifth Edition)*, ser. Optics and Photonics. Boston: Academic Press, 2013, pp. 27–56. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123970237000024

[112] J. d. Hoyo, A. R. de la Cruz, E. Grace, A. Ferrer, J. Siegel, A. Pasquazi, G. Assanto, and J. Solis, "Rapid assessment of nonlinear optical propagation effects in dielectrics," *Scientific Reports*, vol. 5, p. 7650, Jan. 2015. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4288215/

[113] E. Grace, "GAFFE A toolbox for solving evolutionary nonlinear PDEs," 2009. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/24016-gaffe-a-toolbox-for-solving-evolutionary-nonlinear-pdes

[114] E. Esarey and M. Pilloff, "Trapping and acceleration in nonlinear plasma waves," *Physics of Plasmas*, vol. 2, no. 5, pp. 1432–1436, May 1995. [Online]. Available: https://aip.scitation.org/doi/10.1063/1.871358

[115] N. Kroll, P. Morton, and M. Rosenbluth, "Free-electron lasers with variable parameter wigglers," *IEEE Journal of Quantum Electronics*, vol. 17, no. 8, pp. 1436–1468, Aug. 1981. [Online]. Available: https://ieeexplore.ieee.org/document/1071285

[116] N. Sudar, D. Garzella, and P. Musumeci, "Inverse Free Electron Laser Separatrix Crossing for Energy Gain and Stability," p. THPMK028, Jun. 2018.

[117] W. Paul, "Electromagnetic traps for charged and neutral particles," *Reviews of Modern Physics*, vol. 62, no. 3, pp. 531–540, Jul. 1990. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.62.531

[118] David Cesar, "Dielectric laser acceleration: research towards an accelerator on a chip," Breckenridge, CO, 2018. [Online]. Available: https://aac2018.txcorp.com/event/1/contributions/261/

[119] L. D. Landau and E. M. Lifshitz, *Mechanics: Volume 1*, 3rd ed. Amsterdam u.a: Butterworth-Heinemann, Jan. 1976.

[120] B. V. Chirikov, "Research concerning the theory of non-linear resonance and stochasticity," 1971. [Online]. Available: http://inspirehep.net/record/898561/

[121] A. Ody, P. Musumeci, J. Maxson, D. Cesar, R. J. England, and K. P. Wootton, "Flat electron beam sources for DLA accelerators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 865, pp. 75–83, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900216310877

[122] J. Holmes, S. Henderson, and Y. Zhang, "Transverse beam optics," Vanderbilt, 2009. [Online]. Available: http://uspas.fnal.gov/materials/09VU/Lecture7.pdf

[123] K. T. McDonald, "Radiation Pressure of a Monochromatic Plane Wave on a Flat Mirror," p. 19. [Online]. Available: http://www.physics.princeton.edu/~mcdonald/examples/pressure.pdf

[124] J. Hebling, K.-L. Yeh, M. C. Hoffmann, B. Bartal, and K. A. Nelson, "Generation of high-power terahertz pulses by tilted-pulse-front excitation and their application possibilities," *JOSA B*, vol. 25, no. 7, pp. B6–B19, Jul. 2008. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=josab-25-7-B6

[125] D. Kreier and P. Baum, "Avoiding temporal distortions in tilted pulses," *Optics Letters*, vol. 37, no. 12, pp. 2373–2375, Jun. 2012. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=ol-37-12-2373

[126] P. Poole, S. Trendafilov, G. Shvets, D. Smith, and E. Chowdhury, "Femtosecond laser damage threshold of pulse compression gratings for petawatt scale laser systems," *Optics Express*, vol. 21, no. 22, pp. 26 341–26 351, Nov. 2013. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=oe-21-22-26341

[127] A. G. Kostenbauder, "Ray-pulse matrices: a rational treatment for dispersive optical systems," *IEEE Journal of Quantum Electronics*, vol. 26, no. 6, pp. 1148–1157, Jun. 1990.

