

UCLA

UCLA Electronic Theses and Dissertations

Title

Computational Modeling of Laser Wakefield Acceleration with Spatiotemporally Structured Laser Pulses

Permalink

<https://escholarship.org/uc/item/63w095dm>

ISBN

9798277446812

Author

Pierce, Jacob

Publication Date

2026-03-20

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Modeling of Laser Wakefield Acceleration
with Spatiotemporally Structured Laser Pulses

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

by

Jacob R. Pierce

2025

© Copyright by
Jacob R. Pierce
2025

ABSTRACT OF THE DISSERTATION

Computational Modeling of Laser Wakefield Acceleration with Spatiotemporally Structured Laser Pulses

by

Jacob R. Pierce

Doctor of Philosophy in Physics

University of California, Los Angeles, 2025

Professor Warren B. Mori, Chair

Plasma-based acceleration (PBA) refers to a class of advanced accelerator concepts in which the electromagnetic fields sustained in a plasma wave are used to accelerate charged particle beams. The strength of these fields can be orders of magnitude higher than those achievable in conventional radio-frequency accelerators. Thus, PBA offers the prospect of dramatically reducing the accelerator size and cost for a given beam energy. This makes PBA attractive for a variety of accelerator-based applications including high-energy collider physics, microscopy using free electron lasers, radiotherapy for cancer treatment, and ion implantation in semiconductors. Laser wakefield acceleration (LWFA), in which an intense laser pulse drives the plasma wave, is one of two concepts of interest for PBA. In this dissertation, new results on LWFA and new capabilities for its simulation are presented through a combination of theory and particle-in-cell computer simulations.

In LWFA, the acceleration length and the total energy gain for an accelerating beam are critical system parameters. Lu et al. [1] provided estimates for how these parameters scale in the nonlinear, self-guided regime as a function of basic quantities such as the laser

wavelength, laser power, and plasma density. It was assumed that the laser is self-guided over a length greater than the length over which the beam dephases with respect to the wake, but it was recognized that this assumption may break down for sufficiently low plasma density. New results on this physics are presented in this dissertation. Using the quasistatic particle-in-cell (PIC) code QPAD, a parameter scan of simulations is carried out with a fixed laser vector potential, a laser profile which scales with the plasma skin depth c/ω_p , and a varying ratio of the laser frequency ω_0 to the plasma frequency ω_p . It is found that the Lu et al. scaling laws are accurate for densities above a threshold value, below which the self-guiding length is less than the dephasing length and scales as $L_{\text{guide}} \approx 25z_R$ where z_R is the Rayleigh range. This results in similar scalings for the acceleration length $L_{\text{acc}} \approx 25z_R$ and energy gain $\Delta E \approx 25z_R(\omega_p/c)m_e c^2$ for sufficiently low density. The results indicate that near-term 10 PW lasers may be able to generate 30-GeV electron beams in a single 2-m plasma stage. They also inform longer-term design considerations for LWFA in the context of a linear collider.

Spatiotemporally structured laser pulses, which feature coupled spacetime-dependent properties, have been proposed to enable or enhance a variety of laser-plasma interactions, including LWFA. Many of these methods require an accurate understanding of the linear propagation of the pulse. Previous work on linear pulse propagation has mostly considered the transverse or longitudinal evolution of the laser pulse separately, but these must be considered simultaneously to describe spatiotemporally structured pulses. To address this need, theory is developed to describe the linear evolution of the full spatiotemporal structure of a laser pulse in a linear dielectric. The approach uses the dispersion relation to derive integral expressions for the evolution of the pulse. Taylor expansion of the dispersion relation leads to systematic estimates for length scales over which spatiotemporal coupling emerges for a laser pulse with an envelope which is initially separable transversely and longitudinally. These are derived in vacuum and in cold, uniform plasma. The results are also described from the standpoint of single photons propagating within the pulse. Simulations confirm accuracy

of the integral formulation and its predictions for the length scales over which spatiotemporal coupling emerges.

Exploration of applications of spatiotemporally structured laser pulses requires an understanding of the kinds of spatiotemporal structure that can exist in nature. It is also desirable to be able to describe a given kind of spatiotemporal structure with a mathematical model that captures its essential features. In this dissertation, a general formalism describing spatiotemporally structured laser pulses is introduced which meets both of these demands. This concept of arbitrarily structured laser (ASTRL) pulses employs a superposition of prescribed pulses to create a desired field structure. The ASTRL formalism is capable of describing a variety of kinds of spatiotemporal structure, including pulses with an evolving focal point, spot size, orbital angular momentum, or polarization. Several examples are presented to illustrate the versatility of ASTRL pulses to address a broad range of laser-based applications, including LWFA, inertial confinement fusion, nanophotonics, and attosecond physics.

Laser wakefield acceleration (LWFA) may enable the next generation of TeV-scale lepton colliders. Current designs are based on the use of multiple LWFA stages to overcome limitations on the energy gain achievable in a single stage. The use of stages, however, introduces challenges such as alignment, adiabatic matching between stages, and a lower average accelerating gradient. Flying focus pulses, a specific type of spatiotemporally structured pulse, have been proposed to circumvent dephasing. This would reduce both the number of stages and the total accelerator length required for a target energy by enabling operation at higher plasma densities. Adding to these concepts, LWFA driven by a discrete flying focus based on the ASTRL formalism is proposed. A sequence of laser pulses with staggered focal points and delays drives a plasma wave in which an electron beam experiences a near-constant accelerating gradient over distances beyond those attainable with a conventional pulse. Simulations with non-optimized parameters demonstrate that a discrete flying focus with a total energy of 150 J can transfer 40 GeV per electron to a 50-pC beam in a single

30-cm stage, corresponding to 50 dephasing lengths. Extension of the concept to positron acceleration through the use of a discrete set of higher-order Laguerre-Gaussian pulses is also presented.

Fully relativistic particle-in-cell (PIC) simulations are crucial for understanding a variety of processes in plasma physics, including LWFA. Modern supercomputers based on graphics processing units (GPUs) offer the potential to perform PIC simulations of unprecedented scale, but require robust and feature-rich codes that can fully leverage their computational resources. To address this demand, GPU acceleration was implemented in the PIC code OSIRIS. An overview of the algorithm, which features a CUDA extension to the underlying Fortran architecture, is given. Detailed performance benchmarks for thermal plasmas are presented, which demonstrate excellent weak scaling on NERSC’s Perlmutter supercomputer and high levels of absolute performance. The robustness of the code to model a variety of physical systems is demonstrated via simulations of Weibel filamentation and laser-wakefield acceleration run with dynamic load balancing. Finally, measurements and analysis of energy consumption indicate that the GPU algorithm is up to 14 times faster and 7 times more energy efficient than the optimized CPU algorithm on a node-to-node basis. The described development addresses the PIC simulation community’s computational demands both by contributing a robust and performant GPU-accelerated PIC code and by providing insight into efficient use of GPU hardware.

The dissertation of Jacob R. Pierce is approved.

E. Paulo Alves

Troy A. Carter

Chandra J. Joshi

Warren B. Mori, Committee Chair

University of California, Los Angeles

2025

For my grandfather, Raymond Barile.

TABLE OF CONTENTS

1	Introduction	1
1.1	Laser wakefield acceleration: a brief overview	1
1.2	Laser wakefield acceleration: a brief history	4
1.3	Overcoming dephasing in LWFA with spatiotemporally structured laser pulses	10
1.4	Arbitrarily Structured Laser Pulses	12
1.5	Methodology: the particle-in-cell simulation algorithm	13
1.6	Dissertation outline	16
2	Laser Wakefield Acceleration in the Blowout Regime: Theory and Computational Methods	18
2.1	Kinetic description of plasma	19
2.2	Kinetic description of LWFA	23
2.3	The ponderomotive guiding center approximation	25
2.3.1	Particle evolution	27
2.3.2	Laser evolution	28
2.3.3	Wakefield evolution	30
2.3.4	Summary	31
2.4	The quasistatic approximation	31
2.4.1	Particle evolution	32
2.4.2	Laser evolution	34
2.4.3	Wakefield evolution	34
2.4.4	Summary	38

2.5	Dimensionless quantities	38
2.6	Structure of the plasma wake	40
2.7	Laser evolution and its consequences: matching, depletion, and dephasing . .	53
2.8	Beam loading	57
2.9	Witness beam evolution	61
2.10	Simulation of laser wakefield acceleration in the blowout regime	66
2.10.1	The particle-in-cell algorithm	66
2.10.2	The ponderomotive guiding center PIC algorithm	69
2.10.3	Quasistatic PIC algorithms	70
2.11	Summary	71
3	On the Scaling Laws for the Acceleration Length and Energy Gain in the Blowout Regime of LWFA in Uniform Plasma	72
3.1	Introduction	73
3.2	Results	75
3.2.1	Simulation Setup	75
3.2.2	Evolution of the laser and wake	78
3.2.3	Particle Acceleration	83
3.3	Discussion and conclusions	87
4	Linear Propagation of Laser Pulses in Cold, Uniform Plasma	92
4.1	Integral solutions to linear partial differential equations using dispersion relations	94
4.1.1	Formulation in terms of the envelope	96
4.1.2	Phase and group velocity	97
4.1.3	The choice of t versus z as the propagation variable	97

4.2	Linear propagation of laser pulses in vacuum	98
4.2.1	First-order propagation	99
4.2.2	Second-order propagation: diffraction and paraxial beams	100
4.2.3	Third-order propagation: lowest-order spatiotemporal coupling	102
4.2.4	Comparison between theory and simulation	103
4.2.5	Geometric understanding of the length-scale for spatiotemporal coupling	105
4.3	Linear propagation of laser pulses in cold, uniform plasma	106
4.3.1	First-order propagation	106
4.3.2	Second-order propagation: group-velocity-delayed paraxial beams and dispersive longitudinal broadening	107
4.3.3	Plasma-induced local frequency evolution	108
4.3.4	Third-order propagation: vacuum-like spatiotemporal coupling and additional longitudinal broadening	109
4.3.5	Comparison between theory and simulation	110
4.4	Recovery of Envelope PDEs from the Integral Formulation	110
4.5	Summary	112
5	Arbitrarily Structured Laser Pulses	113
5.1	Introduction	114
5.2	Theoretical formalism	115
5.3	Design of ASTRL pulses	118
5.4	Closed-form ASTRL pulses	123
5.5	Experimental approaches to synthesizing ASTRL pulses	129
5.6	Accuracy of Equation 5.6	132

5.7	Length scale for emergence of spatiotemporal coupling	133
5.8	Separable envelope decomposition of arbitrary laser fields	134
5.9	Discussion	135
6	Laser Wakefield Acceleration with Arbitrarily Structured Laser Pulses .	138
6.1	Introduction	139
6.2	Theory	141
6.3	Simulations: electron acceleration	149
6.3.1	Linear Regime	149
6.3.2	Nonlinear Regime	152
6.3.3	Acceleration in the Nonlinear Regime	155
6.4	Higher-Order Effects of Linear Laser Propagation	158
6.5	Simulation details	159
6.6	Comparison of nonlinear LWFA in QPAD and OSIRIS	160
6.7	Simulations: positron acceleration using an $\ell = 1$ Laguerre-Gaussian ASTRL driver	164
6.7.1	Linear Regime	165
6.7.2	Nonlinear Regime	165
6.8	Discussion	169
7	Acceleration of the Particle-in-Cell Code OSIRIS with Graphics Processing Units	172
7.1	Introduction	173
7.2	Methodology	175
7.2.1	GPU programming framework	175

7.2.2	Tile-based domain decomposition	176
7.2.3	Particle chunk pool	177
7.2.4	Particle pushing and current deposition	180
7.2.5	Particle sorting and boundary conditions	182
7.3	Results	184
7.3.1	Thermal plasma benchmarks	185
7.3.2	Simulations of more complex plasma systems	188
7.3.3	Energy efficiency	191
7.4	Discussion	193
8	Summary	196
A	Detailed Guide to Running Lorentz-Boosted Simulations of LWFA . . .	201
A.1	Introduction	201
A.1.1	What is the maximum γ that can be applied?	203
A.1.2	When is a Lorentz-boosted simulation beneficial?	203
A.1.3	Brief review and current status in the field	204
A.2	Initialization of lorentz-boosted simulations	205
A.2.1	Field solvers	205
A.2.2	Boundary conditions	206
A.2.3	Simulation box boundaries	206
A.2.4	Resolution: Δ'_z , Δ'_x , Δ'_t , and particles per cell	208
A.2.5	Estimation of the simulation cost	210
A.2.6	Background plasma initialization	210

A.2.7	Electromagnetic field initialization (including lasers)	212
A.2.8	Particle beam initialization	215
A.3	Boosted diagnostics	218
A.3.1	Code details	219
A.3.2	Other applications of boosted diagnostics	220
References		222

LIST OF FIGURES

1.1	Visualization of laser wakefield acceleration in the nonlinear blowout regime. A laser pulse (red) injected into a plasma drives a wake in the electron plasma density (blue). The force field within the plasma wake (light gray arrows) for electrons moving forward near the speed of light simultaneously accelerates and focuses the electrons. Accelerating electrons can be either externally injected into the system (green) or “self-injected” from the back of the plasma wake (dark blue). The wavelength of the wake is on the order of $10c/\omega_p$. This data was generated from a 3D particle-in-cell simulation using the code OSIRIS [2].	2
1.2	A flying focus laser pulse. The laser pulse features two unusual properties: (1) a controllable range of propagation which exceeds the Rayleigh range and (2) controllable envelope velocity which may be different from the speed of light. In this case, the flying focus is created by sending a chirped laser pulse through a lens with a large chromatic aberration. The data, measurements from an experiment of this optical configuration, show a flying focus with a subluminal velocity propagating over about 100 Rayleigh ranges. Figure from Ref. [3]. Used with permission from Springer Nature.	11
1.3	Schematic depicting a single time step in the fully electromagnetic particle-in-cell (PIC) algorithm. Used with permission from Ref. [4].	14

- 2.1 3D PIC simulation of LWFA initialized according to Eqs. 2.22 and 2.23. The laser electric field is shown in red, an external electron beam in green, and the number density of the plasma electrons in blue. (a) The initial condition at $\omega_p t = 0$. (b) The state of the system at $\omega_p t = 40$. In this case, $k_0/k_p = 10$, $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 2$. The piecewise-polynomial profile used in Ref. [1] is used for the $g(z)$ function in Eq. 2.23. The parameters for this system are the same as for the case shown in Fig. 1.1. The simulation box moves forward at the speed of light. 24
- 3.1 Snapshots of the laser envelope and electron charge density at densities ranging from $n_0 = 1 \times 10^{19} \text{ cm}^{-3}$ down to $n_0 = 2 \times 10^{16} \text{ cm}^{-3}$ at a fixed $z = 0.7L_d$, where L_d depends on the density for each case. If the evolution of the laser and wake depended on the dimensionless parameter z/L_d , then all the images would appear the same. This is approximately true for the 5 cases with $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$. For the cases with $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$, it is evident that $L_{\text{guide}} < L_d$ and the self-guiding length follows a weaker scaling with density than $L_{\text{guide}} \propto L_d \propto n_0^{-1}$. 79
- 3.2 Snapshots of the laser envelope and electron charge density as in Fig. 3.1, but at fixed $z = 20z_R$. The cases with $n_0 \geq 1 \times 10^{18} \text{ cm}^{-3}$ are fully depleted at this stage. On the other hand, the three lowest-density cases are nearly indistinguishable in structure. This suggests that $L_{\text{guide}} \propto z_R$ at these lower densities. 80
- 3.3 On-axis longitudinal electric field in normalized units for cases ranging from $n_0 = 1 \times 10^{19} \text{ cm}^{-3}$ down to $1 \times 10^{15} \text{ cm}^{-3}$. (a) At $z = 0.7L_d$, it is clear that the normalized wake amplitude decreases with density. The constancy of the position of the zero-crossing indicates that $\beta_w - 1 \propto \left(\frac{\omega_p}{\omega_0}\right)^2$ as in Lu et al. (b) At $z = 20z_R$, the amplitude of the cases with $n_0 \leq 5 \times 10^{17} \text{ cm}^{-3}$ is nearly identical. This is consistent with the scaling $L_{\text{guide}} \propto z_R$ at these lower densities. 81

- 3.4 Evolution of the maximum of $e|\hat{E}|/mc\omega_0$ for cases ranging from $n_0 = 1 \times 10^{19} \text{ cm}^{-3}$ down to $1 \times 10^{15} \text{ cm}^{-3}$. (a) When plotted against z/L_d , it is clear that the cases with $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$ have $L_{\text{guide}} \approx 1.5L_d$ and exhibit similar dynamics on this length scale. The cases with $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$ do not propagate as far as L_d follow a different scaling. (b) When plotted against z/z_R , we see that $L_{\text{guide}} \approx 25z_R$ for $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$. The fit is the expected evolution under vacuum diffraction with a pulse with an effective Rayleigh range $z_{R,\text{eff}} = 12.5z_R$. 82
- 3.5 Illustration of the beam (gray) and the slice of the beam selected (white box) for examining the energy gain of each simulation. The same slice can be used because the structure of the wake at early time is nearly identical. A more sophisticated analysis using the ξ slice of maximum energy gain for each simulation yielded similar results. The simulation shown is the $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$ case at $z = 0.5z_R$. 84
- 3.6 (a) Evolution of the maximum Lorentz factor for particles in the region $k_p\xi \in [-11, -10.9]$ and $k_pr \in [0, 1]$. The energy gain for $n_0 \geq 2 \times 10^{17} \text{ cm}^{-3}$ is approximately $\Delta\gamma = 1.2\Delta\gamma_{\text{Lu}}$. For $n_0 \leq 5 \times 10^{16} \text{ cm}^{-3}$, the acceleration length and energy gain are significantly lower than the predictions from Lu et al. (b) Beam current for electrons in this region. The data in (a) are plotted up to the lowest distance at which $\Lambda \leq \Lambda_0/2$ 85
- 3.7 The maximum energy gain $\Delta E \equiv \max_z \Delta\gamma(z)mc^2$ and acceleration length $L_{\text{acc}} \equiv \arg \max_z \Delta\gamma(z)$ are extracted from the data in Fig. 3.6 and respectively plotted in (a) and (b). (a): the energy gain is accurately modeled by Lu et al. for $n_0 \geq 5 \times 10^{16} \text{ cm}^{-3}$. $\Delta\gamma = 25k_pz_R$ models the data well for $n_0 \leq 2 \times 10^{16} \text{ cm}^{-3}$. (b): the Lu et al. model is almost exact for the acceleration length for $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$, within a factor of 2 down to $n_0 = 5 \times 10^{16} \text{ cm}^{-3}$. The acceleration length is more accurately described by the model $L_{\text{acc}} = 25z_R$ for $n_0 \leq 2 \times 10^{16} \text{ cm}^{-3}$ 86

4.1	Comparison between theory and simulation and illustration of the effect of propagation beyond τ_{xxz} . (a) $ A $ for the initial condition at $ct = -5z_R$. (b-e) Solutions at $ct = 5z_R$ calculated by simulation in OSIRIS, integrating using the full $\omega(\mathbf{k})$ (Eq. 4.10), integrating the third-order approximation (Eq. 4.46) and integrating using the second-order approximation (Eq. 4.36). Significant curvature in the true solution is consistent with the estimate $\Delta t = 2\tau_{xxz}$. (f-k) Analogous plots for a laser pulse with a duration 5 times larger, so that $\Delta t = 0.4\tau_{xxz}$. In this case, the pulse structure remains approximately separable, consistent with theory.	104
4.2	Geometry leading to the development of spatiotemporal coupling as evident in Fig. 4.1	106
4.3	Demonstration of spatiotemporal coupling and additional longitudinal broadening induced by propagation in plasma in 3D. The initial condition and propagation length are the same as in the vacuum cases shown Fig. 4.1, but now a plasma is added such that $k_0/k_p = 10$. ($k_p = 0$ in Fig. 4.1). (a-e) The effects of plasma-induced dispersion manifest as broadening of the longitudinal envelope such that the bow structure in vacuum becomes a triangular structure in plasma. The agreement between OSIRIS and the FFT with the full $\omega(\mathbf{k})$ supports the theory, as in the vacuum case. (f-k) In this case plasma-induced dispersion is negligible; the evolution looks just as in vacuum except for the recession of the pulse with respect to ξ because $v_g < 1$	111

- 5.1 (a): On-axis evolution of the envelope of an ASTRL flying focus pulse found by evaluating Eq. 5.7 with $k_0 w_0 = 40$, $\omega_0 \tau_0 = 20$, and $\beta_f = 1.05$. The dashed line indicates the trajectory $z = v_f t$. (b) The envelope from a quasi-3D FDTD simulation for the same parameters. The two are in excellent agreement. (c) – (e): A generalized flying focus constructed using the ASTRL formalism in which the focal point oscillates transversely as it moves along the optical axis. The propagation was simulated in 2D using the FDTD method. The pulse was initialized at $ct/z_R = -10$ using Eq. 5.7 with $x_{\perp \text{osc}} = w_0$, $k_{\text{osc}} = 1/z_R$, $\beta_f = 1.05$, $L = 20z_R$, $k_0 w_0 = 40$, and $\omega_0 \tau_0 = 20$. From left to right, the frames show the pulse at $ct/z_R = -4, 0$, and 4 119
- 5.2 Examples of ASTRL pulses constructed from superpositions of laser pulses focused to the same point ($s_0(\eta) = 0$): (a) A pulse everywhere linearly polarized, but with a time-dependent polarization angle. The duration of each pulse in the superposition is $\omega_0 \tau_0 = 20$. (b) A laser pulse with a spot size that continuously decreases in time with $k_0 w_{0\text{min}} = 40$, $w_{0\text{max}} = 4w_{0\text{min}}$, $\omega_0 \tau_0 = 40$, and $\Delta_{\text{max}} = 20c\tau_0$. (c) Time slices of the transverse electric field of a laser pulse whose angular momentum continuously varies in ξ . (d) A lattice of STOVs with $k_0 w_0 = 40$ and $\omega_0 \tau_0 = 20$. All examples are evaluated using Eq. 5.7 and shown at the focal point $s = 0$. . . 121

- 5.3 QPAD simulations illustrating the spatiotemporal structure of flying focus ASTRL pulses described the closed-form Eq. 5.12, as well as differences in propagation resulting from the dependence of the duration on the normalized focal velocity β_f (Eq. 5.17). (a) A flying focus with $\beta_f = 1.05$ at $t = 0$. (b) Final state at $c\Delta t = 10z_R$. The state is mirror-symmetric (with minor deviations) because it is described accurately by Eq. 5.12. In this case, $L_{\text{stc}} = 80z_R \gg c\Delta t$. (c) A flying focus with the same focal range but with $\beta_f = 1.005$. The structure is the same as in (a) except for compression by a factor of $10\times$ in the longitudinal direction. (d) At $c\Delta t = 10z_R$, there is significant distortion from Eq. 5.17, consistent with the estimate $L_{\text{stc}} = 8z_R \lesssim c\Delta t$. Pulse parameters are in the text. 128
- 5.4 An experimental concept to produce new classes of ASTRL pulses. Each pulse is independently manipulated by a separate optical assembly, labeled “OA_{*j*}”, which can alter the amplitude, transverse profile, frequency, delay, or other properties. The pulses are then coherently combined to form an ASTRL pulse. 130
- 5.5 (a,b) Discrete approximations of Eq. 5.7 for the $\beta_f = 1.05$ flying focus pulse shown in Fig. 1 with $N = 10$ and $N = 20$ pulses, respectively; the continuum is well approximated by 20 pulses, or 2 pulses per z_R . (c) Peak on-axis amplitude for the $N = 10$ and 20 cases, showing oscillation for $N = 10$. (d) Relative amplitude oscillation, defined as $(\max |A| - \min |A|)/(\max |A| + \min |A|)$ over the interval $[-z_R, z_R]$. The relative oscillation decreases exponentially with N , which determines the number of pulses required for a given tolerance. 131

5.6 (a,b): On-axis evolution of the vector envelope for a negative-velocity flying focus with structured polarization. The components of the vector envelope $\mathbf{A}$ are defined by $A_j \equiv \hat{\mathbf{e}}_j \cdot \int d\eta B_\eta(\xi) C_\eta(\mathbf{x}_\perp, s) \hat{\mathbf{e}}(\eta)$. The parameters are identical to those in Fig. 5.1 except $\beta_f = -1$ and the polarization evolves continuously. Specifically, $\hat{\mathbf{e}}_x(\eta) = \cos(k_{\text{osc}}\eta L)$ and $\hat{\mathbf{e}}_y(\eta) = \sin(k_{\text{osc}}\eta L)$, where $k_{\text{osc}} = (3/2)z_R^{-1}$. The white dashed lines indicate the focal trajectory $z = -ct$. Both plots are evaluated using Eq. 5.7. 136

6.1 Comparison of LWFA with a conventional pulse and a discrete flying focus in the linear regime. (a) A single pulse traveling at its group velocity (black) drives a plasma wave (gray), whose longitudinal field E_z accelerates a highly relativistic electron (cyan). The pulse and longitudinal field slip backward in $\xi = z - ct$ as the pulse propagates from $z = 0$ to $2L_d$. As a result, the electron moves from an accelerating to a decelerating longitudinal field, which limits its energy gain—a process referred to as dephasing. (b) A sequence of pulses (red, green, blue) with appropriately selected delays and focal points drives a plasma wave (gray) whose longitudinal field remains stationary in ξ . This eliminates dephasing by ensuring that the electron (cyan) remains collocated with the maximum accelerating field over the entire distance. In both (a) and (b), the field of the laser pulse is calculated using Eq. (6.7), and the longitudinal field is calculated using linear theory [5, 6]. The specific parameters, chosen for illustration, are $k_0/k_p = 40$, $k_p\tau_0 = 0.5$, $f_j = jL_d$, and $\Delta_j = (1 - v_g/c)f_j = (\frac{\pi}{2})jk_p^{-1}$. The dephasing length is $k_pL_d = 1600\pi$, the single pulse in (a) has $k_pz_R = \infty$, and each pulse in (b) has $k_pz_R = 320$ 142

6.2 Evolution of a discrete flying focus and driven plasma wave in the linear regime ($a_0 = 0.1$). (a) The on-axis ($|\mathbf{x}_\perp| = 0$) envelope of the DFF pulse (red and blue) and longitudinal electric field of the plasma wave (black and gray) have nearly identical profiles in $\xi = z - ct$ after the pulse has propagated 5 (solid) and 20 (dashed) dephasing lengths. (b,c) The spatiotemporal structure of the DFF envelope (top) and longitudinal field (bottom) are also nearly identical at these propagation distances. From left to right in ξ , the variation in the transverse width of the DFF (i.e., the X structure) results from ξ slices that have already passed through focus, are at or near focus, and have yet to come into focus, respectively. This slice-dependent focusing is also the source of the pre and postcursors on the envelopes in (a). In this example, $N = 250 = N_{\min} \gg N_{\text{overlap}}$. Due to significant overlap between individual pulses, the spot size of the overall envelope $k_p w_{\text{eff}} = 9$ is larger than that of the individual pulses $k_p w_0 = 4$ 150

- 6.3 Dependence of the on-axis ($|\mathbf{x}_\perp| = 0$) longitudinal field E_z on the number of pulses driving the plasma wave in the (a,b,c) linear and (d,e,f) nonlinear regimes. (a) With a single pulse ($N = 1$) in the linear regime ($a_0 = 0.1$), the maximum accelerating field (dashed black line) recedes and drops in amplitude due to the subluminal group velocity and diffraction of the pulse. (b) For a DFF composed of $N = 40 < N_{\min} = 250$ pulses, the plasma wave persists over the entire propagation range of 25 dephasing lengths, but the large separation between pulses results in an irregular longitudinal field. (c) A DFF with $N = 250 = N_{\min}$ pulses drives a plasma wave with a smooth longitudinal field and a phase velocity $v_w = c$. (d) With a single pulse ($N = 1$) in the nonlinear regime ($a_0 = 4$), the maximum accelerating field (dashed black line) recedes and drops in amplitude due to a combination of the subluminal group velocity and depletion. (e,f) Similar to the linear regime, a DFF composed of too few pulses ($N = 20 < N_{\min} = 50$) produces an irregular field, while a DFF composed of a sufficient number of pulses ($N = 40 \approx N_{\min}$) produces a smooth field. 153
- 6.4 Acceleration of a 50 pC electron beam from 1 GeV to 41 GeV over 32 cm, or 50 dephasing lengths, using a discrete flying focus with tailored delays. (a) Electron density of the background plasma and electron beam normalized to n_0 . The beam profile and accelerating structure are preserved over the entire length of the accelerator. Note that the beam density saturates the color map. (b) Mean slice energy of the beam. At the end of the accelerator ($z = 32$ cm), the spectrum is nearly monoenergetic with an energy spread $\sigma_\gamma/\bar{\gamma} = 0.04$. (c) Normalized slice emittance. The emittance does not grow as the beam is accelerated. 156

- 6.5 The electron densities and electric field envelopes from (left) QPAD and (right) OSIRIS simulations run with the same physical parameters. The results for a single pulse ($N = 1$) are shown in (a) and (b), while the results for a DFF with $N = 9$ pulses are shown in (c) and (d). The agreement between QPAD and OSIRIS is excellent. In (c) and (d), local depletion and redshifting causes the light to slip backward, where it is temporarily trapped by the plasma wave. This prevents the rear sheath from closing. Over longer distances, the sheath undergoes multiple cycles of opening and closing as the trapped light first accumulates and then escapes out the back. [See Figs. 6.3(e) and (f)]. The quasistatic model used in QPAD does not capture self-injection, which is observed in the OSIRIS simulations (d). 162
- 6.6 Linear wakefields driven by a flying focus described by an ASTRAL flying focus with $C_j(z, \mathbf{x}_\perp)$ given by an $(\ell, p) = (1, 0)$ Laguerre-Gaussian mode, $k_p w_0 = 4$ for each pulse, $k_p L = 20,000 \approx 10L_d$, and $N = 80$. Δ_j and f_j are chosen in the same way as for the electron acceleration cases. Top panels show the state at $z = 5000 \approx 2.5L_d$ and $z = 15,000 \approx 7.5L_d$. (a) Amplitude of the laser envelope for the $m = 1$ mode. The laser envelope peaks off-axis and retains the same structure throughout its propagation. The characteristic x-shape and precursor are visible, as with an $(\ell, p) = (0, 0)$ profile. (b-d) $m = 0$ mode of the electron number density, focusing field, and accelerating field. The structure is maintained over the 5 dephasing lengths for each quantity. However, this configuration cannot accelerate positrons because the focusing and accelerating fields vanish on the axis. 166

6.7	Nonlinear wakefields driven by a flying focus described by an ASTRL flying focus with an $(\ell, p) = (1, 0)$ Laguerre-Gaussian transverse profile for each pulse. Each pulse has $a_{0j} = 20$ and $k_p w_0 = 4$. Alternating polarization is used. Top panels show the state at $z = 5000 \approx 2.5L_d$ and $z = 15,000 \approx 7.5L_d$. (a) Amplitude of the laser envelope for the $m = 1$ mode. The evolution is significantly more complex than in the linear regime and appears turbulent. (b) Unlike in the linear regime, plasma electrons are pushed by the laser onto the axis. Multiple off-axis wakes are visible. (c-d) The blue region for the focusing field and red region of the accelerating field near $k_p \xi = -20$ are favorable for acceleration of positrons. This structure is maintained throughout the propagation of the laser.	167
6.8	Time evolution of lineouts from the simulation in Fig. 6.7 for (a) the accelerating field on-axis ($k_p r = 0$) and (b) the focusing field at $k_p r = -20$. The accelerating and focusing fields persist over 10 dephasing lengths. Recession of the wake is visible in the E_z lineout. The discreteness of the $N = 16$ pulses is evident. . . .	168
7.1	Schematic of the CPU memory layout, chunk pool data structure, and species chunk manager. The CPU species object stores contiguous arrays of particle position relative to nearest cell, momentum, charge, and cell index ($\mathbf{x}$, $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{ix}$) for each tile. The species chunk manager, whose state is maintained on both CPU and GPU, stores pointers to chunks of device memory for each tile. Chunks contiguously store all four data fields for a fixed number of particles. Each species chunk manager may transfer pointers to chunks to and from the chunk pool with nearly zero overhead, enabling arbitrary load-balancing between tiles with nearly zero overhead. Particle transfer between the CPU species object and species chunk manager is achieved through an object for buffered data movements, discussed in Section 7.2.2.	178

7.2	Comprehensive thermal plasma benchmarks run on Perlmutter. (a): Weak scaling tests for the cases described in the text, showing near-perfect scaling up to 4096 GPUs on Perlmutter. (b) Parameter scan showing dependence of throughput on particles per cell and grid sizes. (c) Comparison of absolute performance on one GPU and one core of one CPU for different interpolation orders. All simulations are run in 3D with quadratic interpolation and 512 particles per cell unless otherwise specified.	185
7.3	Examples of spatially inhomogeneous plasma simulations run on Perlmutter. (a): 3D simulation of the Weibel instability on 4096 GPUs, demonstrating the tangling of magnetic field lines. (b): charge density from a 2D simulation of LWFA in the nonlinear regime with dynamic load balancing. The black lines indicate the computational boundaries of different MPI ranks.	189
7.4	Energy consumption per particle push for the GPU algorithm and for the CPU algorithm. The comparison is between a simulation on a Perlmutter GPU node with 4 GPUs and a Perlmutter CPU node with 128 CPU cores. The GPU node is ~ 14 times faster than the CPU node and uses ~ 2 times more energy, leading to an overall energy efficiency ~ 7 times higher. Further detail of what is being measured is in the text.	192

7.5	Pictorial explanation of an idea to improve performance for computational problems featuring a stationary plasma in a moving window, as in laser wakefield acceleration. The black box is the computational domain, moving from left to right, and the gray numbered boxes are tiles. In the current algorithm, the tiles move with the moving window. This results in a stationary plasma particle, shown in red, effectively moving across the tiles. In an improved algorithm, stationary tiles could be used. In this case, the tiles are stationary in space and drift with respect to the computational domain. This would in principle make the performance for drifting plasmas the same with and without a moving window. Once the computational domain passes the leftmost tile, that tile can be repositioned on the right side of the domain.	195
-----	---	-----

LIST OF TABLES

1.1	A brief chronology of research in laser wakefield acceleration and plasma simulation showing development of theory, simulation, and experiment. The emphasis is on work related to this dissertation. Years correspond to the date of publication. The final three entries marked with a * correspond to Chapters 5, 6, and 7 in this dissertation.	6
2.1	Summary of how quantities are normalized according to electron plasma parameters.	39
3.1	Summary of the relevant physical and dimensionless quantities associated with the simulations. For each simulation, $k_p w_0 = 4$ and $k_p \tau_{\text{rise}} = k_p \tau_{\text{fall}} = 4$. Values are rounded to 2 significant digits for legibility. The simulations used more precision for the ratio ω_0/ω_p , which was calculated from the exact values of n_0 in the first column. Any row within the table can be rescaled to a different laser wavelength and plasma density with the same ω_0/ω_p	77
3.2	Tabulation of the simulated values for the energy gain and acceleration length as plotted in Fig. 3.7. Other simulation parameters can be cross-referenced with Table 3.1.	87
6.1	Parameters for the simulations presented in Figs. 6.2 and 6.3. The normalized vector potential of each pulse a_{0j} is scaled to achieve the desired a_0 for the overall envelope. Note that for the profile choices defined in Eqs. (6.8) and (6.9), $a_{0j} = \mathcal{A}_{0j}$. The values for $f_{j+1} - f_j$ and $\Delta_{j+1} - \Delta_j$ are respectively listed for the $N > 1$ cases. Time and space are normalized by ω_p^{-1} and k_p^{-1} where indicated. .	151
7.1	Absolute throughput in gigaparticles per second for the GPU and CPU algorithms on one GPU and one core of one CPU with different interpolation orders. The benchmarks are measured on a thermal plasma.	188

ACKNOWLEDGMENTS

This dissertation owes its existence and its particular character to a large number of people. Several have had very direct contributions. Probably the most significant is my advisor Warren Mori, whom I acknowledge for funding my doctoral research, finding problems for me to work on, always asking me to give my best effort, putting an enormous amount of effort into reviewing and providing feedback on this dissertation, and giving me space to work independently. John Palastro, Kyle Miller, and Roman Lee also had significant and direct roles in this dissertation. They collectively provided major contributions to the work and writing of Refs. [7], [8], and [9], which are reproduced and extended in Chapters 5, 6, and 7. Along with Warren, John, Kyle, Roman, and myself, these works were also collectively co-authored by Fei Li, Bernardo Malaca, Dillon Ramsey, George Vieira, Kale Weichman, Maria Almanza, Adam Tableman, Viktor Decyk, Ricardo Fonseca, and Paulo Alves; each made important intellectual contributions. I am grateful to have had the privilege of working with all of you. Several institutions also played a direct role. I acknowledge the U.S. Department of Energy, the National Science Foundation, and the Laboratory for Laser Energetics for funding this research through grants. Also, most of the simulations in this dissertation were run on supercomputers at the National Energy Research Scientific Computing Center (NERSC). I acknowledge these resources, as well as Frank Tsung and Warren for obtaining allocations on them.

Many others had contributions which were less direct but still significant. The work in this dissertation builds on previous work by countless researchers, especially that of Asher Davidson and Wei Lu for the work in Chapter 3. The simulations in this dissertation were made possible by the developers of OSIRIS and QPAD. Chapter 3 probably would not have been written without the encouragement of Chan Joshi. Troy Carter took time out of his vacation to serve on the committee for my thesis defense! The many scientists I encountered during my doctoral studies, especially Warren, John, Paulo, Frederico Fiuza, Nuno Loureiro,

and Bill Dorland, taught me much about plasma physics, science, and life in general. What I learned from them is reflected in these pages. The same can be said of my peers at UCLA. Madox McGrae-Menge and I shared many fruitful discussions on plasma physics and rock climbing. Roman Lee and I spent many hours together in the Meyers Room debugging code, discussing the big questions, and occasionally even doing physics. Roman also contributed to my personal development by taking me down far steeper ski terrain and up significantly more hazardous climbing than I had realized I was signing up for. My roommates Casey Barkan and Tyler Carbin provided support, feedback, and inspiration. I am grateful to have known and worked with many other peers at UCLA, including Thamine Dalichaouch, Qianqian Su, Lance Hildebrand, Yujian Zhao, Sarah Chase, Zack Pine, Maria Almanza, Gabi Guttormsen, Vijay Patel, Elias Hansen, Albert Countryman, Abijah Simon, Rachel Kaufman, and Jack Lubin.

The many great teachers and mentors I have had throughout my life have also had an indirect role in this dissertation. I am blessed to have had many great teachers of math and physics. I acknowledge in particular my high-school math teacher David McCord, who taught me not only a great deal of math but also how to explore and enjoy math. Scott Wakely, Cheng Chin, Guy Savard, and Damiano Caprioli all gave me the opportunity to work in their research groups while I was an undergraduate. While working in Scott's lab, I was enormously fortunate to have been taught to program in C and C++ by Remy Prechelt. This was extremely helpful during my doctoral research. Remy also converted me to mechanical keyboards and the Workman keyboard layout, both of which have kept my fingers healthy while typing over the last decade. I also acknowledge those, especially my teachers from high school and T. S. Eliot, who have helped me try to learn to use words (particularly in English).

This work would not have been possible without the love and support of my friends and family. I acknowledge my mother Mary in particular for always supporting me, especially during my most difficult times. I also acknowledge my father Scott for his support and my

step-father Mark for generally being solid. All of my grandparents, especially my grandfather Raymond Barile, were supportive of my journey in physics. My brother Alex has been a constant inspiration for me and a living reminder of the beauty and necessity of humor. I thank Emilio Gomez, Irene Manahan, Rama Bauer, Andrea Arias, Mark Villano, Gilbert Martinez, Br. Joseph, and Melissa Maricich for their support and encouragement. Noah Rennert and Alex Smith helped me wrestle with questions of meaning and purpose which inevitably arose throughout this work.

I also wish to acknowledge several less visible groups and people. A number of people in administrative roles, especially Andrew Soderlund, Michelle Phan, Stephanie Krilov, and Craig Murayama, helped with reimbursements and finicky ethernet ports. I acknowledge those working in facilities management and custodial services for the Physics and Astronomy Building, especially Jules Berger, for keeping the building running and clean. I also acknowledge the architects, builders, and groundskeepers of UCLA for their work in creating and maintaining the beautiful campus. The beauty of the campus has been one of the highlights of my time at UCLA and I am fairly confident that it is reflected in this dissertation.

Finally, I must acknowledge the cause of existence, without which this dissertation and I myself certainly could not exist. The Abrahamic religions hold that this cause is an immanent and transcendent person. I acknowledge this possibility.

VITA

- 2019–2026 Graduate Student Researcher and Teaching Assistant, University of California, Los Angeles, Los Angeles, CA
- 2020 MA in physics, University of California, Los Angeles, Los Angeles, CA
- 2019 BA physics with honors and BS in mathematics, University of Chicago, Chicago, IL

PUBLICATIONS

M. McGrae-Menge, J. R. Pierce, and E. P. Alves. Embedding physical symmetries into machine-learned reduced plasma physics models via data augmentation. *Physical Review Research* 8, 013200 (2026).

J. R. Pierce, K. G. Miller, F. Li, J. P. Palastro, and W. B. Mori. Laser wakefield acceleration driven by a discrete flying focus. *Physical Review Accelerators and Beams* 28, 101303 (2025).

R. P. Lee, J. R. Pierce, K. G. Miller, M. Almanza, A. Tableman, V. K. Decyk, R. A. Fonseca, E. P. Alves, and W. B. Mori. Acceleration of the particle-in-cell code OSIRIS with graphics processing units. *Journal of Plasma Physics* 91 (1), E8 (2025).

K. Weichman et al. Analytic Pulse Technique for Computational Electromagnetics. *Computer Physics Communications* 298, 109096 (2024).

D. Li, K. G. Miller, J. R. Pierce, W. B. Mori, A. G. R. Thomas, and J. P. Palastro. Spatiotemporal control of high-intensity laser pulses with a plasma lens. *Physical Review Research* 6 (1), 013272 (2024).

B. Malaca et al. Coherence and superradiance from a plasma-based quasiparticle accelerator. *Nature Photonics* 18 (1), 39-45 (2024).

K. G. Miller, J. R. Pierce, M. V. Ambat, J. L. Shaw, K. Weichman, W. B. Mori, D. H. Froula, and J. P. Palastro. Dephasingless Laser Wakefield Acceleration in the Bubble Regime. *Scientific Reports* 13 (1), 21306 (2023).

X. Xu et al. Generation of Ultrabright and Low Energy Spread Electron Beams in Laser Wakefield Acceleration in a Uniform Plasma. *Physical Review Accelerators and Beams* 26 (11), 111302 (2023).

J. R. Pierce et al. Arbitrarily Structured Laser Pulses. *Physical Review Research* 5 (1), 013085 (2023).

D. Ramsey et al. Exact Solutions for the Electromagnetic Fields of a Flying Focus. *Physical Review A* 107 (1), 013513 (2023).

M. T. Burkey et al. Improved Limit on Tensor Currents in the Weak Interaction from ^8Li β Decay. *Physical Review Letters* 128 (20), 202502 (2022)

F. Li et al. Ultrabright electron bunch injection in a plasma wakefield driven by a superluminal flying focus electron beam. *Physical Review Letters* 128 (17), 174803 (2022).

M. T. Burkey et al. Precision β - ν correlation measurements with the Beta-decay Paul Trap. *Hyperfine Interactions* 240, 36 (2019).

CHAPTER 1

Introduction

This chapter provides an overview of the context and results of this dissertation. This includes the history and status of laser wakefield acceleration (LWFA), the use of spatiotemporally structured laser pulses in LWFA, the particle-in-cell algorithm for simulation of LWFA, and a summary of the results of the dissertation. The chapter is written primarily at a conceptual level. It was written with the goal of being intelligible to a lay audience. Chapter 2 serves a similar purpose from a mathematical standpoint. In addition, this chapter serves as a kind of glue for the dissertation. Chapters 5, 6, and 7, which reproduce material published in peer-reviewed journals and are the key scientific contributions of this dissertation, are largely unmodified from their published forms. Hence, at first glance there is little explicit connection between them. This chapter provides this connection.

1.1 Laser wakefield acceleration: a brief overview

The central focus of this dissertation is laser wakefield acceleration (LWFA). Here a qualitative overview of the process is provided. Figure 1.1 is a visualization of the key aspects of the dynamics in the so-called blowout regime. In the blowout regime, the wake takes the form of a sequence of bubbles trailing the laser. Both the wake and laser move forward (to the right, in this case) at close to the speed of light. Within the bubbles, the Lorentz force is directed radially inward for negatively charged particles. Thus, negative charges copropagating within the first half of each bubble will be continuously focused toward the axis by the transverse component of the force and accelerated by the longitudinal component the force. The particles

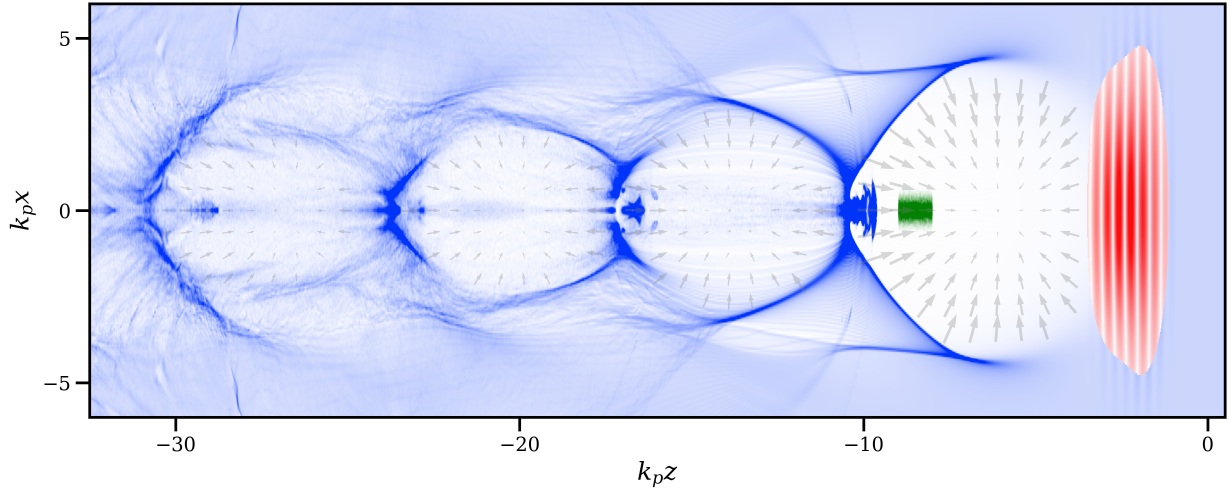


Figure 1.1: Visualization of laser wakefield acceleration in the nonlinear blowout regime. A laser pulse (red) injected into a plasma drives a wake in the electron plasma density (blue). The force field within the plasma wake (light gray arrows) for electrons moving forward near the speed of light simultaneously accelerates and focuses the electrons. Accelerating electrons can be either externally injected into the system (green) or “self-injected” from the back of the plasma wake (dark blue). The wavelength of the wake is on the order of $10c/\omega_p$. This data was generated from a 3D particle-in-cell simulation using the code OSIRIS [2].

that undergo acceleration can be either externally injected as a preexisting particle beam or, in the case of electrons, can be injected from the background plasma using a variety of methods (e.g., [10, 11, 12, 13, 14]). When using a conventional laser pulse as in Figure 1.1, the energy gain of the accelerating particles is limited by the finite energy of the laser, which manifests as diffraction and pump depletion, as well as dephasing of the accelerating particles with respect to the wake. These issues are discussed mathematically in Chapter 2.

LWFA is closely related to plasma wakefield acceleration (PWFA) [15], in which a particle beam drives the wake. LWFA and PWFA are the main examples of what is generally called plasma-based acceleration (PBA). In both LWFA and PWFA, the magnitude of the accelerating force is typically on the order of 10–1000 GeV/m, orders of magnitude higher than

can be sustained in conventional RF accelerators such as synchrotrons and cyclotrons. The maximum accelerating field of RF accelerators is fundamentally limited by the ability of the component structures to withstand ionization and arc discharges. In contrast, plasma-based accelerators can sustain arbitrarily large electric fields because they are already ionized. Because the accelerating gradient in plasma-based accelerators is so much higher than in RF accelerators, PBA offers the opportunity to build more compact and cost-effective particle accelerators for specified particle beam properties. This could lower the size and cost for many applications of particle accelerators, including a next-generation linear collider for high-energy particle physics research, generation of both coherent and incoherent x-ray radiation for high-resolution microscopy, semiconductor ion implantation, and radiotherapy for treatment of various cancers [16]. Because of this high potential impact, plasma-based acceleration has been extensively investigated theoretically, computationally, and experimentally since its discovery nearly a half-century ago.

Although LWFA has much potential, it also has many challenges that must be overcome for its application to be practical in a next-generation linear collider or x-ray free electron laser (XFEL). One is positron acceleration [17]. Positrons can be accelerated easily in the linear regime, but in this case the accelerating gradient is significantly lower. Positrons can also be accelerated in the blowout regime within the on-axis electron density spikes in Fig. 1.1, but the volume in which this is possible is much smaller than that for electrons within the bubbles. Finding ways to accelerate positrons with LWFA at high accelerating gradients and in larger phase volumes is an active area of research [17]. For electron acceleration in the blowout regime, there are also a number of challenges. Certain applications, most notably a next-generation linear collider, require an output beam energy far higher than can be produced by a single LWFA stage using a conventional laser pulse. One way to address this is to accelerate the particle beam through a sequence of independent LWFA stages [18]. Staging is technically very challenging because the diameter of the plasma wake is typically on the length scale of 10–1000 μm , far smaller the length scale of the accelerating region of

conventional RF accelerators. This is exacerbated by jitter and drift in laser systems. Staging is also challenging because the accelerating particle beam must be adiabatically coupled between the vacuum and plasma in every stage through a tailored plasma density profile [19]. Another possibility to overcome single-stage energy limitations is the use of spatiotemporally structured laser pulses to extend the acceleration length of a single LWFA stage. This is a major focus of this dissertation and is described in more detail in Section 1.3.

As mentioned above, LWFA is closely related to PWFA. There are many similarities in the physics underlying the two methods, but also many practical differences. For the sake of limiting the scope of this work, LWFA is the focus. However, many of the same ideas explored in this dissertation can be applied to PWFA. For example, flying focus laser pulses, discussed in Section 1.3, translate directly to flying focus particle beams, as in Ref. [20]. Exploration of the interaction of spatiotemporally structured particle beams with matter is a natural extension of the work in this dissertation, particularly the ideas of ASTRAL pulses (Chapter 5).

1.2 Laser wakefield acceleration: a brief history

In this section, we provide a brief overview of the history of research on laser wakefield acceleration. The purpose is to provide context for the specific research undertaken in this dissertation. A more complete history of the field is provided in Ref. [16]. Table 1.1 shows a brief chronology of research in laser wakefield acceleration including significant developments in theory, simulation, and experiment. The table is not intended to be complete, but to provide sufficient context to understand this dissertation.

Year	Event
1959	Oscar Buneman publishes the first computer plasma simulations [21]
1962	John Dawson publishes electrostatic sheet model for computer simulation of plasma [22]
1971	Demonstration of fully electromagnetic PIC simulations [23]
1979	Tajima and Dawson publish <i>Laser Electron Accelerator</i> , introducing the concept of LWFA and showing the first PIC simulations of LWFA [24]
1993	First experimental demonstration of LWFA using beat wave excitation; 7 MeV was transferred to an electron beam over 1 cm [25]
1997	First PIC code featuring the quasistatic and ponderomotive-guiding-center approximations (WAKE) [26]
2002	Demonstration in simulations of quasi-monoenergetic electron beams via LWFA in the blowout regime [27]
2004	Three articles published in Nature experimentally demonstrate monoenergetic ~ 100 MeV electron beams accelerated using LWFA [28, 29, 30]
2004	First 3D simulations of LWFA on centimeter scales in the nonlinear blowout regime, demonstrating acceleration of self-injected electrons to ~ 1 GeV [31]
2007	Lu et al. publish phenomenological theory of LWFA in the self-guided regime [1]
2007	First experimental demonstration of acceleration of electrons to 1 GeV using LWFA [32]
2008	First GPU-parallelized PIC simulations [33]
2014	First simulations demonstrating positron acceleration in LWFA using a Laguerre-Gaussian transverse laser profile [34]

2014	Experimental demonstration of acceleration of a self-injected monoenergetic 4 GeV electron beam over 10 cm [35]
2018	Experimental demonstration of the chromatic flying focus [3]
2019	Experimental demonstration of acceleration of a self-injected monoenergetic 8 GeV electron beam over 20 cm [36]
2020	Axiparabola-echelon flying focus laser proposed as a path toward single-stage acceleration of electron beams to 1 TeV in ~ 1 m in LWFA [37, 38]
2021	Development of QPAD, a quasi-static PIC code based on azimuthal mode decomposition [39]
2023	Concept of ASTRL pulses introduced [7]*
2025	GPU Acceleration of the Particle-in-Cell Code OSIRIS [9]*
2025	Simulation demonstration of 40 GeV electron acceleration in 30 cm using ASTRL pulses [8]*

Table 1.1: A brief chronology of research in laser wakefield acceleration and plasma simulation showing development of theory, simulation, and experiment. The emphasis is on work related to this dissertation. Years correspond to the date of publication. The final three entries marked with a * correspond to Chapters 5, 6, and 7 in this dissertation.

Research on laser wakefield acceleration began in 1979 with Tajima and Dawson’s publication of *Laser Electron Accelerator* [24]. Since then, experimental, theoretical, and computational capabilities have rapidly developed. At the time of writing, the highest-energy electron beams produced in a single LWFA stage are on the order of ~ 10 GeV in a ~ 20 cm channel (results from 2019) [36]. In addition, electron beams of sufficient quality (charge, energy, energy spread, and angular divergence) to drive an FEL at ultraviolet wavelengths have been generated from an LWFA at a lower beam energies [40, 41, 42, 43]. The experimental progress at higher energies has spurred exploration of new concepts that could be applied

to the development of a LWFA-based linear collider [18]. In 2020, two independent groups proposed using an axiparabola-echelon pair to overcome the dephasing limit of a single LWFA stage [37, 38]. Both works proposed acceleration of electrons to TeV energies in meter-scale plasmas. This development set the conceptual foundation for the physics research conducted by the author which is presented in this dissertation. In particular, investigations of other kinds of laser pulses that could be used as a driver of LWFA led to the author’s research on the concept of Arbitrarily Structured Laser (ASTRL) pulses, presented in Ref. [7]. ASTRL pulses were later applied to LWFA in Ref. [8]. These two publications are reproduced in Chapters 5 and 6.