[128] G. Marcus, "Spatial and temporal pulse propagation for dispersive paraxial optical systems," *Optics Express*, vol. 24, no. 7, pp. 7752–7766, Apr. 2016. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=oe-24-7-7752

[129] S. Akturk, X. Gu, P. Gabolde, and R. Trebino, "The general theory of first-order spatio-temporal distortions of Gaussian pulses and beams," *Optics Express*, vol. 13, no. 21, pp. 8642–8661, Oct. 2005. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=oe-13-21-8642

[130] G. Pariente, V. Gallet, A. Borot, O. Gobert, and F. Quéré, "Space–time characterization of ultra-intense femtosecond laser beams," *Nature Photonics*, vol. 10, no. 8, pp. 547–553, Aug. 2016. [Online]. Available: https://www.nature.com/articles/nphoton.2016.140

[131] R. Trebino, *Frequency-Resolved Optical Gating: The Measurement of.* Springer Science + Business Media, 2000. [Online]. Available: http://www.springer.com/us/book/9781402070662

[132] G. Dugan, A. Misuri, and W. Leemans, "Design and performance estimates for the l'OASIS experiment magnetic spectrometers," Nov. 2001. [Online]. Available: https://escholarship.org/uc/item/98h5r889

[133] S. e. Agostinelli, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, Jul. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900203013688

[134] M. J. Berger, "ESTAR, PSTAR, and ASTAR: Computer programs for calculating stopping-power and range tables for electrons, protons, and helium ions," Tech. Rep., Dec. 1992. [Online]. Available: http://adsabs.harvard.edu/abs/1992esta.rept.....B

[135] C. M. Scoby, R. K. Li, and P. Musumeci, "Effect of an ultrafast laser induced plasma on a relativistic electron beam to determine temporal overlap in pump–probe experiments," *Ultramicroscopy*, vol. 127, pp. 14–18, Apr. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399112001854

[136] M. Kozák, J. McNeur, K. J. Leedle, H. Deng, N. Schönenberger, A. Ruehl, I. Hartl, H. Hoogland, R. Holzwarth, J. S. Harris, R. L. Byer, and P. Hommelhoff, "Transverse and longitudinal characterization of electron beams using interaction with optical near-fields," *Optics Letters*, vol. 41, no. 15, pp. 3435–3438, Aug. 2016. [Online]. Available: http://www.osapublishing.org/abstract.cfm?uri=ol-41-15-3435

[137] D. A. Fish, A. M. Brinicombe, E. R. Pike, and J. G. Walker, "Blind deconvolution by means of the Richardson–Lucy algorithm," *JOSA A*, vol. 12, no. 1, pp. 58–65,

Jan. 1995. [Online]. Available: https://www.osapublishing.org/josaa/abstract.cfm?uri=josaa-12-1-58

[138] S. Y. Lee, *Accelerator Physics*, 3rd ed. Singapore: World Scientific Publishing Company, Nov. 2011. [Online]. Available: http://dx.doi.org/10.1142/8335

[139] O. Scherzer, "The Theoretical Resolution Limit of the Electron Microscope," *Journal of Applied Physics*, vol. 20, no. 1, pp. 20–29, Jan. 1949. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.1698233

[140] V. E. Cosslett, "High voltage electron microscopy," *Science Progress (1933-)*, vol. 55, no. 217, pp. 15–34, 1967. [Online]. Available: https://www.jstor.org/stable/43419618

[141] F. Phillipp, R. Höschen, M. Osaki, G. Möbus, and M. Rühle, "New high-voltage atomic resolution microscope approaching 1 Å point resolution installed in Stuttgart," *Ultramicroscopy*, vol. 56, no. 1, pp. 1–10, Nov. 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0304399194901414

[142] H. H. Rose, "Historical aspects of aberration correction," *Journal of Electron Microscopy*, vol. 58, no. 3, pp. 77–85, Jun. 2009. [Online]. Available: https://academic.oup.com/jmicro/article/58/3/77/1989077

[143] S. Tsujino, P. Das Kanungo, M. Monshipouri, C. Lee, and R. J. D. Miller, "Measurement of transverse emittance and coherence of double-gate field emitter array cathodes," *Nature Communications*, vol. 7, p. 13976, Dec. 2016. [Online]. Available: https://www.nature.com/articles/ncomms13976

[144] R. Li and P. Musumeci, "Single-Shot MeV Transmission Electron Microscopy with Picosecond Temporal Resolution," *Physical Review Applied*, vol. 2, no. 2, p. 024003, Aug. 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.2.024003

[145] E. Pirez, D. Alesini, J. Maxson, and P. Musumeci, "New 1.4 Cell RF Photoinjector Design for High Brightness Beam Generation," p. THA2CO03, Jan. 2017.