Reference [16] describes the development of plasma-based acceleration as “a tale of synergy between theory, simulation, and experiment.” Table 1.1 demonstrates this important point. For example, the first experimental demonstration of LWFA 1993 in Ref. [25] was preceded by a decade of theoretical and computational studies. As another example, consider the 2002 publication of Pukhov and Meyer-ter-Vehn [27]. They presented PIC simulation results of ultra-intense and high power lasers (~ 200 TW) propagating through mildly underdense plasmas. It was found that these lasers created wakes in a nonlinear blowout regime (already identified for particle beam drivers [44]), that was subsequently labeled the bubble regime. The wake was sufficiently large that copious numbers of background electrons were trapped and accelerated. These electrons eventually formed a quasi-monoenergetic beam as they rotated in the axial phase space. Two years later in 2004 it was shown in simulations that similar physics could occur for lower laser powers and lower plasma densities [31]. And nearly simultaneously in 2004, three experimental groups independently demonstrated the generation of monoenergetic electron beams with ~ 100 MeV energy using lower intensities and powers (10 TW) than the work of Pukhov and Meyer-ter-Vehn [28, 29, 30]. Each of these experimental teams relied on particle-in-cell (PIC) simulations to interpret their experiments.

The synergy of simulations and experiments in which simulations provide the ability to explore the physics of LWFA before experiments are conducted relates directly to this

dissertation: theoretical and computational work on the use of spatiotemporal structured laser drivers for LWFA is preceding experiments that are now under consideration [37, 38, 45, 8]. In particular, at the time of writing, the National Science Foundation is funding an \$18 million design study on the feasibility of creating the NSF OPAL laser system at the University of Rochester [46]. One of the proposed flagship experiments on NSF OPAL is the acceleration of 100-GeV-class electron beams using LWFA with spatiotemporally structured laser drivers [47]. This potential experimental reality has become apparent through numerical simulations.

Computational modeling of laser wakefield acceleration is possible through several simulation methodologies. The most notable of these is the particle-in-cell (PIC) algorithm. An overview of these methodologies is given in Sec. 1.5. Here a high-level historical overview is provided, following the more detailed account in Ref. [48]. The first plasma simulation codes were written in the late 1950s and early 1960s, notably by Buneman [21] and Dawson [22]. The earliest codes were one-dimensional “sheet models,” which simulate the pairwise Coulomb interaction of all the plasma particles and are based on unique nature of the Coulomb force in 1D. Langdon and Dawson later developed an electromagnetic sheet model in the late 1960s [49]. Sheet models exaggerate the effects of particle collisions compared to those of real plasmas unless the number of simulation particles matches the number of physical particles. This issue and the restriction of sheet models to 1D were addressed through the development of electrostatic PIC codes, including those by Hockney [50], Dawson, Hsi, and Shanny [51], and Birdsall and Fuss [52]. In these methods, particles do not directly interact through Coulomb’s law. Instead, their charges are deposited onto a grid to compute the charge density, from which the electric field is obtained by inverting Gauss’s law. The magnetic field is assumed to not evolve. In the 1970s, fully electromagnetic PIC codes emerged. Early codes were developed by Morse and Nielson [53], Boris [23], Langdon and Lasinski [54], and Buneman, Barnes, Green, and Nielson [55]. Birdsall and Langdon developed early supporting theory [56] and later wrote a standard textbook providing a complete mathematical description of the PIC method and references to the early history [57]. In these codes, both the charge

and current densities of the plasma are computed by depositing them on a grid; the full set of Maxwell's equations is then evolved to capture the full dynamical evolution of the electric and magnetic fields. A half-century later, PIC codes are still the gold standard for simulation of kinetic and nonlinear processes in plasmas. Active research is still being carried out on improvement of the basic components of the algorithm including the field solver [58, 59, 60, 61, 62], particle pusher [63, 64, 65], charge/current deposition, and field interpolation. Furthermore, research continues on developing PIC algorithms that are more computationally efficient, described in Sec. 1.5 and Appendix A.

The PIC algorithm has been essential for modeling LWFA in particular because of its ability to numerically handle the complex evolution of fields and particles. These typically feature crossings of particle trajectories which cannot be accurately modeled using fluid codes. The first simulations of LWFA by Tajima and Dawson were carried out using the PIC algorithm with 5120 electrons for about 1000 timesteps in one spatial dimension (1D) [24]. As computing capabilities increased, both through increase in performance of individual computing hardware as well as through creation of supercomputers, the ability to model plasma physics, including LWFA, dramatically increased. For example, the main simulation from Ref. [1] featured about 500 million particles pushed for 3 million timesteps in 3 spatial dimensions. The developing hardware resources have required development of the software implementing the PIC algorithm. For example, the version of OSIRIS in Ref. [1] was parallelized to run on many CPUs using MPI. The past decades have seen the emergence of GPU-based architectures which have enabled computing power to continue to increase exponentially. However, this increase in computing power has arisen at the expense of software complexity. New data structures and new parallelism strategies have been required for PIC codes to run effectively on GPUs. One of the first efforts appeared in the literature in 2008 (Ref. [33]). Efficient implementation of the PIC algorithm on GPUs has remained an active area of research since then. This is the topic for the final publication of the author appearing in Table 1.1, namely Ref. [9]. This material is presented in Chapter 7.

1.3 Overcoming dephasing in LWFA with spatiotemporally structured laser pulses

The previous section discussed the historical background in which exploration of spatiotemporally structured laser pulses in LWFA emerged. In this section, more detail on these ideas is provided. We begin with a conceptual explanation of the physics of one type of “flying focus” laser pulse, the primary type of spatiotemporally structured laser pulse that has been considered. Figure 1.2 shows a chromatic flying focus. In this specific case, a chirped pulse is focused by a lens with a sufficiently large chromatic aberration. The lens focuses different colors to different points in space. Thus, the varying local color of the laser pulse from the frequency chirp results in different longitudinal slices of the pulse being focused to different longitudinal positions. The resulting pulse has two properties that differ from a conventional pulse. First, the propagation length of the laser pulse is determined by the difference in focal points between the highest and lowest frequencies of the laser pulse. In general, this can be much larger than the length scale for propagation of a single color, which is the Rayleigh range. Second, the intensity peak of the laser pulse can propagate at a controllable velocity which is different than the speed of light in vacuum. This can be understood conceptually by considering a laser pulse that consists of two frequencies, ω_1 and ω_2 , at longitudinal slices ξ_1 and ξ_2 and focused to positions z_1 and z_2 . Because each slice propagates at the speed of light, the intensity peak of the laser will effectively propagate at a velocity $v_f = \Delta z / \Delta t = c[\Delta z / (\Delta z - \Delta \xi)]$. The velocity v_f can be different than c and can be customized through the parameters Δz and $\Delta \xi$. The effect is analogous to the arbitrary propagation velocity that can be achieved by a crowd doing “the wave.” Negative-velocity pulses are even possible given appropriate parameters.

The ability to control the velocity and propagation length of a laser pulse has application to a wide variety of laser-matter interactions, including LWFA. In LWFA, dephasing of the wake with respect to the particle beam occurs with a conventional pulse because the laser

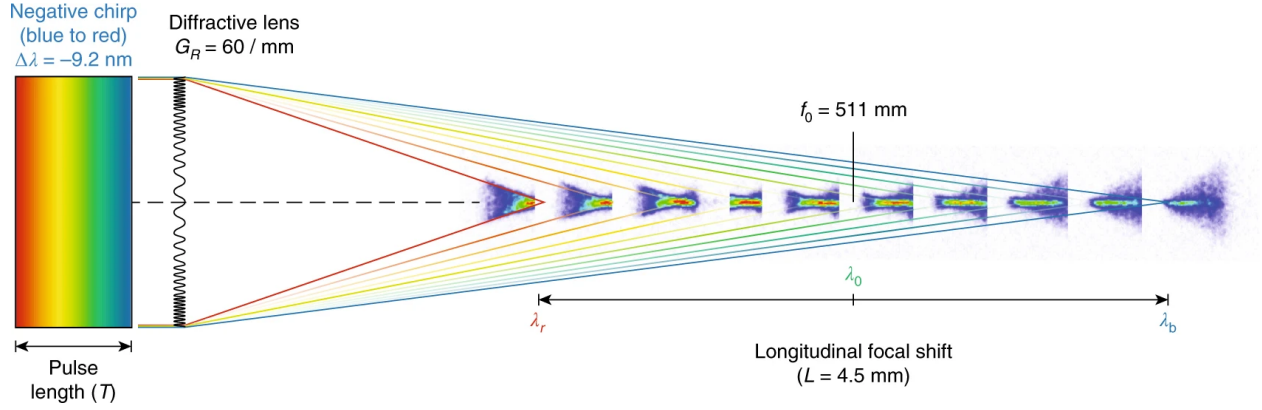


Figure 1.2: A flying focus laser pulse. The laser pulse features two unusual properties: (1) a controllable range of propagation which exceeds the Rayleigh range and (2) controllable envelope velocity which may be different from the speed of light. In this case, the flying focus is created by sending a chirped laser pulse through a lens with a large chromatic aberration. The data, measurements from an experiment of this optical configuration, show a flying focus with a subluminal velocity propagating over about 100 Rayleigh ranges. Figure from Ref. [3]. Used with permission from Springer Nature.

pulse moves at close to the group velocity (less than the speed of light) while the accelerating particles move at nearly the speed of light. This places a limit on the acceleration length. This limit can potentially be overcome by customizing the velocity of the intensity peak of the driver to be exactly the speed of light in plasma. The acceleration length is then determined by the spread in focal points within the pulse and can in principle be arbitrarily large. These are the ideas that were proposed in Refs. [37, 38]. These works proposed using a flying focus formed using an axiparabola-echelon pair to accelerate electrons to TeV energies in a single plasma stage of several meters. This optical configuration and the resulting pulse structure of an axiparabola-echelon pair differ in some ways from those of the chromatic flying focus shown in Figure 1.2, but the properties of an extended focal region and customizable envelope velocity are shared. The chromatic flying focus is shown because the author finds it slightly easier to explain.

1.4 Arbitrarily Structured Laser Pulses

The experimental confirmation that flying focus laser pulses could be generated [3] led to theoretical and computational studies exploring their application in a variety of laser-plasma interactions, including Raman amplification [66], generation of ionization fronts [67], photon acceleration [68], and laser wakefield acceleration [37, 38]. These early simulations studied the fields predicted to be generated by specific optical configurations, namely the flying focus produced by (1) sending a chirped pulse through a chromatic lens [3] and (2) reflecting a pulse off an axiparabola and an appropriately tailored echelon [37]. Later, it was realized that simpler theoretical models could be used to capture the essence of the flying focus, making simulations simpler and calculations more tractable. This approach was used to study the application of flying focus pulses in a variety of physical scenarios including vacuum electron acceleration [69], photon acceleration [70], nonlinear Thomson scattering [71], radiation reaction [72], charged-particle transport [73], and vacuum birefringence [74].

In order to computationally model LWFA with a flying focus pulse, the author, like several others, sought solutions of Maxwell’s equations that captured the essential features of the flying focus while being analytically simpler than the exact fields of the proposed optical configurations. This led to the realization that approximate flying focus solutions could naturally be generated by taking superpositions of conventional laser pulses with different delays and focal points. It became clear that this method of generating solutions generalized the concept of a flying focus: while a flying focus has a dynamic focal point, this method could be used to mathematically build solutions that had other dynamically varying quantities such as the spot size, polarization, and angular momentum. Because of the great flexibility of this method, we named these solutions “Arbitrarily Structured Laser Pulses,” abbreviated to “ASTRL pulses.” The name and acronym are due to John Palastro. The work was published in Ref. [7]. To formulate this concept, it was necessary to develop an understanding of the full spatiotemporal evolution of conventional laser pulses in linear media. This led to the

results presented in Chapter 4.

The utility of the ASTRL framework is multifaceted. First, it provides a convenient way to initialize and launch structured light in simulations (e.g., flying focus pulses into PIC simulations for LWFA). By capturing the essential features of a given type of structured light, it can shed light on fundamental aspects of the laser-matter interaction in question. The ASTRL framework also serves as a unifying mathematical and conceptual language for describing various types of structured light. For example, the ideas of pulses with evolving focal points, polarization, and spot size were already present in the literature, but it was not recognized that there was a natural connection between them. The simple nature of ASTRL pulses also allows rapid conceptualization of new kinds of structured light. For example, with this concept in mind, one can readily conceive of a flying focus pulse with dynamic polarization. Finally, in addition to serving as a theoretical tool (e.g., for broadly understanding the physics of LWFA with a flying focus driver), ASTRL pulses have the potential to become an experimental reality. The simple idea of superposing pulses with different properties naturally motivates an experimental configuration, discussed in Chapter 5.

1.5 Methodology: the particle-in-cell simulation algorithm

In this section, we provide an overview of the particle-in-cell (PIC) algorithm with a focus on its connection to research on LWFA. Since the latter was first conceived and simulated by Tajima and Dawson in 1979 (Ref. [24]), the particle-in-cell (PIC) algorithm has been the workhorse methodology for its simulation. Fully electromagnetic PIC simulations capture the full self-consistent evolution of the laser and the plasma. Fluid descriptions can have difficulty capturing relativistic effects, modeling regions of particle-crossing, and modeling the discontinuities that naturally form in laser-driven plasma wakes. PIC simulations capture all these features and provide insight into the dynamics of LWFA. In simulations, time-resolved diagnostics of the entire system can be output at arbitrary intervals. Such probing of the

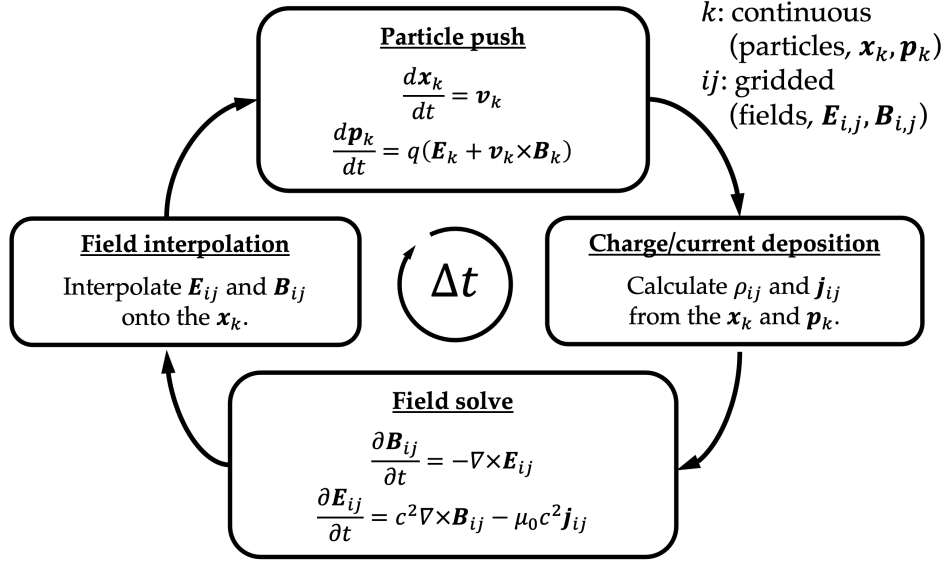


Figure 1.3: Schematic depicting a single time step in the fully electromagnetic particle-in-cell (PIC) algorithm. Used with permission from Ref. [4].

system is impossible in real experiments. Simulations are also a natural platform to test new ideas, such as injection methods or novel laser pulse drivers, to determine whether the ideas are promising before investing resources to test them in an experiment. Examples of the interplay between simulation (especially with PIC), theory, and experiment were discussed in Section 1.2.

Figure 1.3 shows a breakdown of the key steps in the fully electromagnetic PIC algorithm. The plasma is represented as a collection of particles with positions that vary continuously. The electromagnetic fields are discretized on a grid. At a given timestep, the current electromagnetic fields are used to evolve, or “push”, the particle trajectories using $\mathbf{F} = m\mathbf{a}$ and the usual Lorentz force $\mathbf{F} = q[\mathbf{E} + \mathbf{v} \times \mathbf{B}]$. From the new positions and velocities of the particles, the charge density ρ and current density $\mathbf{J}$ can be computed. Using these quantities, the electric and magnetic fields can be advanced further in time. With the new fields, the process can be repeated. This is a very basic overview of the method. Important details and subtleties can be found in the literature (e.g. Ref. [75, 76]). Some are discussed in more

detail in Chapter 2.

Several developments of the fully electromagnetic PIC algorithm and its implementation are important to the work in this dissertation. One is the development of the PIC code OSIRIS [2]. OSIRIS serves an important role in the plasma simulation community as a stable, performant code which is built for extension. The research in this dissertation was possible through developments implemented in OSIRIS by other scholars before the author arrived at UCLA. One example is the quasi-3D simulation algorithm [77, 78]. In this method, particle positions are stored without a grid in 3D, while all grid quantities (including $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{J}$) are stored in 2D grids as components of an azimuthal Fourier expansion. This allows 3D plasma physics to be accurately simulated with far less cost computational than full 3D simulations for scenarios where there is a high degree of azimuthal symmetry. This simulation geometry naturally applies to laser wakefield acceleration. Another example is the development of customized finite-difference time domain (FDTD) field solvers, especially the “dual solver” introduced in Ref. [65]. This field solver corrects for numerical dispersion, or inaccuracy in wave propagation induced by numerics, of laser pulses propagating in plasma. This correction is important for many of the simulations run in this dissertation.

Another important capability is the quasistatic PIC algorithm [26], implemented in codes such as WAKE, QuickPIC, QPAD, and HiPACE++ [26, 79, 80, 39, 81, 82]. In these simulations, time-averaged equations are used to evolve the laser and the plasma. The simulation timestep resolves the timescale for evolution of the laser envelope rather than the laser frequency. This results in a drastic speedup compared to fully electromagnetic PIC. Quasistatic PIC codes can have numerical issues caused by inaccuracy in the rear of highly-nonlinear plasma wakes. The quasistatic PIC code QPAD (QuickPIC with Azimuthal Decomposition) was essential for the work in Chapters 3 and 6, which would not have been possible with fully electromagnetic PIC simulations because of computational constraints. More mathematical details on the quasistatic algorithm, which is also a useful theoretical framework, are provided in Chapter 2.

GPU acceleration of the PIC algorithm is another simulation capability directly related to this dissertation. GPUs offer the potential to provide enormous speedups over CPUs, but the implementation of the PIC algorithm on GPUs presents a number of nontrivial challenges. Research on the PIC algorithm on GPUs and a practical implementation in OSIRIS were a major part of the author’s doctoral research. Results from this effort were published in Ref. [9] and are described in this dissertation in Chapter 7. This development may be helpful for computational modeling of plasmas, including LWFA with spatiotemporally structured laser pulses, in future work.

The final simulation capability we discuss is the so-called “Lorentz-boosted” simulation method for LWFA [83, 84, 85]. In this method, the PIC simulation is run in a Lorentz frame boosted with respect to the lab frame in the laser’s direction of propagation. Under such a transformation, the laser is doppler-redshifted and the plasma is Lorentz-contracted. This combination reduces the scale separation of the system in the boosted frame and ideally results in speedup proportional to γ^2 with respect to the moving-window lab-frame simulation, where γ is the boost Lorentz factor. In practice, there are several challenges to modeling LWFA in a Lorentz-boosted frame. Chief among these is the numerical Cherenkov instability (NCI) [86, 87]. Overcoming the NCI was a major effort by many researchers during the 2010s [88, 89, 90, 91, 92, 93, 94, 95, 59, 96]. In practice, boosted-frame simulations were not used for the core chapters of this dissertation, but were an active focus of the author’s research. Through this work, it was discovered that the dual solver (Ref. [61]) effectively suppresses the NCI. Another aspect of the author’s doctoral research was implementing boosted-frame simulation capabilities in OSIRIS. A summary of these developments and a practical guide to running Lorentz-boosted simulations are provided in Appendix A.

1.6 Dissertation outline

An outline of the remainder of this dissertation is as follows:

- Chapter 2 contains an overview of the physics of LWFA and the mathematics underlying its simulation using the particle-in-cell method.
- Chapter 3 presents a study on the scaling laws for the acceleration length and energy gain for LWFA in the blowout regime in uniform plasma.
- Chapter 4 contains a mathematical formulation of linear propagation of laser pulses in general linear dielectrics, including plasma. This formulation provides a basis for the design of ASTRL pulses.
- Chapter 5 contains the theory and simulations of arbitrarily structured laser pulses in vacuum (Ref. [7]).
- Chapter 6 contains theory and simulations for the application of ASTRL pulses to LWFA (Ref. [8]).
- Chapter 7 is an overview of GPU acceleration of the particle-in-cell code OSIRIS (Ref. [9]).
- Chapter 8 contains a summary of this dissertation and a discussion of directions for future work.
- Appendix A contains a detailed, practical guide for running Lorentz-boosted simulations of LWFA.

CHAPTER 2

Laser Wakefield Acceleration in the Blowout Regime:

Theory and Computational Methods

This chapter presents mathematical theory describing laser wakefield acceleration (LWFA) in the blowout regime and the computational methods used for its simulation. In the context of the dissertation, this chapter serves as a mathematical backbone for Chapter 6, which synthesizes the concepts presented here of the plasma wake structure, laser evolution, and witness beam evolution. It also introduces the key computational methods used throughout the dissertation. More generally, this chapter serves as a summary of the key physics of LWFA in the blowout regime. It is written to be as simple as possible while remaining mathematically rigorous. It includes perspectives of the author which he found useful to understand key concepts and is written to serve as both a primer and a reference.

The blowout regime is also often referred to as the “bubble regime”. Blowout refers to the nearly complete expulsion of electrons from the axis when the laser amplitude is sufficiently large, as in Fig. 1.1. This is the most relevant regime for most applications of LWFA for electron acceleration, including the potential creation of the electron arm of a next-generation linear collider. Although analysis of the wake properties in the linear regime provides insight and intuition, this chapter focuses on the blowout regime because of its attractiveness for electron acceleration. However, general concepts applicable to both the linear and nonlinear regimes are discussed. Understanding the physics of laser evolution in the linear regime is critical for the work in this dissertation. Linear propagation of the laser is considered more fully in Chapter 4.

Most of the chapter consists of distillation of theory emerging from the UCLA Plasma Simulation Group, especially Refs. [97], [98], [1], and [99]. In this sense, the chapter also serves as a kind of review of the thought and theory of the UCLA Plasma Simulation Group. The chapter also includes details on the specific particle-in-cell codes (OSIRIS and QPAD) and simulation features that enabled the research in this dissertation.

A note on variables and notation: the quasistatic theory in this chapter is presented in terms of the variables $\xi = z - ct$ and $s = z$. In much of the literature the choice $\xi = ct - z$ was adopted. This results in signs being flipped for terms in several equations. The author have made this choice because he found it to be simpler to conceptualize. This variable choice is also used throughout the rest of the dissertation. The notation $D_x f(x) = \frac{df}{dx}$ is used for total derivatives. In addition, the indicator function $I_{[a,b]}(x)$, defined as

$$I_{[a,b]}(x) \equiv \begin{cases} 1, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases} \quad (2.1)$$

is frequently used in this chapter in order to streamline the notation. Note that $I_{[a,b]}(x) = \eta(x - a) - \eta(x - b)$ where η is the Heaviside step function (which can itself be expressed as $\eta(x) = I_{[0,\infty)}(x)$).

2.1 Kinetic description of plasma

We begin by considering a fully ionized plasma for which quantum effects are negligible. The most accurate and complete mathematical description of such a system is Newton's law for N point particles with the Lorentz force:

$$\dot{\mathbf{x}}_j = \mathbf{v}_j \quad (2.2)$$

$$\dot{\mathbf{p}}_j = q_j \left[\mathbf{E} + \frac{\mathbf{v}_j}{c} \times \mathbf{B} \right] \quad (2.3)$$

where $\mathbf{p}_j$ and $\mathbf{v}_j$ are related by

$$\mathbf{p}_j = m_j \gamma_j \mathbf{v}_j \quad (2.4)$$

and the Lorentz factor is given by

$$\gamma_j = \left[1 + \left(\frac{\mathbf{p}_j^2}{m^2 c^2} \right) \right]^{1/2}. \quad (2.5)$$

The Lorentz factor can also be expressed as $\gamma_j = [1 - \mathbf{v}_j^2/c^2]^{-1/2}$. CGS units, which are typically used in the literature, are used here. In Eq. 2.3, the electromagnetic fields evolve self-consistently through Maxwell's equations:

$$\nabla \cdot \mathbf{E} = 4\pi\rho \quad (2.6)$$

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B} \quad (2.7)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (2.8)$$

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \partial_t \mathbf{E} \quad (2.9)$$

Here, the charge density ρ and current density $\mathbf{J}$ are determined from the motion of the point particles:

$$\rho(t, \mathbf{x}) = \sum_j q_j \delta(\mathbf{x} - \mathbf{x}_j(t)) \quad (2.10)$$

$$\mathbf{J}(t, \mathbf{x}) = \sum_j q_j \mathbf{v}_j(t) \delta(\mathbf{x} - \mathbf{x}_j(t)) \quad (2.11)$$

Together, Eqs. 2.2–2.11 constitute a complete system of differential equations. The physics of ionization and recombination are neglected, but this description can be modified to include them. Aside from these and quantum effects, this system is the highest-level mathematical description possible for a plasma and includes the effects of collisions (near-neighbor particle interactions). In practice, this system is both analytically and computationally intractable.

In the point-particle description, the state of the plasma can be fully specified by the function

$$F_s(t, \mathbf{x}, \mathbf{p}) \equiv \sum_{j=1}^{N_s} \delta(\mathbf{x} - \mathbf{x}_j(t)) \delta(\mathbf{p} - \mathbf{p}_j(t)) \quad (2.12)$$

where s indexes each species of identical particles (e.g., electrons) with charge q_s and mass m_s . In terms of F_s , Eqs. 2.2 and 2.3 can be recast as an equation known as the Klimontovich

equation [100]:

$$\partial_t F_s + \mathbf{v} \cdot \nabla_{\mathbf{x}} F_s + \frac{q_s}{m_s} \left[\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right] \cdot \nabla_{\mathbf{p}} F_s = 0 \quad (2.13)$$

This equation, containing the same information as Eqs. 2.2 and 2.3, is thus also analytically and computationally intractable.

On the other hand, if collective effects dominate the plasma such that nearest-neighbor particle collisions are negligible, then one can derive from Eq. 2.13 the Vlasov equation as an approximation:

$$\partial_t f_s + \mathbf{v} \cdot \nabla_{\mathbf{x}} f_s + \frac{q_s}{m_s} \left[\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right] \cdot \nabla_{\mathbf{p}} f_s = 0 \quad (2.14)$$

This equation has the exact same form as the Klimontovich equation, but is repeated to stress the difference between F_s and f_s . The latter, referred to as the particle distribution function, can be arbitrary and continuous rather than a sum of delta functions as in Eq. 2.12. It represents the probability density function for the distribution of particles in the $(\mathbf{x}, \mathbf{p})$ phase space. For example, $f(0, \mathbf{x}, \mathbf{p}) = \exp(-\mathbf{x}^2/\sigma_x^2) \exp(-\mathbf{p}^2/\sigma_p^2)$ is a valid initial distribution function, though not an equilibrium. The Vlasov equation couples to Maxwell's equations through the continuum version of Eqs. 2.10 and 2.11:

$$\rho(t, \mathbf{x}) = \sum_s q_s \int f_s(t, \mathbf{x}, \mathbf{p}) d\mathbf{p} \quad (2.15)$$

$$\mathbf{J}(t, \mathbf{x}) = \sum_s q_s \int \mathbf{v} f_s(t, \mathbf{x}, \mathbf{p}) d\mathbf{p} \quad (2.16)$$

The Vlasov-Maxwell system, constituted by Eqs. 2.6–2.9 and 2.14–2.16, is a complete description of a classical collisionless plasma. Unlike the point-particle description of plasma, the Vlasov-Maxwell system is analytically tractable in certain situations (e.g., linear Landau damping). For most plasmas of interest, however, it is also analytically intractable. But for many problems, its solution can be approximated through the particle-in-cell method, which simulates the particle-based description for finite-size particles.

Mathematically, the PIC algorithm begins by writing Newton's law for finite-size particles:

$$\dot{\mathbf{p}}_j = q_j \int \left[\mathbf{E}(t, \mathbf{x}') + \frac{\mathbf{v}_j}{c} \times \mathbf{B}(t, \mathbf{x}') \right] S(\mathbf{x}_j - \mathbf{x}') d\mathbf{x}' \quad (2.17)$$

Here, $S(\mathbf{x})$ is referred to as the particle shape function, assumed for simplicity to be the same for all particles. The plasma source terms for finite-size particles can be written as

$$\rho(t, \mathbf{x}) = \sum_j q_j S(\mathbf{x} - \mathbf{x}_j(t)) \quad (2.18)$$

$$\mathbf{J}(t, \mathbf{x}) = \sum_j q_j \mathbf{v}_j(t) S(\mathbf{x} - \mathbf{x}_j(t)) \quad (2.19)$$

Equations 2.17–2.19, together with Maxwell’s equations (Eqs. 2.6–2.9) and Eq. 2.2, constitute a complete system. In the PIC algorithm, these are simulated by discretizing the grid quantities (either in real or Fourier space) and discretizing the evolution of both fields and particles in time.

It is worth noting that the effects of collisions are still included in this description. Macroparticle-macroparticle interactions occur on length scales on the order of or larger than that of the particle shape function. For example, Refs. [101, 102, 103] use the PIC algorithm to self-consistently simulate particle collisions for finite-size particles. The effects become more pronounced when the inter-macroparticle spacing is resolved by the grid (i.e., less than one macroparticle per cell on average). It is also possible to approximate the effect of collisions without resolving the inter-macroparticle spacing by using Monte-Carlo algorithms, as in Ref. [104]. Effects of field ionization can also be approximated, as in Ref. [105].

The analogue of Eq. 2.12 can be written for finite-size particles:

$$F_s(t, \mathbf{x}, \mathbf{p}) = \sum_j S(\mathbf{x} - \mathbf{x}_j(t)) \delta(\mathbf{p} - \mathbf{p}_j(t)) \quad (2.20)$$

However, unlike for point particles, the Klimontovich equation does not hold without a nonzero right-hand-side for finite-size particles, i.e.,

$$\partial_t F_s + \mathbf{v} \cdot \nabla_{\mathbf{x}} F_s + \frac{q_s}{m_s} \left[\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right] \cdot \nabla_{\mathbf{p}} F_s \neq 0 \quad (2.21)$$

with F_s given by Eq. 2.20. Note that PIC is neither a discretization of the Vlasov equation nor a solution of the “Klimontovich equation for finite-size particles”. These ideas are discussed in Ref. [106].

2.2 Kinetic description of LWFA

The Vlasov-Maxwell system is appropriate for describing LWFA as collisions are negligible in this process. What is the simplest initial condition for the Vlasov-Maxwell system that gives rise to LWFA in the blowout regime? We consider an initially cold electron-ion plasma and a conventional laser pulse in focus at the plasma entrance. We assume that the ions are immobile and consider only the evolution of the electrons. We will drop subscripts for f , m , and q , which subsequently refer to electrons. The following initial condition captures these assumptions and is the simplest mathematical initialization that gives rise to LWFA in the Vlasov-Maxwell system:

$$f(0, \mathbf{x}, \mathbf{p}) = n_0 H(z) \delta(\mathbf{p}) \quad (2.22)$$

$$\mathbf{E}_\perp(0, \mathbf{x}) = g(z) e^{-\mathbf{x}_\perp^2/w_0^2} e^{ik_0 z} \hat{\mathbf{e}} \quad (2.23)$$

where n_0 is the electron number density, H is the Heaviside step function, $g(z)$ is the longitudinal profile of the laser envelope chosen so that it vanishes in the plasma region $z \geq 0$, w_0 is the initial spot size, k_0 is the laser wavenumber, and $\hat{\mathbf{e}} \cdot \hat{\mathbf{z}} = 0$. $\mathbf{B}$ and E_z are also nonzero initially and are chosen to satisfy Maxwell's equations given the choice of $\mathbf{E}_\perp$. The parameters must be chosen so that $\omega_0 > \omega_p$, where $\omega_0 \equiv ck_0$ is the laser frequency and $\omega_p = \sqrt{4\pi n_0 e^2/m}$ is the plasma frequency. This is required so that the laser does not reflect at the vacuum-plasma interface. The interaction of the laser with the plasma must be sufficiently strong to drive out the electrons. In practice, this requires $a_0 \gtrsim 2$ and $k_p w_0 \gtrsim 2$, where $a_0 \equiv eE_0/mc\omega_0$ is the initial peak normalized vector potential and $E_0 \equiv \max_{\mathbf{x}} |\mathbf{E}_\perp(0, \mathbf{x})|$ is the initial peak laser amplitude. An external electron beam can be added to f behind the laser, but this is not required for excitation of the wake. In addition, Eqs. 2.22 and 2.23 can be modified to include various effects such as nonzero temperature, spatial density gradients, more complex plasma distribution functions, and more complex laser profiles. For example, the work in Chapter 6 modifies Eq. 2.23 to use a flying focus laser profile for the transverse electric field, the functional form of which is derived in Chapter 5.

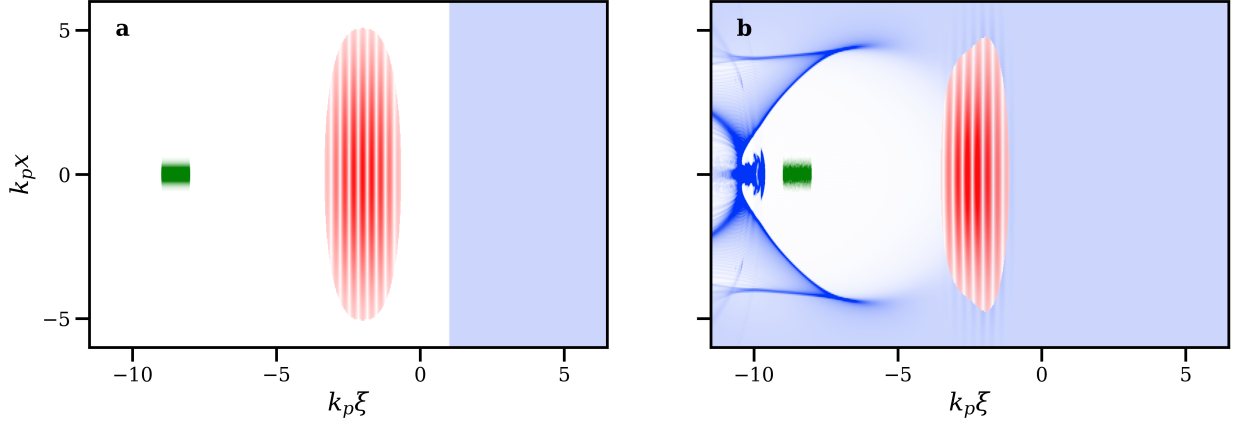


Figure 2.1: 3D PIC simulation of LWFA initialized according to Eqs. 2.22 and 2.23. The laser electric field is shown in red, an external electron beam in green, and the number density of the plasma electrons in blue. (a) The initial condition at $\omega_p t = 0$. (b) The state of the system at $\omega_p t = 40$. In this case, $k_0/k_p = 10$, $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 2$. The piecewise-polynomial profile used in Ref. [1] is used for the $g(z)$ function in Eq. 2.23. The parameters for this system are the same as for the case shown in Fig. 1.1. The simulation box moves forward at the speed of light.

The Vlasov-Maxwell system does not admit explicit solutions for the initial conditions in Eqs. 2.22 and 2.23 in the blowout regime. However, its solution can be approximated numerically. The particle-in-cell (PIC) algorithm, discussed in more detail in Sec. 2.10, is one of the primary methods for approximating solutions of this system. In this method, a large number of finite-size particles move in response to the self-consistent electromagnetic fields generated through their motion. Figure 2.1(a–b) shows a 3D PIC simulation of LWFA for the same parameters as in Fig. 1.1 run using OSIRIS. The initial condition from Eqs. 2.22 and 2.23 is shown in Fig. 2.1(a). The laser electric field is shown in red, an external electron beam in green, and the number density of the plasma electrons in blue. Figure 2.1(b) shows the state of the system at $\omega_p t = 40$ in a simulation domain moving forward at the speed of light. The laser and electron beam both propagate forward at close to the speed of light, so their locations in the box are approximately the same. The excess plasma upstream of the laser is

included for illustration; only a narrow slab in front of the laser is required computationally.

Although the exact results shown in Figure 2.1(b) cannot be derived by hand, the figure clearly shows some universal characteristics of the plasma response: the laser pushes plasma electrons forward and outward, resulting in the formation of a narrow sheath (dark blue). The expulsion of electrons leads to an ion cavity (white). The electrons in the sheath are eventually pulled back to the axis by the ions, resulting in a bubble structure. A number of basic questions can naturally be asked: why is the shape of the bubble roughly spherical? Can the trajectories of the electrons along the sheath be described? What are the electromagnetic fields inside the bubble? What is the length scale over which the laser drives this process? How do the particle beams feed back on the wake? How do particle beams evolve? How should the properties of an externally injected electron bunch be chosen? The remainder of this chapter provides an overview of theoretical results that address these questions.

2.3 The ponderomotive guiding center approximation

The kinetic description of LWFA can be reduced to a more analytically tractable form through the ponderomotive guiding center (PGC) approximation (coined in Ref. [107]) and the quasistatic approximation. In addition to providing the ability to use the PGC approximation in full PIC simulations and enabling quasistatic PIC simulations, discussed in Sec. 2.10, the equations derived from this approach provide a natural starting point for phenomenological description of the physics of LWFA in the blowout regime. This theoretical framework was introduced in Ref. [26], building on previous published work. The main results of Ref. [26] are summarized below. The goal of this section is to explain the key ideas, without necessarily deriving all of them, as clearly as possible. Reference [26] can be studied for a rigorous derivation of the results. In this presentation, the PGC and quasistatic approximations are separated into different sections so that one can clearly see the results of each. To compartmentalize, the particle evolution, laser evolution, and wakefield evolution

are further separated within each approximation.

The PGC approximation begins by decomposing all relevant quantities into rapidly- and slowly-oscillating terms. These terms respectively vary in time on the frequency scales of the laser frequency ω_0 and the plasma frequency ω_p . This decomposition is applied to the potentials $\phi = \tilde{\phi} + \bar{\phi}$ and $\mathbf{A} = \tilde{\mathbf{A}} + \bar{\mathbf{A}}$, the fields $\mathbf{E} = \tilde{\mathbf{E}} + \bar{\mathbf{E}}$ and $\mathbf{B} = \tilde{\mathbf{B}} + \bar{\mathbf{B}}$, the plasma particle trajectories $\mathbf{x}_j = \tilde{\mathbf{x}}_j + \bar{\mathbf{x}}_j$ and momenta $\mathbf{p}_j = \tilde{\mathbf{p}}_j + \bar{\mathbf{p}}_j$, and the plasma source terms $\rho = \tilde{\rho} + \bar{\rho}$ and $\mathbf{J} = \tilde{\mathbf{J}} + \bar{\mathbf{J}}$. For each quantity, a tilde denotes rapid variation and a bar denotes slow variation (the same convention as in Ref. [26]). In addition, the wake potential is defined as

$$\psi \equiv \phi - A_z \quad (2.24)$$

and is also decomposed as $\psi = \tilde{\psi} + \bar{\psi}$. The variable ψ is useful because the gauge-invariant combinations $\phi - A_z$ and $\bar{\phi} - \bar{A}_z$ frequently occur in equations descending from this line of analysis. Finally, the rapidly-varying part of the transverse vector potential (associated with the laser) is expressed as an envelope $\hat{A}$ modulated by a carrier wave:

$$\tilde{\mathbf{A}}_{\perp}(t, z, \mathbf{x}_{\perp}) = \frac{1}{2}\hat{A}(t, z, \mathbf{x}_{\perp})e^{ik_0(z-ct)}\hat{\mathbf{e}} + \text{c.c.} \quad (2.25)$$

The factor of $\frac{1}{2}$ in Eq. 2.25 is not included in this definition in the original work Ref. [26], which results in a few factors being different in the resulting equations. Reference [81] includes this factor of $\frac{1}{2}$. Equation 2.25 by itself makes no assumption that the laser evolution is linear and hence contains no reference to phase or group velocity. When the plasma response is linear, this physics is encoded in the phase and amplitude of $\hat{A}$. Similarly, any shift in the wavenumber of the laser pulse is encoded in the phase of $\hat{A}$.

Using all these definitions and assumptions, the PGC approximation is applied by starting with the point-particle description of plasma and expanding the equations in powers of ω_p/ω_0 . This leads to differential equations for the evolution of the slowly-varying particle momenta $\bar{\mathbf{p}}$ and the laser envelope $\hat{A}$.

2.3.1 Particle evolution

We begin by considering evolution of the particles. For ease of notation, the results are derived for a single species of plasma particles of charge q and mass m . The results can easily be extended to multiple species by generalizing $q \rightarrow q_j$ and $m \rightarrow m_j$. Starting with Eqs. 2.2 and 2.3, it was shown in Ref. [26] that the slowly-varying part of each particle's position (the “guiding center”) evolves as

$$\dot{\mathbf{x}}_j = \bar{\mathbf{v}}_j \quad (2.26)$$

and the equation of motion for the slowly-varying part of the particle momentum is

$$\dot{\mathbf{p}}_j = q \left[\bar{\mathbf{E}} + \frac{\bar{\mathbf{v}}_j}{c} \times \bar{\mathbf{B}} \right] - \frac{1}{4} \frac{q^2}{\bar{\gamma}_j m c^2} \nabla |\hat{A}|^2 \quad (2.27)$$

where $\bar{\mathbf{p}}_j$ and $\bar{\mathbf{v}}_j$ are related by

$$\bar{\mathbf{p}}_j = m \bar{\gamma}_j \bar{\mathbf{v}}_j \quad (2.28)$$

and the time-averaged Lorentz factor is given by

$$\bar{\gamma}_j = \left[1 + \frac{1}{m^2 c^2} \left(\bar{\mathbf{p}}_j^2 + \frac{1}{2} \left| \frac{q}{c} \hat{A} \right|^2 \right) \right]^{1/2}. \quad (2.29)$$

Although Eqs. 2.26–2.29 appear simple and intuitive, their derivation is somewhat complicated. The full details are given in Ref. [26].

Equation. 2.27 shows that the evolution of the slowly-varying part of the particle momenta has the form of the Lorentz force equation with the full fields $\mathbf{E}$ and $\mathbf{B}$ replaced by the slowly-varying fields $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$ and the presence of an additional term, referred to as the ponderomotive force:

$$\mathbf{F}_p \equiv -\frac{1}{4} \frac{q^2}{\bar{\gamma}_j m c^2} \nabla |\hat{A}|^2 \quad (2.30)$$

The ponderomotive force causes particles to move away from regions of high laser intensity, independently of the sign of the charge q . The bubble structure of the plasma wakefield can be qualitatively understood in terms of the ponderomotive force: electrons are pushed forward and outward by the laser. For a sufficiently intense laser pulse, electrons are nearly completely

expelled in the region of the laser. For parameters of interest, the movement of ions is negligible. The result is an evacuated region behind the laser populated by approximately stationary ions. Electrons are then attracted back to the axis by these ions, resulting in the bubble closing.

Equation. 2.29 comes from the fact that the rapidly-varying transverse momentum can be approximated in terms of the rapidly-varying vector potential from the conservation of canonical momentum:

$$\tilde{\mathbf{p}}_{\perp} \approx -\frac{q}{c}\tilde{\mathbf{A}}_{\perp}, \quad (2.31)$$

where the vector potential and particle momentum are both taken to be zero upstream of the laser. This relation assumes that the length scale of the transverse motion of the particle is much smaller than the length scale of the transverse variation of $\tilde{\mathbf{A}}_{\perp}$. This approximation provides a constitutive relation which will be used to write a closed equation for the evolution of the laser envelope in the next subsection.

2.3.2 Laser evolution

Next we consider the evolution of the laser envelope within the PGC approximation. The wave equations for the full potentials in the Lorenz gauge are

$$\nabla^2 \phi - \frac{1}{c^2} \partial_t^2 \phi = -4\pi \rho \quad (2.32)$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \partial_t^2 \mathbf{A} = -\frac{4\pi}{c} \mathbf{J} \quad (2.33)$$

along with the gauge condition

$$\nabla \cdot \mathbf{A} + \frac{1}{c} \partial_t \phi = 0. \quad (2.34)$$

Taking the rapidly-varying part of Eq. 2.33, we obtain

$$\nabla^2 \tilde{\mathbf{A}}_{\perp} - \frac{1}{c^2} \partial_t^2 \tilde{\mathbf{A}}_{\perp} = -\frac{4\pi}{c} \tilde{\mathbf{J}}_{\perp} \quad (2.35)$$

For finite-size particles, the rapidly-varying part of the current density may be computed from Eq. 2.19 by approximating $\tilde{\mathbf{v}}_j \approx \tilde{\mathbf{p}}_j/m\bar{\gamma}_j$, using Eq. 2.31 to obtain $\tilde{\mathbf{v}}_{\perp j} \approx -q\tilde{\mathbf{A}}_{\perp}/mc\bar{\gamma}_j$,

and approximating $\mathbf{x}_j \approx \bar{\mathbf{x}}_j$. This yields

$$\tilde{\mathbf{J}}_{\perp} \approx \chi \tilde{\mathbf{A}}_{\perp} \quad (2.36)$$

where the plasma susceptibility χ is defined as

$$\chi(t, \mathbf{x}) \equiv - \sum_j \frac{q^2}{mc} \frac{S(\mathbf{x} - \bar{\mathbf{x}}_j(t))}{\bar{\gamma}_j(t)} \quad (2.37)$$

Thus, Eq. 2.35 becomes

$$\nabla^2 \tilde{\mathbf{A}}_{\perp} - \frac{1}{c^2} \partial_t^2 \tilde{\mathbf{A}}_{\perp} = -\frac{4\pi}{c} \chi \tilde{\mathbf{A}}_{\perp} \quad (2.38)$$

We now transform into (s, ξ) coordinates, applying the Galilean transformation

$$s = z \quad (2.39)$$

$$\xi = z - ct \quad (2.40)$$

Note that this is purely a mathematical transformation, not a Lorentz transformation.

Using the multivariable chain rule, the derivatives transform as $\partial_z = \partial_z(s)\partial_s + \partial_z(\xi)\partial_{\xi}$ and $\partial_t = \partial_t(s)\partial_s + \partial_t(\xi)\partial_{\xi}$, which may be computed as

$$\partial_z = \partial_s + \partial_{\xi} \quad (2.41)$$

$$\partial_t = -c\partial_{\xi} \quad (2.42)$$

Applying this transformation to Eq. 2.38 and substituting in the definition of the laser envelope $\hat{A}$ in Eq. 2.25, we obtain

$$[\nabla_{\perp}^2 + \partial_s^2 + 2ik_0\partial_s + 2\partial_s\partial_{\xi}] \hat{A} = -\frac{4\pi}{c} \chi \hat{A}. \quad (2.43)$$

Although the information in Eq. 2.43 is the same as in Eq. 2.38, Eq. 2.43 is more appropriate computationally because it involves only slowly-varying quantities. The fact that χ varies slowly is key. The left-hand side of Eq. 2.43 would be the same if the full wave equation Eq. 2.33 were transformed using the Galilean transformation and the definition of $\hat{A}$ with no assumptions about the particle motion. However, this would provide no computational

advantage without a way to express $\mathbf{J}$ in terms of slowly-varying quantities as is done here (i.e., $\hat{A}$ and $\bar{\mathbf{x}}$).

In Eq. 2.43, the same symbol $\hat{A}$ is used to represent the vector potential envelope before and after the transformation. In order to compute the resulting differential equation in the transformed variables, the author finds it easier to temporarily notate the transformed vector potential envelope with a new symbol $\hat{A}_{\xi s}$ which satisfies $\hat{A}(t(\xi, s), z(\xi, s), \mathbf{x}_\perp) = \hat{A}_{\xi s}(s, \xi, \mathbf{x}_\perp)$. After the transformation, the symbol $\hat{A}_{\xi s}$ can be relabeled as $\hat{A}$ in order to reduce the number of symbols. It is important to point out that $\hat{A}(s, \xi, \mathbf{x}_\perp)$ and $\hat{A}(t, z, \mathbf{x}_\perp)$ implicitly have different functional dependencies on their respective independent variables. The meaning is usually clear from the context.

2.3.3 Wakefield evolution

To complete the PGC description, we derive equations for the slowly-varying potentials $\bar{\phi}$ and $\bar{\mathbf{A}}$ and slowly-varying fields (also known as the “wakefields”) $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$. Taking the slowly-varying part of the wave equations for the potentials (Eq. 2.32 and Eq. 2.33), the slowly-varying potentials satisfy

$$\nabla^2 \bar{\phi} - \frac{1}{c^2} \partial_t^2 \bar{\phi} = -4\pi \bar{\rho} \quad (2.44)$$

$$\nabla^2 \bar{\mathbf{A}} - \frac{1}{c^2} \partial_t^2 \bar{\mathbf{A}} = -\frac{4\pi}{c} \bar{\mathbf{J}} \quad (2.45)$$

because all the terms in the wave equations for the potentials are linear. For finite-size particles, the slowly-varying plasma source terms $\bar{\rho}$ and $\bar{\mathbf{J}}$ can be expressed to lowest order as

$$\bar{\rho}(t, \mathbf{x}) = \sum_j q S(\mathbf{x} - \bar{\mathbf{x}}_j(t)) \quad (2.46)$$

$$\bar{\mathbf{J}}(t, \mathbf{x}) = \sum_j q \bar{\mathbf{v}}_j(t) S(\mathbf{x} - \bar{\mathbf{x}}_j(t)) \quad (2.47)$$

So the slowly-varying potentials obey wave equations of the same form as those of the full potentials, except now the source terms come from depositing the macroparticles at their

slowly-varying positions and with their slowly-varying velocities. Similarly, the slowly-varying fields have the same form as Maxwell's equations (Eq. 2.6–2.9) with the slowly-varying source terms given above.

2.3.4 Summary

In summary, Eq. 2.27 governs the evolution of the slowly-varying part of the particle trajectories. The force in this equation includes the Lorentz force from the slowly-varying fields $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$ and the ponderomotive force $\bar{\mathbf{F}}_p$, the latter of which depends on the laser envelope $\hat{A}$. The laser envelope evolves according to Eq. 2.43. The slowly-varying fields are derived from the slowly-varying potentials $\bar{\phi}$ and $\bar{\mathbf{A}}$. Both evolve according to wave equations that are of the same form as the wave equations for the full potentials, but with the slowly-varying plasma source terms. The slowly-varying plasma source terms for the potential wave equations and the laser envelope equation can be computed in terms of the slowly-varying particle positions and momenta. So far, the only approximation applied is $k_p/k_0 \ll 1$. The equations derived so far are sufficient to implement the PGC approximation into the standard PIC algorithm, as in Ref. [107]. This is discussed in Sec. 2.10.

2.4 The quasistatic approximation

The equations for the slowly-varying fields are further simplified through the quasistatic approximation. The physical picture for this approximation is that the laser profile remains approximately constant during the time it takes to pass by a plasma particle. The result of this picture is that the laser envelope, electromagnetic fields of the wake, and slowly-varying plasma source terms (1) copropagate at close to the speed of light and (2) evolve over length scales much larger than the length of the wake. These assumptions typically hold in LWFA and PWFA. The approximation is handled mathematically by changing variables to

$(\xi, s) = (z - ct, z)$ and applying the approximation

$$\partial_s \ll \partial_\xi \quad (2.48)$$

for the laser envelope, the slowly-varying fields, and the slowly-varying plasma source terms. Equation 2.48 is the quasistatic approximation. Here, s represents the distance the laser has propagated into the plasma and ξ represents the relative position within a box moving forward at the speed of light. For each of these quantities, the approximation is that the length scale for evolution in s is much larger than the variation in ξ .

The consequences of this approximation are worked out in this section. As before, the approximation is applied separately to the plasma particles, laser envelope, and wakefields.

2.4.1 Particle evolution

The equations for the evolution of $\bar{\mathbf{p}}_j$ and $\bar{\mathbf{x}}_j$ (Eqs. 2.26 and 2.27) are often expressed in terms of the ξ coordinate of the particle, $\bar{\xi}_j(t) = \bar{z}_j(t) - ct$. This enables the particle to be numerically advanced along constant intervals of $\Delta\xi$. This is natural in quasistatic PIC codes because it keeps the macroparticle $\bar{\xi}$ coordinates aligned with the ξ values on which the grid quantities are defined.

As in the discussion on the notation for the vector potential envelope discussed in Sec. 2.3, the author finds it easier to temporarily notate the momentum in the transformed coordinates with a new symbol $\bar{\mathbf{p}}_\xi$ which satisfies $\bar{\mathbf{p}}(t) = \bar{\mathbf{p}}_\xi(\bar{\xi}(t))$. (The j subscript is temporarily omitted.) The time derivative in Eq. 2.27 then becomes $\dot{\bar{\mathbf{p}}}(t) = D_t \bar{\mathbf{p}}_\xi(\bar{\xi}(t))$, where D_t denotes a total derivative with respect to t . Applying the chain rule, one obtains $\dot{\bar{\mathbf{p}}} = (\bar{v}_z - c) D_{\bar{\xi}} \bar{\mathbf{p}}_\xi$. One can then relabel $\bar{\mathbf{p}}_\xi$ as $\bar{\mathbf{p}}$, understanding that $\bar{\mathbf{p}}(t)$ and $\bar{\mathbf{p}}(\bar{\xi})$ have different functional dependencies on their respective independent variables.