[146] H. Rose, *Geometrical Charged-Particle Optics*, ser. Springer Series in Optical Sciences. Berlin Heidelberg: Springer-Verlag, 2009. [Online]. Available: //www.springer.com/us/book/9783642099441

[147] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, 2nd ed. Springer US, 2009. [Online]. Available: https://www.springer.com/us/book/9780387765006

[148] D. E. Groom and S. R. Klein, "Passage of particles through matter," *The European Physical Journal C*, vol. 15, no. 1-4, pp. 163–173, Mar. 2000. [Online]. Available: http://www.springerlink.com/index/10.1007/BF02683419

[149] S. M. Seltzer and M. J. Berger, “Improved procedure for calculating the collision stopping power of elements and compounds for electrons and positrons,” *The International Journal of Applied Radiation and Isotopes*, vol. 35, no. 7, pp. 665–676, Jul. 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0020708X84901133

[150] A. E. Burgess, “The Rose model, revisited,” *JOSA A*, vol. 16, no. 3, pp. 633–646, Mar. 1999. [Online]. Available: https://www.osapublishing.org/josaa/abstract.cfm?uri=josaa-16-3-633

[151] A. Rose, “The Sensitivity Performance of the Human Eye on an Absolute Scale*,” *JOSA*, vol. 38, no. 2, pp. 196–208, Feb. 1948. [Online]. Available: https://www.osapublishing.org/josa/abstract.cfm?uri=josa-38-2-196

[152] R. K. Li, K. G. Roberts, C. M. Scoby, H. To, and P. Musumeci, “Nanometer emittance ultralow charge beams from rf photoinjectors,” *Physical Review Special Topics - Accelerators and Beams*, vol. 15, no. 9, p. 090702, Sep. 2012. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevSTAB.15.090702

[153] K. Halbach, “Design of permanent multipole magnets with oriented rare earth cobalt material,” *Nuclear Instruments and Methods*, vol. 169, no. 1, pp. 1–10, Feb. 1980. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0029554X80900944

[154] P. Elleaume, O. Chubar, and J. Chavanne, “Computing 3d magnetic fields from insertion devices,” in *Proceedings of the 1997 Particle Accelerator Conference (Cat. No.97CH36167)*, vol. 3, May 1997, pp. 3509–3511 vol.3.

[155] Z. Wolf, “A Vibrating Wire System For Quadrupole Fiducialization,” SLAC National Accelerator Lab., Menlo Park, CA (United States), Tech. Rep. SLAC-TN-10-087, Dec. 2010. [Online]. Available: https://www.osti.gov/biblio/1000391

[156] J. K. Lim, P. Frigola, G. Travish, J. B. Rosenzweig, S. G. Anderson, W. J. Brown, J. S. Jacob, C. L. Robbins, and A. M. Tremaine, “Adjustable, short focal length permanent-magnet quadrupole based electron beam final focus system,” *Physical Review Special Topics - Accelerators and Beams*, vol. 8, no. 7, p. 072401, Jul. 2005. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.8.072401

[157] P. W. Hawkes, “Quadrupole Optics,” in *Springer Tracts in Modern Physics, Volume 42*, ser. Springer Tracts in Modern Physics, P. W. Hawkes, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1966, pp. 1–126. [Online]. Available: https://doi.org/10.1007/BFb0045898

[158] M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius, and K. Urban, “A spherical-aberration-corrected 200kv transmission electron microscope,” *Ultramicroscopy*, vol. 75, no. 1, pp. 53–60, Oct. 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399198000485

[159] R. L. Gluckstern, "Scalar Potential for Charge Distributions with Ellipsoidal Symmetry," 1986.