Transforming Eq. 2.27 into the (ξ, s) variables, one obtains

$$(\bar{v}_{zj} - c) D_{\bar{\xi}_j} \bar{\mathbf{p}}_j = q \left[\bar{\mathbf{E}} + \frac{\bar{\mathbf{v}}_j}{c} \times \bar{\mathbf{B}} \right] - \frac{1}{4} \frac{q^2}{mc^2} \frac{\nabla |\hat{A}|^2}{\bar{\gamma}_j} \quad (2.49)$$

where $D_{\bar{\xi}_j}$ is the total derivative with respect to $\bar{\xi}_j$. Similarly, in the same notation, Eq. 2.26 becomes

$$(\bar{v}_{zj} - c)D_{\bar{\xi}_j}\bar{\mathbf{x}}_j = \bar{\mathbf{v}}_j \quad (2.50)$$

Equations 2.49 and 2.50 are natural to use in the context of theory and simulations where the fields are expressed in ξ and $\mathbf{x}_\perp$ at a given s . This is the case in quasistatic PIC simulations and in the theory for the structure of the wake in Sec. 2.6.

Under the quasistatic approximation, it can also be shown that particles satisfying the condition $1 - v_z/c \gg \omega_p/\omega_0$ and starting at rest upstream of the laser possess the following “constant of motion” [26]:

$$\gamma mc - p_z = mc - \frac{q}{c}\bar{\psi} \quad (2.51)$$

where $\bar{\psi}$ is evaluated along the particle trajectory. (The j subscripts are dropped in Eq. 2.51 and the rest of this section for ease of notation.) In Eq. 2.51, both γ and p_z contain both slowly- and rapidly-varying components, but their difference is a slowly-varying quantity. Explicitly, for the sake of clarity:

$$\bar{\gamma}mc - \bar{p}_z = mc - \frac{q}{c}\bar{\psi} \quad (2.52)$$

and

$$\tilde{\gamma}mc - \tilde{p}_z = 0 \quad (2.53)$$

Equation 2.52 implies that $\bar{p}_z$ and $\bar{\gamma}$ can both be expressed in terms of $\bar{\mathbf{p}}_\perp$ and $\bar{\psi}$ by combining Eq. 2.52 with Eq. 2.29:

$$\bar{p}_z = \frac{m^2c^2 + \bar{\mathbf{p}}_\perp^2 + \frac{1}{2}\left|\frac{q}{c}\hat{A}\right|^2 - (mc - \frac{q}{c}\bar{\psi})^2}{2(mc - \frac{q}{c}\bar{\psi})} \quad (2.54)$$

$$\bar{\gamma} = \frac{m^2c^2 + \bar{\mathbf{p}}_\perp^2 + \frac{1}{2}\left|\frac{q}{c}\hat{A}\right|^2 + (mc - \frac{q}{c}\bar{\psi})^2}{2(mc - \frac{q}{c}\bar{\psi})} \quad (2.55)$$

Computationally, these equations are significant because they imply that only $\mathbf{p}_\perp$ needs to be advanced using the equation of motion; p_z does not need to be advanced because it

can be obtained from Eq. 2.54. Dividing Eq. 2.52 by Eq. 2.55, we also obtain the following relationship:

$$1 - \frac{\bar{v}_z}{c} = \frac{\frac{2}{mc}(mc - \frac{q}{c}\bar{\psi})^2}{m^2c^2 + \bar{\mathbf{p}}_\perp^2 + \frac{1}{2}\left|\frac{q}{c}\hat{A}\right|^2 + (mc - \frac{q}{c}\bar{\psi})^2} \quad (2.56)$$

These equations will be helpful in Sec. 2.6.

2.4.2 Laser evolution

Almost all of the work is done in the PGC section for the laser. The $\partial_\xi \partial_s \hat{A}$ in Eq. 2.43 term cannot be dropped if changes in the local laser wavenumber and recession of the pulse due to dispersion are physically relevant [108]. Both of these effects must be captured to model LWFA in the blowout regime. The $\partial_s^2 \hat{A}$ term, however, can typically be dropped.

For completeness, the relationship between the laser vector potential envelope $\hat{A}$ and electric field envelope $\hat{E}$ under the quasistatic approximation is worked out here. We have $\tilde{\mathbf{E}}_\perp = -\partial_t \tilde{\mathbf{A}}_\perp \approx \partial_\xi \tilde{\mathbf{A}}_\perp$. Using $\tilde{\mathbf{A}}_\perp = \hat{A} e^{ik_0 \xi} \hat{\mathbf{e}}$, we obtain $\tilde{\mathbf{E}}_\perp \approx \left[ik_0 \hat{A} + \partial_\xi \hat{A} \right] e^{ik_0 \xi} \hat{\mathbf{e}}$. Hence we obtain

$$\hat{E} \approx ik_0 \hat{A} + \partial_\xi \hat{A}. \quad (2.57)$$

Although the quasistatic equations are more naturally formulated in terms of $\hat{A}$ than $\hat{E}$, the latter can be helpful to look at for comparison with PIC simulations which simulate the full laser electric field or for studying specific phenomena (e.g., laser guiding in Chapter 3).

2.4.3 Wakefield evolution

The most notable simplification under the quasistatic approximation occurs for the evolution of the wakefields. Under the quasistatic approximation, $\partial_z \approx \partial_\xi$. Thus, the wave equations for the slowly-varying potentials in the Lorenz gauge (Eqs. 2.58 and 2.59) become

$$\nabla_\perp^2 \bar{\phi} = -4\pi \bar{\rho} \quad (2.58)$$

$$\nabla_\perp^2 \bar{\mathbf{A}} = -\frac{4\pi}{c} \bar{\mathbf{J}} \quad (2.59)$$

while the gauge condition (Eq. 2.34) becomes

$$\partial_\xi \bar{\psi} + \nabla_\perp \cdot \bar{\mathbf{A}}_\perp = 0 \quad (2.60)$$

Thus, the 3+1D Maxwell's equations approximately reduce to static (hence “quasistatic”) 2D Poisson-like equations. (They are not purely Poisson equations because of the ξ dependence of the source terms, but they are 2D Poisson equations for a fixed ξ .) This greatly simplifies theoretical calculations and numerical simulation. Note that while different slices of the potentials at fixed ξ are differentially decoupled (i.e., there is no ∂_ξ derivative), they remain coupled through the particle trajectories which determine $\bar{\rho}$ and $\bar{\mathbf{J}}$ and which do not obey a differentially decoupled evolution equation. Also note that Eqs. 2.58 and 2.59 can be combined to write a Poisson-like equation for the wake potential:

$$\nabla_\perp^2 \bar{\psi} = -4\pi [\bar{\rho} - \bar{J}_z/c] \quad (2.61)$$

If $\rho - J_z/c$ is specified, then this equation can be solved to obtain $\bar{\psi}$ given $\bar{\rho} - \bar{J}_z/c$ and $\bar{\mathbf{A}}_\perp$ can then be obtained by integrating the gauge condition Eq. 2.60.

Rather than the macroparticle description for the particles leading to the source terms in Eq. 2.18 and 2.19, it is more natural under the quasistatic approximation to describe the plasma as line charges representing a continuum of point particles in the longitudinal direction. Discreteness is retained in the transverse direction. In (ξ, s) coordinates, the charge and current density become

$$\bar{\rho}(s, \xi, \mathbf{x}_\perp) = \sum_j \Lambda_j(s, \xi) S(\mathbf{x}_\perp - \bar{\mathbf{x}}_{\perp j}(s, \xi)) \quad (2.62)$$

$$\bar{\mathbf{J}}(s, \xi, \mathbf{x}_\perp) = \sum_j \Lambda_j(s, \xi) \bar{\mathbf{v}}_j(s, \xi) S(\mathbf{x}_\perp - \bar{\mathbf{x}}_{\perp j}(s, \xi)) \quad (2.63)$$

where $\bar{\mathbf{x}}_{\perp j}(s, \xi)$ and $\bar{\mathbf{v}}_j(s, \xi)$ have been promoted to a continuum in s . They satisfy Eqs. 2.49 and 2.50 with $D_{\bar{\xi}_j} \rightarrow \partial_\xi$ and with s parameterizing the propagation of the system.

The form of the above source terms is constrained by conservation of charge. From the slowly-varying part of the continuity equation for each species, we have

$$\partial_t \bar{\rho} + \nabla \cdot \bar{\mathbf{J}} = 0 \quad (2.64)$$

Applying the change of variables $(\xi, s) = (z - ct, z)$ and the quasistatic approximation ($\partial_s \ll \partial_\xi$), this becomes

$$\partial_\xi(\bar{J}_z - \bar{\rho}) + \nabla_\perp \cdot \bar{\mathbf{J}}_\perp = 0 \quad (2.65)$$

Integrating over the transverse coordinate and taking the plasma source terms to be zero upstream of the laser, this becomes

$$\int (\bar{J}_z - \bar{\rho}) d\mathbf{x}_\perp = 0 \quad (2.66)$$

Thus, $\int (\bar{J}_z - \bar{\rho}) d\mathbf{x}_\perp$ is globally conserved. Inserting the forms of Eq. 2.68 and 2.69, we obtain

$$\partial_\xi \sum_j (\bar{v}_{zj}(s, \xi) - 1) \Lambda_j(s, \xi) = 0 \quad (2.67)$$

This equation can only hold identically if $\Lambda_j(s, \xi) = \Lambda_{0j}(s)/(1 - \bar{v}_z(s, \xi))$, i.e., the ξ dependence of the line charge density is determined by $\bar{v}_z$. $\Lambda_{0j}(s)$ parameterizes the line charge density upstream of the laser for each line charge and varies in s if there is a density gradient. Assuming for simplicity that $\Lambda_{0j}(s) = \Lambda_0(s)$ is the same for each particle, the slowly-varying charge and current density become

$$\bar{\rho}(s, \xi, \mathbf{x}_\perp) = \sum_j \Lambda_0(s) S(\mathbf{x}_\perp - \bar{\mathbf{x}}_{\perp j}(s, \xi)) \left(\frac{1}{1 - \bar{v}_{zj}(s, \xi)} \right) \quad (2.68)$$

$$\bar{\mathbf{J}}(s, \xi, \mathbf{x}_\perp) = \sum_j \Lambda_0(s) \bar{\mathbf{v}}_j(s, \xi) S(\mathbf{x}_\perp - \bar{\mathbf{x}}_{\perp j}(s, \xi)) \left(\frac{1}{1 - \bar{v}_{zj}(s, \xi)} \right) \quad (2.69)$$

The need for these factors of $1/(1 - \bar{v}_{zj})$ was recognized by Mora and Antonsen in Ref. [26]. The connection to conservation of charge was made explicit in Ref. [79]. The factors can be understood as an accounting for the varying inter-particle spacing within the line charges. For example, particles that originate with some inter-particle spacing in a region where $\bar{v}_z = 0$ will be closer together in regions where $\bar{v}_z > 0$. If this factor were not included, the longitudinal density of particles would be the same everywhere as a function of ξ for fixed s .

Under the quasistatic approximation, the equations for the fields in terms of the potentials

also simplify. In general, the full fields are related to the potentials by

$$\mathbf{E} = -\nabla\phi - \frac{1}{c}\partial_t\mathbf{A} \quad (2.70)$$

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (2.71)$$

Therefore the slowly-varying fields satisfy

$$\bar{\mathbf{E}} = -\nabla\bar{\phi} - \frac{1}{c}\partial_t\bar{\mathbf{A}} \quad (2.72)$$

$$\bar{\mathbf{B}} = \nabla \times \bar{\mathbf{A}} \quad (2.73)$$

Under the quasistatic approximation, $\bar{\mathbf{B}}$ has the same form with $\partial_z \rightarrow \partial_\xi$ in the curl operator, while $\bar{\mathbf{E}}$ satisfies

$$\bar{E}_z = -\partial_\xi\bar{\psi} \quad (2.74)$$

$$\bar{\mathbf{E}}_\perp = -\nabla_\perp\bar{\phi} + \partial_\xi\bar{\mathbf{A}}_\perp \quad (2.75)$$

where the wake potential ψ has finally entered. The slowly-varying Lorentz force

$$\bar{\mathbf{F}} = q \left[\bar{\mathbf{E}} + \frac{\bar{\mathbf{v}}}{c} \times \bar{\mathbf{B}} \right] \quad (2.76)$$

can be expressed in terms of the slowly-varying potentials under the quasistatic approximation, in cylindrical coordinates, and under cylindrical symmetry as

$$\bar{F}_z = q \left[-\partial_\xi\bar{\psi} + \left(\frac{\bar{v}_r}{c} \right) \partial_\xi\bar{A}_r - \left(\frac{\bar{v}_r}{c} \right) \partial_r\bar{A}_z \right] \quad (2.77)$$

$$\bar{F}_r = q \left[-\partial_r\bar{\phi} + \left(1 - \frac{\bar{v}_z}{c} \right) \partial_\xi\bar{A}_r + \frac{\bar{v}_z}{c} \partial_r\bar{A}_z \right] \quad (2.78)$$

$$\bar{F}_\varphi = 0 \quad (2.79)$$

In what follows, $\bar{\mathbf{F}}$ always refers to the Lorentz force rather than the total force $\bar{\mathbf{F}} + \bar{\mathbf{F}}_p$, where $\bar{\mathbf{F}}_p$ is the ponderomotive force. In addition, the Lorentz force for beam particles ($v_z \approx c$, $|\mathbf{v}_\perp| \ll c$) can be approximated with no assumption of cylindrical symmetry as

$$\bar{F}_z \approx -q\partial_\xi\bar{\psi} \quad (2.80)$$

$$\bar{\mathbf{F}}_\perp \approx -q\nabla_\perp\bar{\psi} \quad (2.81)$$

Thus, the dynamics of the beam particles are completely determined by the wake potential. The force fields in these equations have the following property:

$$\partial_\xi \bar{\mathbf{F}}_\perp = \nabla_\perp \bar{F}_z. \quad (2.82)$$

This is the so-called Panofsky-Wenzel theorem for relativistic plasma wave wakefields [109].

2.4.4 Summary

In summary, within the quasistatic approximation, the most significant simplification occurs for the slowly-varying fields. Maxwell's equations for the slowly-varying fields reduce to 2D Poisson-like equations, which simplifies theory and computation considerably. The equations developed in this section can be discretized to implement quasistatic PIC algorithms, discussed in more detail in Sec. 2.10.3. Developing a numerical scheme to solve the equations in this section is complicated; the equations in this section are not necessarily the ones discretized. For example, the wave equations for the electric and magnetic fields under the quasistatic approximation are discretized Ref. [80].

2.5 Dimensionless quantities

For most of the rest of this chapter and much of this dissertation, we will use dimensionless quantities to simplify the algebra. Since we only consider electron dynamics, all quantities are naturally normalized using the electron charge e , electron mass m_e , and initial electron density n_0 . These parameters give rise to the plasma frequency $\omega_p = \sqrt{4\pi n_0 e^2 / m}$ and the plasma wavenumber $k_p = \omega_p / c$. In terms of the dimensionless variables, factors of 4π in Maxwell's equations become unity. Table 2.1 shows how all relevant quantities are normalized using these characteristic parameters.

In order to distinguish each quantity from its normalized version, additional notation could be added, e.g., $\mathbf{A}_N \equiv e\mathbf{A}/m_e c^2$. In practice, the same variable can be used to represent

Variable	Dimensionless Quantity
Time	$\omega_p t$
Position	$k_p \mathbf{x}$
Velocity	$\mathbf{v}/c$
Momentum	$\mathbf{p}/m_e c$
Charge	q/e
Mass	m/m_e
Density	n/n_0
Current Density	$\mathbf{J}/ec$
Force	$\mathbf{F}/m_e c \omega_p$
Electric Field	$e\mathbf{E}/m_e c \omega_p$
Magnetic Field	$e\mathbf{B}/m_e c \omega_p$
Scalar potential	$e\phi/m_e c^2$
Vector potential	$e\mathbf{A}/m_e c^2$

Table 2.1: Summary of how quantities are normalized according to electron plasma parameters.

the dimensionless quantity, understanding that dimensional factors can be added back. In addition, because the laser and plasma evolution can be fully described in terms of slowly-varying quantities in the PGC and quasistatic approximations, all equations of relevance only include slowly-varying quantities. Thus, bars can be dropped. For both of these notational choices, the notation is simplified at the expense of possible ambiguity.

As an example of this notation, Eq. 2.56 becomes

$$1 - v_z = \frac{2(1 + \psi)^2}{1 + \mathbf{p}_\perp^2 + \frac{1}{2}|\hat{A}|^2 + (1 + \psi)^2}, \quad (2.83)$$

which is considerably easier to read. This notation is used for most of the following calculations in this chapter. Unnormalized quantities are used occasionally for clarity or for equations

which are easily expressed explicitly in dimensionless form (e.g., $k_p w_0 = 2\sqrt{a_0}$). Words will be used to clarify any ambiguity.

2.6 Structure of the plasma wake

In order to understand how a particle bunch is accelerated in a plasma wake, it is necessary to understand the structure of the plasma wake, including both the plasma charge and current densities and the electromagnetic fields. A predictive theory describing the structure of the plasma wake in the blowout regime was first worked out by Lu et al. in Ref. [97] and expanded on in Ref. [98]. The key finding of this work was that the fields and plasma source terms of the wake can be well-approximated based on the facts that (1) $\int(\rho - J_z) d\mathbf{x}_\perp = 0$ is globally conserved (Eq. 2.66), (2) a simple function can be used to model the $\rho - J_z$ profile and the results are insensitive to the parameterization, (3) the shape of the bubble is well-approximated by the trajectory of a single plasma electron moving within the sheath, and (4) the fields within the bubble can be expressed in terms of the parameters of the model for $\rho - J_z$, including the blowout radius $r_b(\xi)$. Here a simplified version of the analysis in these papers is presented in order to explain and shed light on the calculation. More complex models can be built using the same approach, as in Ref. [110]. A summary of the more general treatment in Refs. [97] and [98] and comments on the slightly different results in Ref. [111] are also provided.

The analysis begins by prescribing a model for the source terms for the plasma wake. After discussing the structure of the sheath, Refs. [97] and [98] used the following approximation for the structure of the plasma source terms including both electrons and ions (in dimensionless

units):

$$\rho(\xi, r) = 1, \quad r < r_b(\xi) \quad (2.84)$$

$$J_z(\xi, r) = 0, \quad r < r_b(\xi) \quad (2.85)$$

$$\rho(\xi, r) - J_z(\xi, r) = \begin{cases} 1, & r \in [0, r_b(\xi)) \\ -n_\Delta, & r \in [r_b(\xi), r_b(\xi) + \Delta) \\ 0, & r \in [r_b(\xi) + \Delta, \infty) \end{cases} \quad (2.86)$$

where $r_b(\xi)$ is the bubble radius and Δ is the thickness of the electron sheath. In words, immobile ions of uniform density are assumed to be present everywhere. No electrons are present for $r \leq r_b(\xi)$, followed by an electron sheath of constant $\rho - J_z$ and thickness Δ , followed by uniform electrons (i.e., quasineutral plasma). Hence $\rho = \rho_e + \rho_i$ and $J_z = J_{ze}$. Cylindrical symmetry is assumed. The sheath thickness Δ can itself be parameterized as a function of ξ , which is done in Refs. [97] and [98] as

$$\Delta(\xi) = \epsilon r_b(\xi) + \Delta_L \quad (2.87)$$

where ϵ and Δ_L were assumed to not vary with ξ for simplicity.¹ Note that ρ and J_z are not independently specified for $r \geq r_b(\xi)$ in this model. It is shown below that their independent specification is not needed globally to calculate the fields within the bubble. Conversely, global calculation of the fields requires global independent specification of ρ and J_z . Independently specifying them globally also requires an additional model for $\mathbf{J}_\perp(\xi, r)$ in order for the continuity equation to be satisfied within the sheath.

In addition to the plasma source terms, those of a particle beam can be added, which is approximated in Refs. [97] and [98] as

$$\rho_b(\xi, r) = -n_b e^{-r^2/\sigma_r^2} e^{-\xi^2/\sigma_z^2} \quad (2.88)$$

$$J_{zb}(\xi, r) \approx \rho_b(\xi, r) \quad (2.89)$$

¹The Δ_L parameter models the linear response and has a width $\sim k_p^{-1}$ while $\epsilon r_b(\xi)$ models the nonlinear response. For $r_b(\xi)$ sufficiently large, the nonlinear part can dominate. In the rear, where $r_b(\xi)$ is small, the linear response can dominate.

where the equation for J_z assumes the beam propagates forward with $v_{zb} \approx 1$. Note that $\rho_b - J_{zb} = O(1/\gamma_b^2) \approx 0$ where $\gamma_b = 1/\sqrt{1 - v_{zb}^2/c^2}$, so so that the beam does not act as source term for ψ . It is assumed that $\sigma_r \ll r_b(\xi)$ for all ξ .

In order to simplify the presentation mathematically, we consider Eq. 2.86 under the limit $\Delta \rightarrow 0$ from the beginning. While this is an unphysical assumption, particularly near the rear of the bubble, it is instructive to consider this limit. The results obtained in Refs. [97] and [98] without taking this limit will be summarized later. Under this limit, Eq. 2.86 becomes

$$\rho(\xi, r) - J_z(\xi, r) = I_{[0, r_b(\xi)]}(r) - n_\Delta r_b(\xi) \delta(r - r_b(\xi)) \quad (2.90)$$

where $I_{[a, b]}(x)$ is the indicator function on $[a, b]$ defined in Eq. 2.1 and the notation n_Δ is retained to distinguish n_Δ from the number density n . We refer to this model as a “delta function sheath” and Lu’s model in Eq. 2.86 as a “step function sheath”. In what follows, we follow the same procedure given in Refs. [97] and [98] except here this delta function profile is used.

In Eq. 2.90, the n_Δ parameter is constrained by conservation of charge. Expressing Eq. 2.66 in cylindrical coordinates and taking the constant of integration to be 0, we have $\int_0^\infty (J_z - \rho) r dr = 0$. Evaluating the integral and solving for n_Δ yields $n_\Delta = \frac{1}{2}$.

The next step is to calculate ϕ , $\mathbf{A}$, and ψ from the Poisson-like equations derived in Sections 2.3 and 2.4 with the source terms in Eqs. 2.84, 2.85, and 2.90. Expressions for ϕ and A_z can be computed independently for $r < r_b(\xi)$ from Eqs. 2.84 and 2.85 using the Poisson-like equations $\nabla_\perp^2 \phi = -\rho$ and $\nabla_\perp^2 A_z = -J_z$ (Eqs. 2.58 and 2.59 for the dimensionless quantities). Integrating the Poisson-like equations from 0 under cylindrical symmetry, we obtain

$$\phi(\xi, r) = \phi_0(\xi) - \frac{r^2}{4} + \Lambda(\xi) \log r, \quad r \leq r_b(\xi) \quad (2.91)$$

$$A_z(\xi, r) = A_{z0}(\xi) + \Lambda(\xi) \log r, \quad r \leq r_b(\xi) \quad (2.92)$$

where ϕ_0 and A_{z0} are the on-axis profiles of ϕ and A_z and $\Lambda(\xi)$ is the beam current defined as

$$\Lambda(\xi) \equiv \int_0^\infty r n_b(\xi, r) dr. \quad (2.93)$$

The on-axis profiles and the structure of ϕ and A_z for $r \geq r_b(\xi)$ cannot be explicitly computed without knowing the global structure of ρ and J_z independently. However, they are not needed because only $\partial_r \phi$ and $\partial_r A_z$ enter into the Lorentz force inside the bubble under the quasistatic approximation (Eqs. 2.77–2.78). These vanish for the on-axis terms.

Next, the Poisson-like equation $\nabla_\perp^2 \psi = -(\rho - J_z)$ (Eq. 2.61) can be integrated to obtain ψ . Writing $\nabla_\perp^2 \psi = \frac{1}{r} \partial_r (r \partial_r \psi)$ and successively integrating from ∞ (instead of from 0 as for ϕ and A_z) and defining $\psi(\infty) = 0$, we obtain

$$\psi(\xi, r) = \begin{cases} \frac{1}{4} [r_b^2(\xi) - r^2], & r \leq r_b(\xi) \\ 0, & r \geq r_b(\xi) \end{cases} \quad (2.94)$$

Note that although ψ and A_z cannot be explicitly specified on-axis, ψ can: $\psi_0(\xi) \equiv \psi(\xi, 0) = \frac{1}{4} r_b(\xi)^2$. Also, note that $\Lambda(\xi)$ does not appear because $\rho_b - J_{zb} = 0$.

Finally, $\mathbf{A}_\perp$ can be computed from the gauge condition $\partial_\xi \psi + \nabla_\perp \cdot \mathbf{A}_\perp = 0$ (Eq. 2.60). Using ψ in Eq. 2.94, expressing $\nabla_\perp \cdot \mathbf{A}_\perp = (1/r) \partial_r (r A_r)$ from the assumption of cylindrical symmetry, and using the fact that $\mathbf{A}_\perp$ vanishes as $r \rightarrow \infty$, $\mathbf{A}_\perp$ may be computed as

$$\mathbf{A}_\perp = \begin{cases} \left(\frac{1}{4}\right) r r_b(\xi) r'_b(\xi) \hat{\mathbf{r}}, & r \leq r_b(\xi) \\ 0, & r \geq r_b(\xi) \end{cases} \quad (2.95)$$

where $r'_b(\xi) \equiv D_\xi r_b(\xi)$.

The next step is to express the force experienced by electrons using the potentials. Inserting the expressions for ϕ , A_z , $\mathbf{A}_\perp$, and ψ into the expressions $F_z = \partial_\xi \psi - v_r \partial_\xi A_r + v_r \partial_r A_z$ and $F_r = \partial_r \phi - (1 - v_z) \partial_\xi A_r - v_z \partial_r A_z$ (Eqs. 2.77 and 2.78, which hold under cylindrical symmetry),

we obtain expressions for the components of the Lorentz force in terms of $r_b(\xi)$:

$$F_z(\xi, r) = \left(\frac{1}{2}\right) r_b(\xi) r'_b(\xi) + \left(\frac{1}{4}\right) r v_r D_\xi [r_b(\xi) r'_b(\xi)] + v_r \frac{\Lambda(\xi)}{r}, \quad r < r_b(\xi) \quad (2.96)$$

$$F_r(\xi, r) = -\left(\frac{1}{2}\right) r + \left(\frac{1}{4}\right) r (v_z - 1) D_\xi [r_b(\xi) r'_b(\xi)] - (v_z - 1) \frac{\Lambda(\xi)}{r}, \quad r < r_b(\xi) \quad (2.97)$$

Note that for beam electrons, with $v_z \approx 1$ and $v_r \approx 0$, the self-force is negligible.

The derivation will continue by finding a differential equation that constrains the form of $r_b(\xi)$. However, it is worth pausing to note two insights that can be obtained from Eqs. 2.96 and 2.97 about the behavior of electron beams (which may be either a drive beam or an accelerating beam) in an azimuthally symmetric wake in the blowout regime. First, for a relativistic electron ($v_z \approx 1$, $v_r \approx 0$), the focusing force is linear in r and independent of ξ (and in particular independent of $r_b(\xi)$):

$$\mathbf{F}_\perp(\xi, r) \approx -\left(\frac{r}{2}\right) \hat{\mathbf{r}} \quad (2.98)$$

This is true both for electrons in a drive beam and for those accelerating in the wake. The specific consequences of this linear transverse focusing force on an accelerating particle beam will be discussed in Sec. 2.9. Second, the accelerating force $F_z \approx \frac{1}{2} r_b(\xi) r'_b(\xi)$ generally depends on the shape of the bubble, but by inspection (Fig. 2.1), the bubble can be approximately described by a sphere, i.e., $r_b(\xi)^2 + \xi^2 = R_b^2$ where R_b is the radius of the bubble, with slight deviation visible near the rear of the wake where $r_b(\xi) \approx 0$. (We will see below that the near-spherical profile of $r_b(\xi)$ is predicted by the theory of Lu et al. using either the delta function sheath or the step function sheath.) Using this functional form, the accelerating force within the bubble may be approximated as

$$F_z(\xi, r) \approx -\frac{\xi}{2} \quad (2.99)$$

Thus, electrons behind the midpoint of the bubble which copropagate with the structure will be continuously accelerated while they are focused onto the axis. The force is independent of r , so all electrons at the same ξ are accelerated at the same rate. However, the equation

for F_z suggests that electrons will also gain a different amount of energy depending on their phase within the wake, leading to an energy spread within an accelerating beam. This poses a challenge for application in a collider or an XFEL, for which it is desirable for the energy spread of the beam to be under 1%. However, this effect can be addressed by “beam loading” of the wake, or tailoring of the beam profile to make the accelerating field constant through feedback of the beam on the wake [112, 99]. This is discussed in Sec. 2.8.

We next show how a differential equation describing the shape of the bubble can be derived, following the approach of Refs. [97] and [98] for the simplified delta function sheath in Eq. 2.90. The key idea is that the equation of motion for plasma particles in Eq. 2.27 is satisfied for electrons moving along the boundary of the plasma sheath. The forces calculated in Eq. 2.96 and Eq. 2.97 are expressed in terms of $r_b(\xi)$. The momentum of the particles along the sheath is also related to $r_b(\xi)$ from the assumption that the trajectory of a single plasma electron traces out the shape of the bubble. This leads to a closed differential equation for $r_b(\xi)$. For simplicity, we perform this derivation in the regions where there is no laser, i.e., $|\hat{A}|^2 = 0$.

It suffices to consider the transverse evolution. The equation for the evolution of the transverse momentum is

$$\dot{\mathbf{p}}_{\perp} = \mathbf{F}_{\perp} \quad (2.100)$$

where $\mathbf{F}_{\perp} = F_r \hat{\mathbf{r}}$ with F_r in Eq. 2.97 evaluated along the inner radius of the sheath, i.e., $r = r_b(\xi)^-$. Subtleties associated with considering the inner instead of outer radius are discussed in the next section. The transverse momentum $\mathbf{p}_{\perp}$ is related to r_b from the constraint that the particle trajectory follows the sheath. We have $\mathbf{p}_{\perp} = \gamma \mathbf{v}_{\perp}$. But along the bubble trajectory, $\mathbf{v}_{\perp} = \dot{r}_b \hat{\mathbf{r}}$. As before, the time derivative can be expressed as $\dot{r}_b = (v_z - 1)r'_b(\xi)$. Combining and using the relation $\gamma - p_z = 1 + \psi$ from Eq. 2.52, we obtain

$$p_r = -(1 + \psi)r'_b(\xi) \quad (2.101)$$

and similarly

$$\dot{p}_r = -(v_z - 1)D_\xi [(1 + \psi)r'_b(\xi)] \quad (2.102)$$

where ψ is evaluated along the particle trajectory, i.e., $\psi = \psi(\xi, r_b(\xi))$. In addition, the $1 - v_z$ factor does not cancel, so an equation for it in terms of $r_b(\xi)$ is needed. Inserting Eq. 2.101 into Eq. 2.56, we obtain

$$1 - v_z = \frac{2(1 + \psi)^2}{1 + (1 + \psi)^2 + [(1 + \psi)r'_b(\xi)]^2} \quad (2.103)$$

Eqs. 2.101–2.103 hold for any ψ . For the delta function sheath, we have $\psi(\xi, r_b(\xi)) = 0$ from Eq. 2.94. This results in simpler algebra than in the case of the step function sheath. Equating Eqs. 2.102 and 2.97 and evaluating the latter at $r = r_b(\xi)^-$ (where $f(r_b(\xi)^-) \equiv \lim_{r \rightarrow r_b(\xi)^-} f(r)$, i.e., the limit from below), we obtain the following second-order differential equation for $r_b(\xi)$:

$$\left(1 + \frac{1}{4}r_b^2\right)r_b'' + \left(\frac{1}{2}\right)r_b(r_b')^2 + \left(\frac{1}{2}\right)r_b = \frac{\Lambda(\xi)}{r_b} \quad (2.104)$$

This equation closes the model. (Strictly speaking, the force outside the sheath, which is nonzero for the delta function, must also be included. This is done in the next subsection. But the calculation up to now illustrates the key flow of logic.)

In the above, the delta function sheath model was used to aid in explication of the calculation, where it was known from the start that such a model cannot accurately describe the wake when $r_b(\xi)$ is small. Here we summarize the results from Refs. [97] and [98] when the step function profile of Eq. 2.86 is used for $\rho - J_z$ in the bubble sheath. The equations for ϕ and A_z are unmodified from Eqs. 2.91 and 2.92. (The on-axis potentials ϕ_0 and A_{z0} are themselves different, but still do not appear in the Lorentz force). The equation for ψ is different:

$$\psi(\xi, r) = \frac{r_b(\xi)^2}{4}(1 + \beta) - \frac{r^2}{4} \quad (2.105)$$

where

$$\beta \equiv \frac{(1 + \alpha)^2 \log [(1 + \alpha)^2]}{(1 + \alpha)^2 - 1} - 1 \quad (2.106)$$

and

$$\alpha \equiv \frac{\Delta}{r_b(\xi)} \quad (2.107)$$

which gives $\alpha = \epsilon + \Delta_L/r_b(\xi)$. It can be shown that $\beta \approx \alpha$ for $\alpha \ll 1$ and $\beta \approx 2 \log \alpha$ for $\alpha \gg 1$. Applying force balance as before, one obtains the following differential equation for r_b , where the laser envelope $\hat{A}$ is also now included for completeness:

$$A(r_b)r_b'' + B(r_b)r_b(r_b')^2 + C(r_b)r_b = \frac{\Lambda(\xi)}{r_b} - \frac{\frac{1}{4}\partial_r|\hat{A}|^2}{1 + \frac{\beta}{4}r_b^2} \quad (2.108)$$

where

$$A(r_b) = 1 + \left[\frac{1}{4} + \frac{\beta}{2} + \frac{1}{8}r_b\partial_{r_b}\beta \right] r_b^2 \quad (2.109)$$

$$B(r_b) = \frac{1}{2} + \frac{3}{4}\beta + \frac{3}{4}r_b\partial_{r_b}\beta + \frac{1}{8}r_b^2\partial_{r_b}^2\beta \quad (2.110)$$

$$C(r_b) = \frac{1}{4} \left[1 + \frac{1 + \frac{|\hat{A}|^2}{2}}{(1 + \frac{\beta}{4}r_b^2)^2} \right] \quad (2.111)$$

and $\hat{A}$ is evaluated along the particle trajectory. As $\Delta \rightarrow 0$, $\beta \rightarrow \alpha = \Delta/r_b$. Hence $\partial_{r_b}\beta \approx -\Delta/r_b^2$ and $\partial_{r_b}^2\beta \approx 2\Delta/r_b^3$. As $\Delta \rightarrow 0$, $\beta \rightarrow 0$, $\partial_{r_b}\beta \rightarrow 0$, and $\partial_{r_b}^2\beta \rightarrow 0$. Thus $A(r_b) \rightarrow 1 + \frac{1}{4}r_b^2$, $B(r_b) \rightarrow \frac{1}{2}$, and $C(r_b) \rightarrow \frac{1}{2}$ so that Eq. 2.104 is recovered from Eq. 2.108. Lu et al. also discuss the ultrarelativistic limit, where $r_b \gg 1$, $\Delta \ll 1$, but $\Delta r_b \gg 1$. In this limit, one obtains the equation

$$\left(1 + \frac{1}{4}r_b^2\right) r_b'' + \left(\frac{1}{2}\right) r_b(r_b')^2 + \left(\frac{1}{4}\right) r_b = \frac{\Lambda(\xi)}{r_b} \quad (2.112)$$

This model is used in the section on beam loading. This equation is the same as for the limit $\Delta \rightarrow 0$ when only considering the inner force except with a term of $r_b/4$ instead of $r_b/2$.

In both the delta function and the step function models for the sheath, $\psi \geq 0$ everywhere. Kinetic simulations show that $\psi \approx -1$ near the rear of the plasma bubble. Thus, both of these models have limited accuracy near the rear of the plasma bubble. If $\rho - J_z = 1$ for $r \leq r_b(\xi)$, then a negative ψ can only be obtained if $\rho - J_z$ is not negative-definite for $\rho - J_z \geq r_b(\xi)$. This condition does not hold in the original step function sheath model

(Eq. 2.86). Assuming the ions remain stationary, this can only occur if $\rho_e - J_{ze} > -1$ for electrons with $r > r_b(\xi) + \Delta$. This is indeed satisfied for electrons with a linear response outside the sheath where $v_z > 0$ and $\rho \approx -1$. Reference [110] shows that this can be modeled using the same approach presented in this section with a multi-sheath parameterization of the $\rho - J_z$ function. In this case, a second step function with $\rho - J_z > 0$ is added for $r \geq r_b(\xi) + \Delta$. For the appropriate parameters, this leads to a solution for ψ which may be negative. This model is capable of more accurately modeling the dynamics near the wake rear.

In summary, the function of the models presented in this section is to provide a predictive theory that identifies the key features of the wake. They provide a theoretical and conceptual framework for understanding key physics including acceleration of a witness beam, beam loading, and the effect of pulse-shaping for high transformer ratios. An area for future work would be to write an algorithm that solves for $r_b(\xi)$ from a differential equation (e.g., Eq. 2.104, Eq. 2.108, or a more complex equation in the case of Ref. [110]). Using $r_b(\xi)$, one could then evolve beam particles, advance the laser driver, and repeat the process. This method would not be as accurate as PIC or quasistatic PIC simulations. However, it could be reasonably accurate in certain situations and appears to not have been deeply explored.

Further analysis of the delta function sheath model

The author originally investigated the delta function sheath model (Eq. 2.90) in order to better understand and explicate the calculations in Refs. [97] and [98], as presented in the previous section. Afterward, it was realized that these calculations could potentially shed light on a discrepancy that exists between Refs. [97] and [98] and Ref. [111], where similar but slightly different differential equations for $r_b(\xi)$ are found. Here, the details of the delta function sheath model are worked out in more detail and a brief discussion of the discrepancy is provided.

Unlike in the step function sheath model, ρ and J_z can be specified when using the delta

function sheath model with no further assumptions because all the electrons are in the sheath:

$$\rho(\xi, r) = I_{[0, r_b(\xi)]}(r) + \rho_S(\xi) r_b(\xi) \delta(r - r_b(\xi)) \quad (2.113)$$

$$J_z(\xi, r) = J_{zS}(\xi) r_b(\xi) \delta(r - r_b(\xi)) \quad (2.114)$$

where $\rho_S(\xi)$ and $J_{zS}(\xi)$ account for variation of the line charge density within the sheath. Reference [111] uses the same model but with the notation $\rho_0(\xi)$ instead of $\rho_S(\xi)$.

The continuity equation leads to $\int_0^\infty (J_z - \rho) r dr = 0$ as before, which now provides

$$\rho_S(\xi) - J_{zS}(\xi) = -\frac{1}{2} \quad (2.115)$$

The global expressions for ψ and $\mathbf{A}_\perp$ are unmodified from Eqs. 2.94 and 2.95. On the other hand, ϕ and A_z can now each be computed globally from the global structure of ρ and J_z . Here the calculation is shown explicitly for ϕ for clarity. Neglecting any particle beams for now, the Poisson-like equation $\nabla_\perp^2 \phi = -\rho$ (Eq. 2.58) under cylindrical symmetry provides

$$\frac{1}{r} \partial_r (r \partial_r \phi) = -\rho_S(\xi) r_b(\xi) \delta(r - r_b(\xi)) - \begin{cases} 1, & r \leq r_b(\xi) \\ 0, & r > r_b(\xi) \end{cases} \quad (2.116)$$

Multiplying by r and integrating from 0, we obtain

$$r \partial_r \phi = \begin{cases} -\frac{r^2}{2}, & r \leq r_b(\xi) \\ -\frac{r_b(\xi)^2}{2} - \rho_S(\xi) r_b(\xi)^2, & r > r_b(\xi) \end{cases} \quad (2.117)$$

Dividing by r and integrating from 0 yields

$$\phi(\xi, r) = \phi_0(\xi) + \begin{cases} -\frac{r^2}{4}, & r \leq r_b(\xi) \\ -\frac{r_b(\xi)^2}{4} - \left[\frac{r_b(\xi)^2}{2} + \rho_S(\xi) r_b(\xi)^2 \right] \log \left(\frac{r}{r_b(\xi)} \right), & r > r_b(\xi) \end{cases} \quad (2.118)$$

where $\phi_0(\xi) \equiv \phi(\xi, 0)$ as before. Adding in the potential contributed by the beam and performing the analogous integrations for A_z , we obtain

$$\phi(\xi, r) = \phi_0(\xi) + \Lambda(\xi) \log r + \begin{cases} -\frac{r^2}{4}, & r \leq r_b(\xi) \\ -\frac{r_b(\xi)^2}{4} - \left[\frac{r_b(\xi)^2}{2} + \rho_S(\xi) r_b(\xi)^2 \right] \log \left(\frac{r}{r_b(\xi)} \right), & r > r_b(\xi) \end{cases} \quad (2.119)$$

and

$$A_z(\xi, r) = A_{z0}(\xi) + \Lambda(\xi) \log r + \begin{cases} 0, & r \leq r_b(\xi) \\ -J_{zS}(\xi) r_b(\xi)^2 \log\left(\frac{r}{r_b(\xi)}\right), & r > r_b(\xi) \end{cases} \quad (2.120)$$

which agree with Eqs. 2.91 and 2.92 for $r < r_b(\xi)$.

Using these potentials, the radial component of force computed on the inner side of the sheath, $F_r(\xi, r_b(\xi)^-)$, is the same as in Eq. 2.97. However, the outer component can now be computed as

$$F_r(\xi, r_b(\xi)^+) = (v_z - 1)J_{zS}(\xi)r_b(\xi) + (v_z - 1) \left(\frac{1}{4}\right) [3r_b(\xi)r_b'(\xi)^2 + r_b(\xi)^2 r_b''(\xi)] - (v_z - 1) \frac{\Lambda(\xi)}{r} \quad (2.121)$$

Further progress requires an expression for $J_{zS}(\xi)$. Expressions for $J_{zS}(\xi)$, $\rho_S(\xi)$, and all kinematic quantities in terms of $r_b(\xi)$ can be obtained through the assumptions of the model. These are now computed. From the fact that a particle is assumed to follow the trajectory $r_b(\xi)$, the radial velocity is related to the bubble radius as $v_r = (1 - v_z)r_b'(\xi)$. Hence, $p_r = \gamma v_r = \gamma(1 - v_z)r_b'(\xi)$. But $\gamma - p_z = 1 + \psi$ and $\psi(\xi, r_b(\xi)) = 0$, so

$$p_r(\xi) = r_b'(\xi) \quad (2.122)$$

From Eq. 2.54 and 2.55, we obtain

$$p_z(\xi) = \frac{1}{2} r_b'(\xi)^2 \quad (2.123)$$

$$\gamma(\xi) = 1 + \frac{1}{2} r_b'(\xi)^2 \quad (2.124)$$

Dividing p_r and p_z by γ , we also obtain

$$v_r = \frac{r_b'(\xi)}{1 + \frac{1}{2} r_b'(\xi)^2} \quad (2.125)$$

$$v_z = \frac{\frac{1}{2} r_b'(\xi)^2}{1 + \frac{1}{2} r_b'(\xi)^2} \quad (2.126)$$

Finally, the plasma source terms can be computed from the relationship

$$J_{zS}(\xi) = \rho_S(\xi) v_z(\xi) \quad (2.127)$$

This equation holds because in the delta function sheath model, all the electrons in the sheath at a given ξ have the same $v_z(\xi)$. This is not true in the step function sheath model, for which an analogous equation cannot be written. Combining Eq. 2.127 with Eq. 2.115 yields $J_{zS}(\xi) = -\frac{1}{2}v_z/(1 - v_z)$. Substituting the above equation for v_z yields

$$J_{zS}(\xi) = -\frac{1}{4}r'_b(\xi)^2 \quad (2.128)$$

Though not needed for this calculation, we calculate the charge density for completeness as

$$\rho_S(\xi) = -\frac{1}{2} - \frac{1}{4}r'_b(\xi)^2 \quad (2.129)$$

At this point, we have expressions for both plasma source terms and all the kinematic quantities in terms of $r_b(\xi)$.

Using Eq. 2.128 for $J_{zS}(\xi)$, the force on the outer side of the sheath may be computed as

$$F_r(\xi, r_b(\xi)^+) = (v_z - 1) \left[\frac{1}{2}r_b(\xi)r'_b(\xi)^2 + \frac{1}{4}r_b(\xi)^2r''_b(\xi) \right], \quad (2.130)$$

which is different from $F_r(\xi, r_b(\xi)^-)$ in Eq. 2.97 with $r = r_b(\xi)^-$. Formally, the force to use for the calculation when using the delta function is the average force across the sheath to account for the jump in force: $\bar{F}_r(\xi, r_b(\xi)) = \frac{1}{2} [F_r(\xi, r_b(\xi)^-) + F_r(\xi, r_b(\xi)^+)]$. When this is done, one obtains the following equation for $r_b(\xi)$:

$$\left(1 + \frac{1}{4}r_b^2\right)r''_b + \left(\frac{1}{2}\right)r_b(r'_b)^2 + \left(\frac{1}{4}\right)r_b = \frac{\Lambda(\xi)}{r_b} \quad (2.131)$$

Interestingly, this is the same as the ultrarelativistic limit from the step function sheath model obtained in Eq. 2.112. Note that one obtains a different model if just the inner force or outer force are used because of its discontinuity. At first, this issue appears not to emerge in the step function sheath model because the force is continuous at $r = r_b(\xi)$. However, it is present: the force on the inner radius of the sheath at $r = r_b(\xi)$ is different from the force on the outer sheath at $r = r_b(\xi) + \Delta$. Electrons moving along the inner and outer sheath therefore follow slightly different trajectories. In practice, the profile for the outer sheath radius cannot be calculated without additional assumptions for the step function model.

One could potentially account for this in a more accurate description through self-consistent calculation of the trajectories $r_j(\xi)$ for particles with different initial radii r_{j0} moving within the sheath.

Golovanov et al. (Ref. [111]) assumed the same delta function sheath model for ρ and J_z (i.e., Eqs. 2.113–2.114, the $\Delta \rightarrow 0$ limit of the step function sheath model from the beginning) but followed an approach based on energy conservation instead of force balance to derive a differential equation for r_b . This approach cannot be applied to the step function sheath model without further assumptions because it requires the global structure of the fields and plasma source terms. The kinematic quantities derived in Eqs. 2.122–2.129 enter into this derivation. The model of Golovanov et al. predicts a differential equation of the form

$$\left(1 + \frac{1}{4}r_b^2\right)r_b'' + \left(\frac{1}{r_b} + \frac{r_b}{2}\right)(r_b')^2 + \left(\frac{1}{4}\right)r_b = \frac{\Lambda(\xi)}{r_b} \quad (2.132)$$

The details of the calculation are not presented here, but much of the work is done through the calculation of the potentials and plasma source terms above.

Eq. 2.132 is the same as the averaged Eq. 2.131 but with an additional term $r_b^{-1}(r_b')^2$. Why is there a discrepancy? Mathematically, one way this can be understood is that the model implicitly assumes a form for the electron distribution function: $f(s, \xi, r, p_z, p_r) = n_S(s, \xi)r_b(s, \xi)\delta(r - r_b(s, \xi))p_r^{-1}\delta(p_r - p_r(s, \xi))\delta(p_z - p_z(s, \xi))$ where the functional dependence on s is made explicit for quantities that up to now have assumed s to be constant. Because of the finite number of free parameters and the fact that only the true Vlasov equation satisfies all the fluid equations, this distribution function cannot satisfy all the fluid equations. In fact, one can demonstrate that the equation of motion for a single particle and the fluid momentum equation are inconsistent under the assumption of this form of the distribution function. One can integrate the fluid momentum equation on its own to obtain a different differential equation for r_b than those of Eqs. 2.131 (particle force balance) and 2.132 (energy conservation). A higher-order fluid equation could also be used and obtain yet another different result. The inconsistency does not invalidate the model but necessitates care in its application.

As noted above, in actuality the plasma wake has a linear response outside the narrow region of the sheath consisting primarily of blown-out electrons. This is not captured in the delta function sheath model. In addition, as r_b becomes small in the rear of the bubble or for weak wakes, the assumption $\Delta \ll r_b$ is not accurate. In addition, the delta function model for the sheath cannot capture the negative values of ψ near the back of the wake, regardless of whether the approach of force balance or of energy conservation is used. Nonetheless, any approach based on energy conservation is attractive. The fact that both give nearly the same differential equation is reassuring. In the end, a detailed comparison of how the models compare with kinetic simulations will be the ultimate judge of their accuracy. While the results in this dissertation shed insight into this comparison, this remains an area for near-term investigation.

2.7 Laser evolution and its consequences: matching, depletion, and dephasing

Although the full evolution of a laser evolving according to Eq. 2.43 is analytically intractable in the blowout regime, the laser evolution and its consequences can be understood phenomenologically in terms of a few key concepts: laser matching, pump depletion, and dephasing. Here we summarize the results presented in Ref. [1], which synthesized these concepts and derived estimates for key quantities.

We begin by considering how the laser parameters can be matched to the plasma wake. Intuitively, the laser can be guided within a plasma wake in the same way that light is guided within a coaxial cable. This requires that the index of refraction $\eta = ck/\omega = \sqrt{1 - \left(\frac{\omega_0}{\omega_p}\right)^2 \frac{\Delta n}{n_p}} \approx 1 - \frac{1}{2} \left(\frac{\omega_0}{\omega_p}\right)^2 \frac{1}{1+\psi}$ be larger on axis and that the spot size w_0 of the laser to be comparable to the radius R of the bubble. The radius of the bubble in terms of the laser parameters can be estimated by equating the peak outward transverse ponderomotive force exerted by the laser on plasma electrons to the peak inward transverse force exerted by the wake.

From Eq. 2.30 in dimensionless units, the transverse ponderomotive force is $\mathbf{F}_{p,\perp} = -\frac{1}{4\bar{\gamma}}\nabla_{\perp}|\hat{A}|^2$. For simplicity we approximate the laser envelope as $\hat{A}(\xi, r) = a_0 \exp(-\xi^2/\tau_0^2) \exp(-r^2/w_0^2)$. Also from Eq. 2.29, the average Lorentz factor is $\bar{\gamma} = \sqrt{1 + \mathbf{p}^2 + \frac{1}{2}|\hat{A}|^2}$ and may be estimated as $\bar{\gamma} \approx \frac{1}{\sqrt{2}}|\hat{A}|$ for $a_0 \gtrsim 2$. Inserting the form for $\hat{A}$, we obtain the estimate

$$\mathbf{F}_{p,\perp}(\xi, r) \approx \left(\frac{\sqrt{2}a_0}{w_0^2} \right) r e^{-\xi^2/\tau_0^2} e^{-r^2/w_0^2} \quad (2.133)$$

The maximum of this function occurs at $(\xi, r) = (0, w_0/\sqrt{2})$. Hence we obtain the estimate $\max \mathbf{F}_{p,\perp} \approx \frac{1}{2} \frac{a_0}{w_0}$. On the other hand, the peak transverse force from the wake may be estimated from Eq. 2.97 as $\max |\mathbf{F}_{\perp}| \approx -R/2$. Equating the forces, we obtain the estimate (with factors of k_p inserted for clarity) $k_p R \approx a_0/k_p w_0$. An estimate for the laser spot size for optimal guiding is $w_0 = R$, which leads to the matching condition $k_p w_0 = \sqrt{a_0}$.² Through simulations, the authors of Ref. [1] tested this and found that the following matching condition leads to more stable guiding:

$$k_p w_0 = 2\sqrt{a_0} \quad (2.134)$$

This equation is a starting point for how to choose a_0 given w_0 or vice versa.

The details of the physics of guiding are subtle. One might expect that the front of the pulse is not guided because it does not propagate within the bubble. In practice, it was found that guiding using the matching condition in Eq. 2.134 is robust [1]. Reference [1] offers the interpretation from Ref. [113] that the head of the pulse locally pump depletes, or loses energy through transfer to the wake, before it diffracts, while the rear remains guided. This manifests as the phenomenon of “etching”: the laser pulse develops a sharp rising edge which gradually recedes over the course of propagation. Reference [113] showed using 1D nonlinear calculations that the velocity of “etching” of the laser can be estimated as $v_{\text{etch}} = (\omega_p/\omega_0)^2 c$. The authors of Ref. [1] confirmed through simulations that the estimate for v_{etch} is accurate

²However, this estimate is based on a steady-state assumption. When a harmonic oscillator of frequency ω_0 is driven by a driver with a duration on the order of ω_0^{-1} , the oscillation amplitude is ~ 2 times the equilibrium value. This leads to the estimate $k_p w_0 = 2\sqrt{a_0}$.

in 1D, 2D, and 3D simulations for $a_0 \in [2, 2(k_0/k_p)]$ and provide theoretical justification for the upper bound. The length scale over which all of the laser is depleted, known as the pump depletion length L_{pd} , may be estimated using this etching velocity and the laser duration τ as $L_{\text{pd}} = \tau/v_{\text{etch}}$, leading to the estimate

$$L_{\text{pd}} \approx c\tau \left(\frac{\omega_0}{\omega_p} \right)^2 \quad (2.135)$$

The pump depletion length is a key length scale for the physics of LWFA.

Another important length scale is the dephasing length L_d . The laser group velocity due to linear dispersion may be calculated as $v_g = c\sqrt{1 - (\omega_p/\omega_0)^2}$ so that

$$v_g \approx \left[1 - \frac{1}{2} \left(\frac{\omega_p}{\omega_0} \right)^2 \right] c \quad (2.136)$$

Concepts from the linear regime need to be applied with care to the nonlinear regime, but this is a reasonable estimate. Because of the etching of the laser and its additional recession due to dispersion ($v_g < c$), particles accelerating within the wake will eventually outrun the wake. This phenomenon is known as dephasing. The wake velocity v_w may be estimated as $v_w = v_g - v_{\text{etch}}$, leading to the estimate

$$v_w \approx \left[1 - \frac{3}{2} \left(\frac{\omega_p}{\omega_0} \right)^2 \right] c \quad (2.137)$$

For particles that start at the rear of the wake with velocity $v_z \sim c$, they will dephase when they have advanced $\Delta\xi = R$. But $\Delta\xi = (v_z - v_w)t$. Equating at $ct = L_d$, we find $L_d = Rc/(c - v_w)$. Plugging in Eq. 2.137 and using the estimate $R = 2\sqrt{a_0}$, we obtain the estimate

$$L_d \approx \frac{4}{3} \left(\frac{\omega_0}{\omega_p} \right)^2 \sqrt{a_0} \quad (2.138)$$

This length scale sets a limit on the acceleration length in LWFA in the blowout regime.