[160] M. Riser, *Theory and Design of Charged Particle Beams*, ser. Wiley series in beam physics and accelerator technology. Wiley, 1994. [Online]. Available: https://www.wiley.com/en-us/Theory+and+Design+of+Charged+Particle+Beams-p-9783527617630

[161] R. Goodall and P. G. L. Leach, "Generalised Symmetries and the Ermakov-Lewis Invariant," *Journal of Nonlinear Mathematical Physics*, vol. 12, no. 1, pp. 15–26, Jan. 2005. [Online]. Available: https://doi.org/10.2991/jnmp.2005.12.1.3

[162] V. Coll, J. Dodd, and M. Harrison, "The Archimedean Projection Property," *arXiv e-prints*, vol. 1504, p. arXiv:1504.02941, Apr. 2015. [Online]. Available: http://adsabs.harvard.edu/abs/2015arXiv150402941C

[163] V. Danilov, S. Cousineau, S. Henderson, and J. Holmes, "Self-consistent time dependent two dimensional and three dimensional space charge distributions with linear force," *Physical Review Special Topics - Accelerators and Beams*, vol. 6, no. 9, p. 094202, Sep. 2003. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.6.094202

[164] O. J. Luiten, S. B. van der Geer, M. J. de Loos, F. B. Kiewiet, and M. J. van der Wiel, "How to Realize Uniform Three-Dimensional Ellipsoidal Electron Bunches," *Physical Review Letters*, vol. 93, no. 9, p. 094802, Aug. 2004. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.93.094802

[165] S. B. van der Geer and M. J. de Loos, "General particle tracer." [Online]. Available: http://www.pulsar.nl/gpt

[166] I. Will and G. Klemz, "Generation of flat-top picosecond pulses by coherent pulse stacking in a multicrystal birefringent filter," *Optics Express*, vol. 16, no. 19, pp. 14 922–14 937, Sep. 2008. [Online]. Available: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-16-19-14922

[167] C. Gulliford and I. Bazarov, "New method for generating linear transfer matrices through combined rf and solenoid fields," *Physical Review Special Topics - Accelerators and Beams*, vol. 15, no. 2, p. 024002, Feb. 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.15.024002

[168] D. Dowell, F. Zhou, and J. Schmerge, "Exact cancellation of emittance growth due to coupled transverse dynamics in solenoids and rf couplers," *Physical Review Accelerators and Beams*, vol. 21, no. 1, p. 010101, Jan. 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.21.010101

[169] D. Marx, J. Giner Navarro, D. Cesar, J. Maxson, B. Marchetti, R. Assmann, and P. Musumeci, "Single-shot reconstruction of core 4d phase space of high-brightness electron beams using metal grids," *Physical Review Accelerators*

and Beams, vol. 21, no. 10, p. 102802, Oct. 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.21.102802

[170] U. Niedermayer, T. Egenolf, and O. Boine-Frankenheim, "Beam dynamics analysis of dielectric laser acceleration using a fast 6d tracking scheme," *Physical Review Accelerators and Beams*, vol. 20, no. 11, p. 111302, Nov. 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.20.111302

[171] J. Joannopoulos, S. Johnson, G., J. N. Winn, and R. D. Meade, *Photonic Crystals*, 2nd ed. Princeton University Press, 2008. [Online]. Available: https://press.princeton.edu/titles/8696.html

[172] L. Sudrie, A. Couairon, M. Franco, B. Lamouroux, B. Prade, S. Tzortzakis, and A. Mysyrowicz, "Femtosecond Laser-Induced Damage and Filamentary Propagation in Fused Silica," *Physical Review Letters*, vol. 89, no. 18, p. 186601, Oct. 2002. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.89.186601

[173] M. Sheik-Bahae, A. A. Said, T. H. Wei, D. J. Hagan, and E. W. V. Stryland, "Sensitive measurement of optical nonlinearities using a single beam," *IEEE Journal of Quantum Electronics*, vol. 26, no. 4, pp. 760–769, Apr. 1990.

[174] S. Kim and M. Potasek, "Numerical Study of Short Optical Pulse Propagation in Nonlinear Reverse Saturable Absorbers," *US AFRL*, p. 26, Feb. 2001.

[175] T. Brabec and F. Krausz, "Intense few-cycle laser fields: Frontiers of nonlinear optics," *Reviews of Modern Physics*, vol. 72, no. 2, pp. 545–591, Apr. 2000. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.72.545

[176] L. V. Keldysh, "IONIZATION IN THE FIELD OF A STRONG ELECTROMAGNETIC WAVE," *Zh. Eksperim. i Teor. Fiz.*, vol. Vol: 47, Nov. 1964. [Online]. Available: https://www.osti.gov/biblio/4662394-ionization-field-strong-electromagnetic-wave