The energy gain for an electron accelerating in the bubble can be estimated by multiplying the accelerating gradient by the length of acceleration: $\Delta W \approx F_z L_{\text{acc}}$. We assume acceleration over the full dephasing length, i.e., $L_{\text{acc}} = L_d$. The longitudinal electric field experienced

by the particles can be estimated from Eq. 2.99 as starting at $E_z \approx -R/2$ and ending at $E_z \approx 0$ when the particles reach the bubble center. Therefore, the average gradient may be estimated as

$$\bar{E}_z = \frac{\sqrt{a_0}}{2} \quad (2.139)$$

Putting all this together and using $\Delta W = \Delta\gamma m_e c^2$ leads to an estimate for the change in the accelerating electron's Lorentz factor:

$$\Delta\gamma \approx \frac{2}{3} \left(\frac{\omega_0}{\omega_p} \right)^2 a_0 \quad (2.140)$$

All of Eqs. 2.134–2.140 are derived in Ref. [1].

The estimate for the energy gain in Eq. 2.140 assumes that the acceleration length L_{acc} is the same as the dephasing length L_d . This requires the additional assumption that the laser pulse is guided over a length scale $L_{\text{guide}} \geq L_d$. It was confirmed in Ref. [1] that this assumption holds at a density of $n_0 = 1.5 \times 10^{18} \text{ cm}^{-3}$ with an 800 nm laser with $a_0 = 4$ and $k_p w_0 = 4$, for which the simulated energy gain of 1.5 GeV closely matched the prediction from Eq. 2.140 of 1.6 GeV in about 0.75 cm. (Strictly speaking, the simulation was only run for half of L_d because of computational constraints, but most of the energy gain occurs in the first half of acceleration.) However, this was not verified at lower plasma densities because it was computationally infeasible and beyond the scope of the work. The authors proposed that self-guiding could be maintained as the density is lowered by proportionally increasing a_0 as

$$a_0 \propto \omega_p^{-1/5} \quad (2.141)$$

However, this scaling was not verified in Ref. [1]. It was found in Ref. [114] that the scaling law in Eq. 2.140 remains accurate for densities as low as $n_0 = 2.5 \times 10^{17} \text{ cm}^{-3}$ with a constant $a_0 = 4$, not necessarily following the scaling in Eq. 2.141. Investigation of the length scale for laser guiding at lower plasma densities, and hence the estimate for the energy gain in Eq. 2.140, is a natural direction for future work. Preliminary results are presented in Chapter 3.

The scalings in this section are sufficient to quantitatively establish the problem at the core of the work in Chapter 6. The goal of laser wakefield acceleration is to accelerate

charged particles at accelerating gradients of $\gtrsim 10 - 100$ GeV/m, more than two orders of magnitude higher than possible in conventional RF accelerators ($\lesssim 100$ MeV/m). In the blowout regime, an order-of-magnitude estimate for the accelerating gradient (Eq. 2.139) is, in terms of dimensional quantities, $eE_z/m_e c \omega_p \approx -1$. Hence, $E_z \propto \omega_p$. However, Eq. 2.140 shows that for a fixed laser wavelength (and hence ω_0), the energy gain in a single stage of LWFA in the blowout regime scales as $\Delta\gamma m_e c^2 \propto \omega_p^{-2}$. Thus, as the plasma density is increased, the accelerating gradient increases, but the energy gain decreases. This can be addressed in principle by chaining together an increasing number of LWFA stages to achieve a specified beam energy, but this introduces a number of significant technical challenges. The goal of using a flying focus laser driver in LWFA is to address this problem by extending the acceleration length far beyond the dephasing length for a fixed laser wavelength and plasma density. This allows one to increase the accelerating gradient without decreasing the acceleration length in a single plasma stage. This method is presented using an ASTRL flying focus in Chapter 6.

2.8 Beam loading

The concept of “beam loading” refers to tailoring of the witness beam profile in order to make $F_z \approx -eE_z$ constant within the region of the beam. Why is this necessary? Suppose the longitudinal electric field varies within the region of an accelerating witness beam in a laser wakefield accelerator. Then the beam will acquire an energy spread. This is detrimental because most applications of particle beams require a low energy spread (including collider physics). As shown in Eq. 2.99, the accelerating field in the blowout regime is approximately $F_z(\xi, r) \approx -\xi/2$. At first sight, the ξ -dependence may seem to be a problem. However, this profile does not include feedback from the witness beam itself on the wake, which modifies the fields.

Beam loading was first considered in the linear regime of LWFA in Ref. [115]. Following a

suggestion from S. van der Meer [116], Katsouleas et al. considered a beam with a trapezoidal current profile (i.e., a current profile linear in ξ). The parameters for the bunch can be found by requiring that the wakes from the laser and from the witness beam produce a constant accelerating field when linearly superposed. In the nonlinear regime, the details are more complicated because linear superposition does not apply. Tzoufras et al. provided a breakthrough in Refs. [99] and [117]: it was shown through analytic calculations and supporting simulations that an electron beam with a trapezoidal current profile can flatten the accelerating field in the nonlinear regime, just as in the linear regime. In this section, the main details of this calculation are presented in a slightly simpler way. Although the rest of the dissertation will not make use of the detailed calculations in this section, the calculations illustrate the concept of beam loading which is employed in Chapter 6.

The starting point of Tzoufras et al. is Lu's model for the electron sheath (Refs. [97] and [98]) under the simultaneous limits $\beta \rightarrow 0$ and $\beta r_b^2 \rightarrow \infty$. Under these limits, $\psi \approx \frac{1}{4}[r_b(\xi)^2 - r^2]$. From Eq. 2.80, the force experienced by an electron with $v_z \approx c$ and disjoint from any other particle beams is $F_z \approx \partial_\xi \psi$. Hence we obtain

$$F_z(\xi, r) \approx \frac{1}{2}r_b(\xi)r_b'(\xi) \quad (2.142)$$

This is the same force as in the delta sheath model.

We seek a functional form $r_b(\xi)$ such that $F_z(\xi, r)$ is a constant value F_{z0} for $\xi \in [\xi_-, \xi_+]$. Since $F_z(\xi, r) = \frac{1}{4}D_\xi(r_b(\xi)^2)$, we can immediately integrate in ξ to obtain the required shape of the bubble in the region of the bunch:

$$r_b(\xi)^2 = r_{b+}^2 + F_{z0}(\xi - \xi_+), \quad \xi \in [\xi_-, \xi_+] \quad (2.143)$$

where $r_{b+} \equiv r_b(\xi_+)$. This tells us the required shape of the bubble to produce the flat accelerating field: a paraboloid. The key insight of Ref. [99] is that the beam profile required to produce this $r_b(\xi)$ can be obtained using the differential equation relating the bubble radius and the beam current derived in Refs. [97] and [98]. In the limit of $\beta \rightarrow 0$ and $\beta r_b^2 \rightarrow \infty$, the

differential equation for the radius of the wake (Eq. 2.108) may be approximated as

$$\left(\frac{1}{4}\right) r_b^2 r_b'' + \left(\frac{1}{2}\right) r_b (r_b')^2 + \frac{r_b}{4} = \frac{\Lambda(\xi)}{r_b} \quad (2.144)$$

which is the same as Eq. 2.131 with the r_b'' term dropped. The current profile $\Lambda(\xi)$ is as yet undetermined. Substituting into Eq. 2.144 the equation for $r_b(\xi)$ in Eq. 2.143 and evaluating the derivatives, the following current profile is obtained:

$$\Lambda(\xi) = F_{z0}^2 + \frac{r_{b+}^2}{4} + F_{z0}(\xi - \xi_+), \quad \xi \in [\xi_-, \xi_+] \quad (2.145)$$

This shows that under these approximations, the beam current profile required to flatten the accelerating field is linear in ξ , or trapezoidal. The maximum length of the bunch $L_b = \xi_+ - \xi_-$ can be estimated from Eq. 2.143 by setting $r_b(\xi_-) = 0$. This yields

$$L_b = \frac{r_+^2}{4F_{z0}} \quad (2.146)$$

The current at the bunch head and tail $\Lambda_{\pm} \equiv \Lambda(\xi_{\pm})$ may be evaluated from Eq. 2.145 as

$$\Lambda_+ = F_{z0}^2 + \frac{r_{b+}^2}{4} \quad (2.147)$$

$$\Lambda_- = F_{z0}^2 \quad (2.148)$$

The total charge $Q = 2\pi \int \Lambda(\xi) d\xi$ may then be calculated as $Q = 2\pi \cdot \frac{1}{2}(\Lambda_- + \Lambda_+)L_b$, yielding

$$Q = \frac{\pi}{16} (8F_{z0}^2 + r_{b+}^2) \left(\frac{r_{b+}^2}{F_{z0}} \right) \quad (2.149)$$

Equations. 2.145–2.149 can be used to design a beam profile for the wake driven by a laser or particle beam. A simulation without a beam can initially be run, from which one can read off F_{z0} and r_{b+} at a chosen location for the beam head ξ_+ . The tail position ξ_- can be any value such that $\xi_+ - \xi_- < L_b$ from Eq. 2.146. The beam current at the head and tail are set according to Eq. 2.147 and Eq. 2.148.

The equations can also be re-cast in terms of $R_b \equiv \max_{\xi} r_b(\xi)$ instead of r_{b+} . This can be done by integrating the equation of the bubble trajectory in the region where the beam is

not present ($\xi > \xi_+$). We define the coordinate system so that the maximum radius occurs at $\xi = 0$. Hence $r_b(0) = R_b$ and $r'_b(0) = 0$. It is assumed that the head of the beam is behind the slice of maximum radius, i.e., $\xi_+ < 0$. For $\xi \in [\xi_+, 0]$, the equation for the bubble radius is $r_b r''_b + 2(r'_b)^2 + 1 = 0$. Since $F_z = \frac{1}{2} r_b r'_b$, we have $2\partial_\xi F_z = r_b r''_b + (r'_b)^2$. Eliminating the $r_b r''_b$ term from the differential equation using this identity, writing $\partial_\xi = r'_b \partial_{r_b}$, and eliminating the remaining r'_b factors using $r'_b = 2F_z/r_b$, we obtain the equation

$$\left(\frac{4F_z}{r_b}\right) \partial_{r_b} F_z + \frac{4F_z^2}{r_b^2} + 1 = 0 \quad (2.150)$$

Defining $u \equiv 2F_z/r_b$ and rewriting Eq. 2.150 in terms of u , we obtain

$$u \partial_{r_b} (r_b u) + u^2 + 1 = 0 \quad (2.151)$$

Using the product rule, this can be written as

$$\frac{1}{2} r_b \partial_{r_b} u^2 + u^2 + 1 = 0 \quad (2.152)$$

Defining $v \equiv u^2 + 1$, this reduces to

$$\frac{1}{4} r_b \partial_{r_b} v + v = 0 \quad (2.153)$$

which is separable and can be integrated to obtain

$$\frac{v(\xi)}{v(\xi_0)} = \left(\frac{r_b(\xi)}{r_b(\xi_0)}\right)^{-1/4} \quad (2.154)$$

for any ξ_0 . Now using $F_z(0) = 0$, $r_b(0) = R_b$, $F_z(\xi_+) = F_{z0}$, and $r_b(\xi_+) = r_{b+}$, we obtain

$$\frac{1}{2 \left(\frac{2F_{z0}}{r_{b+}}\right)^2 + 1} = \left(\frac{r_{b+}}{R_b}\right)^4 \quad (2.155)$$

This leads to a quadratic equation for r_{b+} which can be solved to obtain

$$r_{b+}^2 = -4F_{z0}^2 + \sqrt{16F_{z0}^4 + R_b^4} \quad (2.156)$$

This equation allows r_{b+} to be eliminated in favor of R_b in Eqs. 2.145–2.149. Doing so, we obtain

$$\Lambda(\xi) = \sqrt{F_{z0}^4 + \frac{R_b^4}{16}} + F_{z0}(\xi - \xi_+) \quad (2.157)$$

$$\Lambda_+ = \sqrt{F_{z0}^4 + \frac{R_b^4}{16}} \quad (2.158)$$

$$\Lambda_- = F_{z0}^2 \quad (2.159)$$

$$L_b = \frac{1}{F_z} \left[\sqrt{F_{z0}^4 + \frac{R_b^4}{16}} - F_{z0}^2 \right] \quad (2.160)$$

$$Q = \frac{\pi R_b^4}{16 F_{z0}} \quad (2.161)$$

The equation for Q in Eq. 2.161 had a typo in Ref. [99] which was corrected in Ref. [117]. Note that the total beam energy gain, which scales as $Q F_{z0}$, is independent of F_{z0} . However, the energy gain per particle, which scales as F_{z0} , is in competition with the total charge because $Q \propto F_{z0}^{-1}$.

In this flow of calculations, one begins with the desired form for $F_z(\xi, r)$, obtains $r_b(\xi)$ by solving Eq. 2.142, and calculates the required $\Lambda(\xi)$ using Eq. 2.144. This procedure can be repeated using more complex sheath models or non-constant F_z profiles. For example, in Ref. [118], a more accurate model for the sheath in the rear of the bubble was described. This model was demonstrated to improve the accuracy of the beam loading over the simpler model presented in this subsection. For the purposes of this dissertation, the simpler model presented here will suffice.

2.9 Witness beam evolution

In Sec. 2.8, it was assumed that the beam profile is invariant in time. In general, the witness beam will evolve. Here the evolution is considered. Before proceeding, we note that there is a vast literature on this subject. Key concepts include emittance preservation, beam matching, Courant-Snyder parameters, differences between projected and slice parameters,

and differences between the geometric and normalized emittance. The purpose of this section is not to be complete but to derive the beam matching condition and address the preservation of the normalized emittance. A more complete overview of this subject for PBA and definitions of these terms can be found in Ref. [119].

In general, there are two possible approaches: the motion of a single particle can be considered and then averaged over an ensemble of noninteracting particles, or the Vlasov equation can be considered for the beam. Both treatments lead to the same result and are thus equivalent. Here we follow the latter approach. We assume that the beam is chosen self-consistently as in Sec. 2.8 so that F_z is constant in space and time. We also assume the focusing force is linear as in Eq. 2.98. This leads to the system

$$\left\{ \begin{array}{l} \partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \mathbf{F}_{\perp} \cdot \nabla_{\mathbf{p}_{\perp}} f + F_z \partial_{p_z} f = 0 \\ F_z(t, z, r) = F_{z0} \\ \mathbf{F}_{\perp}(t, z, r) = -\frac{1}{2} m c^2 k_p^2 r \hat{\mathbf{r}} \end{array} \right. \quad (2.162)$$

This approach is taken to simplify the ideas as much as possible. One can generalize this approach by starting with the Vlasov equation under the quasistatic approximation and considering each slice of the beam separately. Under this form for $\mathbf{F}_{\perp}$, the motion in x and y is decoupled.

To understand the evolution of the beam, we consider the time evolution of the macroscopic quantities $\langle x^2 \rangle$, $\langle p_x^2 \rangle$, and $\langle x p_x \rangle$, defined by

$$\langle x^2 \rangle \equiv \int x^2 f(t, \mathbf{x}, \mathbf{p}) d\mathbf{x} d\mathbf{p} \quad (2.163)$$

$$\langle p_x^2 \rangle \equiv \int p_x^2 f(t, \mathbf{x}, \mathbf{p}) d\mathbf{x} d\mathbf{p} \quad (2.164)$$

$$\langle x p_x \rangle \equiv \int x p_x f(t, \mathbf{x}, \mathbf{p}) d\mathbf{x} d\mathbf{p} \quad (2.165)$$

These quantities are only functions of time and are referred to as projected quantities for the beam. Other averages are defined as $\langle g(\mathbf{x}, \mathbf{p}) \rangle = \int g(\mathbf{x}, \mathbf{p}) f(t, \mathbf{x}, \mathbf{p}) d\mathbf{x} d\mathbf{p}$. From the Vlasov

equation, we may compute $D_t\langle x^2 \rangle$ as

$$D_t\langle x^2 \rangle = \int x^2 D_t f \, d\mathbf{x} \, d\mathbf{p} \quad (2.166)$$

$$= \int x^2 \left[-v_x \partial_x f - v_y \partial_y f - v_z \partial_z f + \frac{1}{2} m c^2 k_p^2 x \partial_{p_x} f + \frac{1}{2} m c^2 k_p^2 y \partial_{p_y} f - F_{z0} \partial_{p_z} f \right] d\mathbf{x} \, d\mathbf{p} \quad (2.167)$$

$$= \int 2x v_x f \, d\mathbf{x} \, d\mathbf{p} \quad (2.168)$$

$$= 2\langle x v_x \rangle \quad (2.169)$$

where the third line follows from integrating each term by parts. Similarly, we obtain

$$D_t\langle p_x^2 \rangle = -m c^2 k_p^2 \langle x p_x \rangle \quad (2.170)$$

and

$$D_t\langle x p_x \rangle = \langle v_x p_x \rangle - \frac{1}{2} m c^2 k_p^2 \langle x^2 \rangle \quad (2.171)$$

We assume that the p_z dependence of f is strongly peaked in momentum space about $(p_x, p_y, p_z) = (0, 0, p_{z0}(t))$ with spread $(\Delta_{p_x}, \Delta_{p_y}, \Delta_{p_z})$. Under this approximation, the Lorentz factor is roughly constant throughout the beam with $\gamma(\mathbf{p}) \approx p_{z0}(t)/mc \equiv \gamma_b(t)$ and $p_{z0}(t) = p_{z0}(0) + \int_0^t F_{z0}(t') \, dt'$ as long as $\Delta_{p_z}^2 \ll p_{z0}^2$ and $\Delta_{p_x} \ll p_{z0}/2$. We assume these conditions hold. Under this approximation, we have

$$\langle x v_x \rangle = \frac{1}{m \gamma_b(t)} \langle x p_x \rangle \quad (2.172)$$

$$\langle v_x p_x \rangle = \frac{1}{m \gamma_b(t)} \langle p_x^2 \rangle \quad (2.173)$$

Using these identities, we obtain the following equations:

$$D_t\langle x^2 \rangle = -\frac{2}{m^2 c^2 k_p^2 \gamma_b(t)} D_t\langle p_x^2 \rangle \quad (2.174)$$

$$D_t^2\langle p_x^2 \rangle = \frac{1}{2} (m c^2 k_p^2)^2 \langle x^2 \rangle - \frac{c^2 k_p^2}{\gamma_b(t)} \langle p_x^2 \rangle \quad (2.175)$$

These coupled equations lead to oscillatory motion for both $\langle x^2 \rangle$ and $\langle p_x^2 \rangle$ at a characteristic frequency

$$\omega_\beta \equiv \frac{\omega_p}{\sqrt{2\gamma}} \quad (2.176)$$

which is referred to as the betatron frequency.

Is it possible to find solutions for which $D_t^2 \langle p_x^2 \rangle \approx 0$ so that $\langle x^2 \rangle$ and $\langle p_x^2 \rangle$ do not oscillate? We can seek them. Setting the right-hand side terms equal in Eq. 2.175, we obtain the relationship

$$\langle p_x^2 \rangle = \frac{\gamma_b(t) m^2 c^2 k_p^2}{2} \langle x^2 \rangle \quad (2.177)$$

This equation must be self-consistent with Eq. 2.174. Differentiating and substituting the resulting equation for $D_t \langle p_x^2 \rangle$ into Eq. 2.174 yields

$$-\frac{1}{2} \gamma_b(t) m^2 c^2 k_p^2 D_t \langle x^2 \rangle = \frac{1}{2} \gamma_b(t) m^2 c^2 k_p^2 D_t \langle x^2 \rangle + \frac{1}{2} \dot{\gamma}_b(t) m^2 c^2 k_p^2 \langle x^2 \rangle \quad (2.178)$$

Simplifying yields

$$D_t \langle x^2 \rangle = -\frac{1}{2} \left(\frac{\dot{\gamma}_b(t)}{\gamma_b(t)} \right) \langle x^2 \rangle \quad (2.179)$$

Solving by separation of variables and using Eq. 2.177 to find $\langle p_x^2 \rangle$, we obtain

$$\langle x^2 \rangle = \left(\frac{\gamma_b(t)}{\gamma_b(0)} \right)^{-1/2} \langle x^2 \rangle_0 \quad (2.180)$$

$$\langle p_x^2 \rangle = \left(\frac{\gamma_b(t)}{\gamma_b(0)} \right)^{1/2} \langle p_x^2 \rangle_0 \quad (2.181)$$

where

$$\langle p_x^2 \rangle_0 = \left[\frac{1}{2} \gamma_b(0) m^2 c^2 k_p^2 \right] \langle x^2 \rangle_0 \quad (2.182)$$

These solutions exactly satisfy Eq. 2.174. The time dependence of $\langle p_x^2 \rangle$ implies that Eq. 2.175 is not exactly satisfied. Eq. 2.175 implies that the first-order correction to the solutions in Eq. 2.180–2.181 will exhibit oscillatory motion at frequency $\omega^{(1)} = \frac{1}{2} \sqrt{\dot{\gamma}_b / \gamma_b}$. The condition for the solutions in Eq. 2.181–2.181 to be accurate is that $\omega^{(1)} \ll \omega^{(0)} = \omega_\beta$. Assuming F_{z0} is constant, this leads to the condition

$$\frac{F_{z0}}{mc^2 k_p} \ll \sqrt{2\gamma_b} \quad (2.183)$$

The left-hand side is typically of order unity in the nonlinear blowout regime, while γ_b is typically $\gg 1$. Thus, this condition is typically satisfied.

To summarize more practically, Eq. 2.182 is a matching condition. It shows how to design the witness beam: given the beam spot size and Lorentz factor, the transverse momentum is set to the appropriate value (or vice versa). Using the notation $\sigma_{p_x} \equiv \sqrt{\langle p_x^2 \rangle}$ and $\sigma_x \equiv \sqrt{\langle x^2 \rangle}$, this matching condition can be expressed in dimensionless units in the more familiar form

$$\sigma_{p_x} = \gamma_b k_\beta \sigma_x \quad (2.184)$$

where $k_\beta \equiv k_p / \sqrt{2\gamma_b}$ is the betatron wavenumber. From Eq. 2.177, this matching condition will be satisfied for all time if it is satisfied initially. Under this condition, the spot size will not oscillate at a frequency ω_β , but will adiabatically contract as the transverse momentum spread adiabatically increases according to Eqs. 2.180 and 2.181. The matching condition can also be expressed as

$$\sigma_{x'} = k_\beta \sigma_x \quad (2.185)$$

where $x' \equiv v_x/c$ is the notation often used in the accelerator physics community.

We conclude this section by commenting on the normalized emittance ϵ_N , defined as

$$\epsilon_N \equiv \langle x^2 \rangle \langle p_x^2 \rangle - \langle xp_x \rangle^2 \quad (2.186)$$

The normalized emittance is conserved for any beam solution of the system in Eq. 2.162 with a linear focusing force and uniform accelerating field. This can be shown by direct calculation:

$$D_t \epsilon_N^2 = D_t [\langle x^2 \rangle \langle p_x^2 \rangle - \langle xp_x \rangle^2] \quad (2.187)$$

$$= \langle x^2 \rangle D_t \langle p_x^2 \rangle + \langle p_x^2 \rangle D_t \langle x^2 \rangle - 2 \langle xp_x \rangle D_t \langle xp_x \rangle \quad (2.188)$$

$$= \langle x^2 \rangle [-mc^2 k_p^2 \langle xp_x \rangle] + \langle p_x^2 \rangle \left[\frac{2}{m\gamma_b(t)} \langle xp_x \rangle \right] \\ - 2 \langle xp_x \rangle \left[\frac{1}{m\gamma_b(t)} \langle p_x^2 \rangle - \frac{1}{2} mc^2 k_p^2 \langle x^2 \rangle \right] \quad (2.189)$$

$$= 0 \quad (2.190)$$

In particular, ϵ_N is conserved even if the beam is not matched using the matching condition in Eq. 2.182. However, if the beam is not matched, then the resulting oscillations of the beam spot size are problematic. If the beam's spot size is oscillating and there is any energy spread in the beam, then particles with different energies oscillate at different betatron frequencies. This results in phase mixing and leads to growth of the normalized emittance [119]. In addition, if the beam's spot size is oscillating and the beam exits the accelerator at any time other than the moment when $D_t\langle x^2 \rangle \approx 0$, then the beam will diverge upon exiting and will not be able to be transferred to the target or next stage without corrective beam optics.

2.10 Simulation of laser wakefield acceleration in the blowout regime

John Dawson wrote: “There are no God-given correct methods of simulation, there are only approximations with varying advantages and disadvantages” [75]. This is certainly true for simulation of LWFA in the blowout regime. The main simulation methods are introduced at a conceptual level in Chapter 1. This section goes into more detail on their subtleties, advantages, disadvantages, and history.

2.10.1 The particle-in-cell algorithm

The fully electromagnetic particle-in-cell algorithm is the highest-fidelity method for computational modeling of plasmas. It is also the most computationally expensive. The key idea is to approximate the system of Maxwell's equations coupled to the evolution of point particles (Eqs. 2.2–2.11) using finite-size particles and a discretization for the evolution of Maxwell's equations ($\mathbf{E}$, $\mathbf{B}$, ρ , and $\mathbf{J}$ are represented on grids and are advanced discretely in time). In this context, the finite-size particles are referred to as macroparticles. If the effects of point-particle interactions on the length scale below that of the particle shape function are negligible compared to the physics determined by interactions outside this length scale, then finite-size particles provide a reasonable approximation. In particular, for a collisionless

plasma for which the physics is completely dominated by the collective plasma response, PIC is typically the most computationally effective approach. An alternative is a finite-difference discretization of the Vlasov-Maxwell system, but this requires a 6D grid for the distribution function and can become computationally infeasible. The most common implementation of the PIC algorithm is summarized in Fig. 1.3 and the surrounding text. Here more technical details are supplied, especially those of relevance to this dissertation.

Numerical dispersion, which refers to alteration of dispersion properties of waves due to discretization, is an important subtlety of the PIC algorithm. For example, the simplest discretization for the fields is a second-order explicit finite-difference solver on the so-called Yee mesh [120] with uniform grid spacing $\Delta \mathbf{x}$ and a timestep Δt . There have been efforts to develop mesh refinement but these are not discussed here. For a uniform grid spacing, it can be shown that the physical dispersion relation for plane waves propagating in the $\hat{\mathbf{z}}$ direction in vacuum

$$\omega^2 = k^2 \quad (2.191)$$

becomes with the Yee solver

$$[\omega]_t^2 = [k]_{E_z} [k]_{B_z} \quad (2.192)$$

where

$$[\omega]_t = \omega \operatorname{sinc} \left(\frac{\omega \Delta t}{2} \right) \quad (2.193)$$

and

$$[k]_{E_z} = [k]_{B_z} = k \operatorname{sinc} \left(\frac{k \Delta z}{2} \right) \quad (2.194)$$

with $\operatorname{sinc}(x) \equiv \sin(x)/x$ [4]. By Taylor-expanding the sinc function, the solution to the numerical dispersion relation can be estimated as

$$\omega_{\text{Yee}} \approx k + \frac{k^3}{24} [-\Delta z^2 + \Delta t^2] \quad (2.195)$$

This leads to an estimate for the group velocity $v_g = \partial_k \omega$ for the Yee solver of

$$v_{g,\text{Yee}} \approx 1 + \frac{k^2}{8} [-\Delta z^2 + \Delta t^2] \quad (2.196)$$

The first term is the expected vacuum group velocity, $v_g = 1$. The second is an alteration due to numerical dispersion. Thus, for physical problems sensitive to the group velocity of the laser, such as LWFA, effects of numerical dispersion can be problematic. In particular, the Yee solver requires $\Delta t < \Delta z$ as a necessary condition for stability, so the second term is always negative. Thus, the Yee solver results in a forward-propagating wave packet traveling subluminally in vacuum. The group velocity difference is approximately the same when accounting for linear plasma dispersion, though in this case the parameters can be chosen so that the numerical difference is smaller than the physical difference due to linear dispersion. The same ideas also apply when including nonlinear interaction.

The effect of numerical dispersion can typically be controlled by decreasing Δz and Δt , e.g., in Eq. 2.196. Another option is to customize the field solver coefficients. This was the focus of much research carried out by the UCLA Plasma Simulation Group during the 2010s, mostly for the sake of addressing the numerical Cherenkov instability. Details about this latter form of numerical dispersion, which occurs prominently in the context of Lorentz-boosted simulations of LWFA, are discussed in Appendix A. Most noteworthy for this dissertation was the development of the “dual solver”, introduced in Ref. [61]. This finite-difference field solver is designed to reduce numerical dispersion for light propagating along the $\hat{\mathbf{z}}$ direction. This field solver enabled the PIC simulations carried out in Chapters 5 and 6, which would not have been possible with the Yee solver.

PIC simulations of LWFA also can benefit from cylindrical mode expansion [77, 78]. In this description, which is appropriate for systems with a high degree of azimuthal symmetry such as LWFA, grid quantities $(\mathbf{E}, \mathbf{B}, \rho, \mathbf{J})$ are approximated by a truncated Fourier series with respect to ϕ . This allows the 3D structure of the fields to be captured using 2D grids. Macroparticles are still pushed in 3D, but there is a substantial speedup because the number of macroparticles required is significantly lower than in Cartesian 3D geometry (and also because of the speedup of the field solver). This method was used extensively throughout my doctoral research.

Lorentz-boosted simulations [83, 84] are another method that can lead to substantial speedup in PIC simulation of LWFA. This simulation method is discussed in detail in Appendix A. This method is not used in simulations of this dissertation. The PIC algorithm has been implemented in many computer codes over the decades. At the time of writing, the most commonly used PIC codes for modeling LWFA include OSIRIS [2], EPOCH [121], FBPIC [122], WarpX [123], and SMILEI [124].

2.10.2 The ponderomotive guiding center PIC algorithm

The ponderomotive guiding center approximation can be built into the PIC algorithm. The algorithm is the same as the original PIC algorithm with a modified particle pusher for the macroparticles to account for the ponderomotive force (Eq. 2.27) coupled to an additional grid which discretizes $\hat{A}$ and evolves it using Eq. 2.43. The original PIC simulation quantities $\mathbf{E}, \mathbf{B}, \rho, \mathbf{J}, \mathbf{x}, \mathbf{p}$ are then interpreted as the slowly-varying quantities $\bar{\mathbf{E}}, \bar{\mathbf{B}}, \bar{\rho}, \bar{\mathbf{J}}, \bar{\mathbf{x}}, \bar{\mathbf{p}}$. No additional correction is required for deposition of $\mathbf{J}$ and ρ , as in quasistatic PIC codes.

Using this approximation, the laser wavelength and frequency no longer need to be resolved. For typical LWFA simulations, this leads to a speedup of $\approx \left(\frac{k_0}{k_p}\right)^2$ because the longitudinal resolution and temporal resolution both change from $\propto k_0^{-1}$ to $\propto k_p^{-1}$. Hence, it can be significantly faster than the full PIC algorithm. Unlike with quasistatic PIC codes, it retains the ability to model the physics of injection and is not prone to inaccuracy or instability near the back of the wake. The main drawbacks are the possible breakdown of the PGC approximation if the laser becomes significantly redshifted or spatially distorted and sensitivity of the laser solver to numerical instability.

The PGC PIC algorithm, which was historically preceded by the quasistatic PIC algorithm, was first implemented in the code turboWAVE, introduced in Ref. [107]. Later it was implemented in VORPAL [125], OSIRIS [126], ALaDyn [127], and SMILEI [128]. It was also recently shown how this method can be applied in a Lorentz-boosted frame [129]. This method is not employed in this dissertation.

2.10.3 Quasistatic PIC algorithms

Quasistatic PIC algorithms are the third main method for simulation of LWFA in the blowout regime. In these codes, the full quasistatic approximation is applied. The algorithm integrates particle trajectories throughout the entire simulation box using Eq. 2.49 and Eq. 2.50, assuming that all laser and particle beams do not evolve during this time. However, this is complicated by the fact that the slowly-varying fields $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$ and particle positions and momenta are not time-centered as in the full PIC algorithm, so a leapfrog algorithm cannot be employed. The fields are determined as the particles are advanced through an iterative process. For this reason, along with other subtleties with deposition of the particles and parallelization, the quasistatic PIC algorithm is more complex than the full PIC and PGC PIC algorithms. In addition, quasistatic PIC algorithms can break down near the back of the wake, where factors of $1/(1 - v_z)$ appearing in the plasma source terms (Eqs. 2.68 and 2.69) can blow up. In particular, quasistatic PIC codes cannot easily self-consistently model the physics of injection of particles from the plasma into the wake because a different pusher is ordinarily used for the beam particles and plasma particles.

When these limitations do not affect the physics of interest, the quasistatic PIC code can be 10^4 to 10^6 times faster than full PIC with robust accuracy. For LWFA, the smallest longitudinal spatial scale changes from the laser wavelength to the plasma skin depth, as with the PGC PIC algorithm. However, the temporal timescale becomes not ω_p^{-1} , but the length scale for evolution of the laser envelope $\Delta_s \gg 0_p^{-1}$. The speedup relative to full PIC thus scales for LWFA as $\left(\frac{k_0}{k_p}\right)^2 k_p \Delta_s$.

The first quasistatic PIC code was WAKE, introduced in Ref. [26]. Mora and Antonsen recognized the need to weight particle charge and current depositions by a factor of $1/(1 - v_z)$ for plasma particles (see Eqs. 2.68 and 2.69). Later came QuickPIC [79, 80], which was the first fully 3D quasistatic PIC code. HiPACE++ was developed subsequently [82]. QPAD [39, 81] was later introduced, the acronym meaning “QuickPIC with Azimuthal

Decomposition”. This code integrates the azimuthal mode decomposition discussed above with the QuickPIC algorithm. QPAD enabled the simulations in Chapter 6, which would have been computationally infeasible with full PIC.

2.11 Summary

The Vlasov-Maxwell system provides a mathematical starting point for theoretical and computational modeling of LWFA in the blowout regime. However, it is analytically intractable. More tractable equations can be obtained by applying the ponderomotive guiding center approximation and quasistatic approximations. The resulting equations can be used for phenomenological description of various aspects of LWFA in the blowout regime, including the structure of the plasma wake, laser evolution, beam loading, and witness beam evolution. Though analytically intractable, the full Vlasov-Maxwell system can be simulated using the PIC algorithm. This provides the highest-fidelity simulations of LWFA. The equations resulting from application of the PGC approximation and quasistatic approximation can also be simulated using the PGC PIC algorithm and quasistatic PIC algorithm. These methods can be significantly faster than the full PIC method.

CHAPTER 3

On the Scaling Laws for the Acceleration Length and Energy Gain in the Blowout Regime of LWFA in Uniform Plasma

The present chapter is motivated by recent developments at ELI Beamlines [130], where experiments of LWFA with laser powers in the range of 3–10 PW at 150 fs (energy of 0.5–1.5 kJ) and at plasma densities in the range of $10^{16} - 10^{17} \text{ cm}^{-3}$ may soon be possible. The theory of Lu et al. [1], presented in Chapter 2, provides scaling laws for the energy gain and acceleration length. Based on these, one would expect that energy gains of 30–80 GeV in a single several-meter stage of uniform plasma are possible in the near term with the above laser and plasma parameters.

In Lu et al., it was recognized that assumptions made to obtain the scaling laws would not continue to apply as the density was lowered because self-guiding of the laser pulse would not continue to scale. It was suggested that this issue could be compensated by scaling the laser amplitude as $a_0 \propto (n/n_c)^{-1/5}$. However, simulation tools for determining at what plasma density the scaling laws would begin to overestimate the energy gain did not exist until recently.

In this chapter, we use QPAD simulations to quantitatively investigate the scaling laws of Lu et al. as ω_0/ω_p is increased for fixed normalized laser parameters. As expected, these simulations suggest that the scaling laws overpredict the energy gain and acceleration length beyond a sufficiently large ω_0/ω_p . QPAD simulations show that this transition occurs for

$\omega_0/\omega_p \gtrsim 75$ for fixed dimensionless values of $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 4$. Beyond this threshold, the laser self-guiding length L_{guide} (defined for simplicity as the z for which $\max e|\hat{E}|/mc\omega_0$ decreases by a factor of 2 from its initial value), satisfies the scaling $L_{\text{guide}} \approx 25z_R$ and is less than L_d (the laser etching or dephasing length). For these dimensionless parameters, it is found that the acceleration length and energy gain approximately satisfy $L_{\text{acc}} \approx 25z_R$ and $\Delta\gamma \approx 25k_p z_R$ for $\omega_0/\omega_p \gtrsim 250$, where the threshold is different because of subtle details in the evolution of the wake and laser beyond depletion.

These results have consequences both for near-term and long term experiments. In the near-term, the simulations suggest that energy gains of 20–30 GeV may be possible at ELI beamlines. In the long term, they place constraints on the regimes within which LWFA can be efficiently operated, a concern that constrains design choices for an LWFA-based linear collider. The chapter does not include simulations of LWFA driven by flying focus pulses. However, the results are relevant in this context, because the motivation for using flying focus pulses is based on an accurate understanding of the energy gain and acceleration length for LWFA driven by a single laser pulse. This is discussed in the conclusion.

3.1 Introduction

In laser wakefield acceleration (LWFA), the acceleration length and the total energy gain for an accelerating beam are critical system parameters. Models for these parameters as a function of basic quantities such as the laser wavelength, laser power, and plasma density are essential for designing the system. Nearly two decades ago, Lu et al. in Ref. [1] provided estimates for these quantities as a function of the laser and plasma parameters in the self-guided blowout regime. As discussed in Chapter 2, Ref. [1] provided the estimates $\Delta\gamma \approx \frac{2}{3} \left(\frac{\omega_0}{\omega_p} \right)^2 a_0$ and $L_{\text{acc}} = L_d \approx \frac{4}{3} \left(\frac{\omega_0}{\omega_p} \right)^2 \sqrt{a_0}$ where the laser satisfies the matching condition $k_p w_0 = 2\sqrt{a_0}$. It was noted that these estimates also apply when the laser is externally guided by a plasma channel.

A key assumption of the model of Lu et al. is that the length scale for self-guiding of the laser L_{guide} occurs over a distance at least as large as the dephasing length L_d , i.e., $L_{\text{guide}} \geq L_d$. This assumption is explicitly acknowledged in Ref. [1], where it was found to be valid at a frequency ratio $\omega_0/\omega_p \approx 34$ for $a_0 = 4$ using 3D PIC simulations. However, the validity of this assumption for higher ratios of ω_0/ω_p (i.e., lower density for a fixed laser wavelength) was not demonstrated and was left as an open research question. Lu et al. pointed out that they did not expect the scalings to hold for arbitrary plasma densities and argued that it could be addressed by proportionately increasing the laser amplitude as $a_0 \propto (n_0/n_c)^{-1/5}$.

It is important and timely to examine these scaling laws for the energy gain and acceleration length because they inform experimental decisions and exploration of new concepts in LWFA. For example, modern high-power laser facilities such as ELI beamlines are on the verge of exploring previously untested laser-plasma parameters (density in the range of $n_0 \approx 10^{16} - 10^{17} \text{ cm}^{-3}$ for $\lambda_0 \approx 1000 \text{ nm}$ so that $100 \lesssim \omega_0/\omega_p \lesssim 300$). As another example, flying focus concepts have been recently proposed to outperform single-pulse concepts [37, 38]; fair comparison requires an accurate understanding of the physics of LWFA with a single pulse.

In spite of the importance, there have been relatively few studies systematically investigating the accuracy of the scaling laws. A parameter scan in Ref. [131] using quasistatic PIC simulations demonstrated that the Lu et al. scaling laws continue to work in the weakly nonlinear regime ($a_0 \approx 2$) up to $\omega_0/\omega_p = 2000$ when a plasma channel is used to guide the laser. Another parameter scan in Ref. [114] showed that the scaling laws continue to work in uniform plasma in the blowout regime with fixed dimensionless values a_0 , $k_p w_0$, and $\omega_p \tau_0$ ($a_0 = 4.44$ in this case) up to $\omega_0/\omega_p \approx 80$ using fully electromagnetic PIC simulations with azimuthal mode decomposition (quasi-3D geometry) [78]. Isolated cases in Ref. [85] using 3D PIC in a Lorentz-boosted frame demonstrated the accuracy of the scaling laws up to $\omega_0/\omega_p \approx 80$ with $a_0 = 5.8$ in uniform plasma but suggested a slight overestimation at $\omega_0/\omega_p \approx 280$ with $a_0 = 2$ in a plasma channel (53 GeV predicted, 40 GeV simulated) with

external guiding. At the time of writing, parameter scans in the blowout regime in uniform plasma with densities extending down to those relevant for modern multi-petawatt laser systems appear not to be published.

In this chapter, this gap in the literature is addressed. The scaling laws for LWFA in uniform plasma in the blowout regime with fixed dimensionless values of $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 4$ are investigated using quasistatic PIC simulations. We find that for $\omega_0/\omega_p \lesssim 50$ the self-guiding length satisfies $L_{\text{guide}} \approx 1.5L_d$. For $\omega_0/\omega_p \gtrsim 75$, the simulations suggest that $L_{\text{guide}} \approx 25z_R$. These scalings are reflected in the energy gain and acceleration length: for $\omega_0/\omega_p \lesssim 150$, the Lu et al. scalings still hold for these quantities. However, for $\omega_0/\omega_p \gtrsim 250$, the scalings $\Delta\gamma \approx 25k_p z_R$ and $L_{\text{acc}} \approx 25z_R$ better describe the data. The Lu et al. scalings hold up to $\omega_0/\omega_p \approx 250$ even though $L_{\text{guide}} \approx L_d$ at $\omega_0/\omega_p \approx 75$ because of the additional subtle physics of the evolution of the laser and wake beyond guiding. Between the lower and upper thresholds, there is a smooth transition from one scaling to the other for each quantity. The ramifications of these findings are discussed and future directions are provided in the conclusion.

3.2 Results

The results are naturally divided into subsections on (1) the simulation setup, (2) the evolution of the laser and wake, and (3) particle acceleration.

3.2.1 Simulation Setup

The tool we use to study these scaling laws is QPAD [39, 81]. QPAD is a quasistatic PIC code which features azimuthal mode decomposition as in Ref. [78]. It solves for the evolution of the electromagnetic fields and particle trajectories for the wake using the quasistatic approximation. The field quantities and plasma source terms are decomposed into azimuthal Fourier series; the different modes are represented on grids in r and ξ . The quasistatic

approximation and QPAD are discussed in more detail in Chapter 2.

In order to test the scaling laws, a simple parameter scan is carried out. For each simulation, the normalized laser parameters are $a_0 = 4$, $k_p w_0 = 4$, $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 4$. The polynomial profile from Ref. [1] is used for the longitudinal envelope. These parameters were chosen for simplicity and are not optimized. A range of densities from $n_0 = 1 \times 10^{19} \text{ cm}^{-3}$ to $n_0 = 1 \times 10^{15} \text{ cm}^{-3}$ for a physical wavelength of $\lambda_0 = 1000 \text{ nm}$ are tested by using a different frequency ratio ω_0/ω_p for each simulation. For each case, the laser is focused to the entrance of the plasma.

In each simulation, the bounds of the computational domain are $k_p r_{\text{max}} = 204.8$ and $k_p(\xi_{\text{max}} - \xi_{\text{min}}) = 17$. The relatively large transverse box size ($r_{\text{max}} \approx 50w_0$) was chosen to ensure there were no spurious numerical effects at the transverse boundary. The value could probably be decreased. The longitudinal box size is large enough to capture one bubble of the plasma wake. In all cases, a timestep of $k_p \Delta_s \approx 10$ is used. The exact value was shifted slightly in order to get the simulation output diagnostics to correspond to the same number of multiples of L_d or z_R . In all cases, $k_p \Delta_s \in [6, 11]$. This choice was informed by Ref. [8], where it was found that $k_p \Delta_s = 10$ was required for agreement between QPAD and OSIRIS using a flying focus driver with $k_0/k_p = 25$. Also, in each case a longitudinal resolution $k_p \Delta_\xi = 0.01$ was used. This choice was also informed by the longitudinal resolution required in Ref. [8]. The fine longitudinal resolution captures the redshifting of the laser. Sec. 3.3 includes a discussion on these choices. Only one azimuthal mode is used for all simulations.

In each simulation, a matched low-density electron beam is initialized along the axis with $n_b/n_0 = 0.01$ and $k_p \sigma_r = 0.2$. This gives a normalized line-charge density of $\Lambda = 2 \times 10^{-4}$, so feedback from the beam on the wake is negligible. The initial Lorentz factor is $\gamma_{b0} = 2000$. The initial transverse momentum spread of $\sigma_{p_x}/mc = 6.32$ is chosen to satisfy the matching condition of Eq. 2.182: $\sigma_{p_x}^2 = \frac{1}{2} \gamma_b m^2 c^2 k_p^2 \sigma_x^2$.

Table 3.1 summarizes the relevant parameters for the simulations that follow. In each case, a reference wavelenth of $\lambda_0 = 1000 \text{ nm}$ is used. This is to facilitate comparison with parameters

n_0 [cm ⁻³]	λ_0 [nm]	a_0	k_0/k_p	w_0 [μm]	τ_{rise} [fs]	P_L [PW]	E_L [J]
1×10^{19}	1000	4	11	6.7	22	0.016	0.27
5×10^{18}	1000	4	15	9.5	32	0.031	0.77
2×10^{18}	1000	4	24	15	50	0.078	3.0
1×10^{18}	1000	4	33	21	71	0.15	8.7
5×10^{17}	1000	4	47	30	100	0.31	24
2×10^{17}	1000	4	75	48	160	0.78	97
1×10^{17}	1000	4	110	67	220	1.6	270
5×10^{16}	1000	4	150	95	320	3.1	770
2×10^{16}	1000	4	240	150	500	7.8	3000
1×10^{16}	1000	4	330	210	710	16	8600
5×10^{15}	1000	4	470	300	1000	31	24,000
2×10^{15}	1000	4	750	480	1600	78	96,000
1×10^{15}	1000	4	1100	670	2200	160	2.7×10^5

Table 3.1: Summary of the relevant physical and dimensionless quantities associated with the simulations. For each simulation, $k_p w_0 = 4$ and $k_p \tau_{\text{rise}} = k_p \tau_{\text{fall}} = 4$. Values are rounded to 2 significant digits for legibility. The simulations used more precision for the ratio ω_0/ω_p , which was calculated from the exact values of n_0 in the first column. Any row within the table can be rescaled to a different laser wavelength and plasma density with the same ω_0/ω_p .

at the ELI laser facility. These reference densities will be used to identify the simulations in much of what follows, occasionally with reference to the corresponding dimensionless parameter ω_0/ω_p . Each row in Table 3.1 can be rescaled to other laser wavelengths (e.g., 800 nm) and plasma densities for a fixed ω_0/ω_p through the scalings $w_0 \propto k_p^{-1}$, $\tau_{\text{rise}} \propto k_p^{-1}$, $P_L \propto \left(\frac{k_0}{k_p}\right)^2$ (so P_L is invariant), and $E_L \propto \left(\frac{k_0}{k_p}\right)^2 k_p^{-1}$.

At a density $n_0 \leq 10^{16}$ cm⁻³, the characteristic accelerating gradient is $mc\omega_p \approx 10$ GeV/m. For densities below this, the accelerating gradient becomes less than two orders of magnitude larger than maximum gradient in a conventional accelerator (~ 100 MeV/m). Thus, there is a diminishing return to the benefit of using plasma-based acceleration at these densities. In addition, the laser and plasma parameters of the three lowest-density cases are far beyond current experimental capabilities. Nonetheless, we will see below that the cases in the density

range of $10^{15} - 10^{16} \text{ cm}^{-3}$ add clarity to the physics.

3.2.2 Evolution of the laser and wake

We begin by considering the evolution of the laser and wake. Figure 3.1 shows snapshots of the evolution of the laser and wake at $z = 0.7L_d$ for each density. The diagnostic frequency and Δ_s for each simulation were chosen so that all the simulations output data at the same values of z/L_d ; this simplified the analysis. If the evolution of the laser and wake depended only on the dimensionless parameter z/L_d , then all the images would appear the same in these normalized units. At this propagation distance, the wake and laser structures appear similar, though slightly different, for the five cases with $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$. These cases suggest that in this density range, the evolution of the laser and wake depend weakly on k_0/k_p but still evolve on the length scale of L_d . However, for the cases with $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$, there is a clear gradual weakening of the laser and the wake. This indicates that for these cases, $L_{\text{guide}} < L_d$ and the self-guiding length follows a weaker scaling with density than $L_{\text{guide}} \propto L_d \propto n_0^{-1}$.

Figure 3.1 suggests a different scaling for the laser guiding length at lower densities. Figure 3.2 shows the laser and wake evolution for the same cases as in Fig. 3.1 but with each snapshot shown at a fixed $z = 20z_R$. A separate set of simulations with the same laser and plasma parameters but different diagnostic frequency and slightly different resolution in Δ_s were run so that the simulations could easily be compared for fixed ratios of z/z_R . The top three cases with $n_0 \geq 1 \times 10^{18} \text{ cm}^{-3}$, for which $z > L_d$, exhibit complete pump depletion. On the other hand, the structure of the three cases with $n_0 \leq 1 \times 10^{17} \text{ cm}^{-3}$ is almost indistinguishable. This suggests that at these lower densities, the $L_{\text{guide}} \propto z_R$.

To further illustrate this point, the on-axis E_z is plotted in Fig. 3.3 for all the cases in Figs. 3.1 and 3.2 as well as the four additional cases with densities down to $n_0 = 1 \times 10^{15} \text{ cm}^{-3}$. Figure 3.3(a) shows results for these cases at a fixed $z = 0.7L_d$. It is clear that the amplitude of the wake at this fixed z/L_d decreases with density. The alignment of E_z and the shared

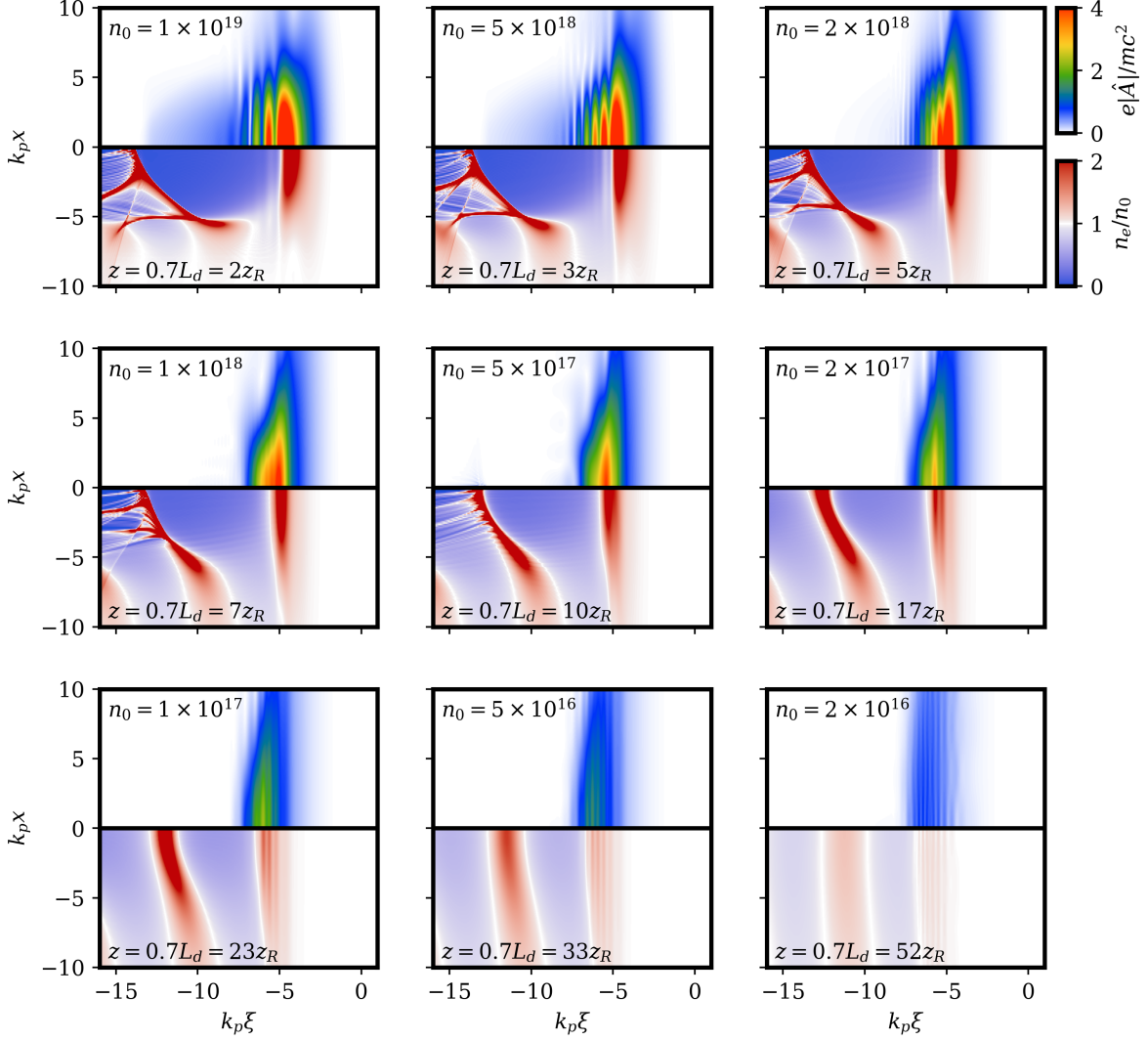


Figure 3.1: Snapshots of the laser envelope and electron charge density at densities ranging from $n_0 = 1 \times 10^{19} \text{ cm}^{-3}$ down to $n_0 = 2 \times 10^{16} \text{ cm}^{-3}$ at a fixed $z = 0.7L_d$, where L_d depends on the density for each case. If the evolution of the laser and wake depended on the dimensionless parameter z/L_d , then all the images would appear the same. This is approximately true for the 5 cases with $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$. For the cases with $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$, it is evident that $L_{\text{guide}} < L_d$ and the self-guiding length follows a weaker scaling with density than $L_{\text{guide}} \propto L_d \propto n_0^{-1}$.

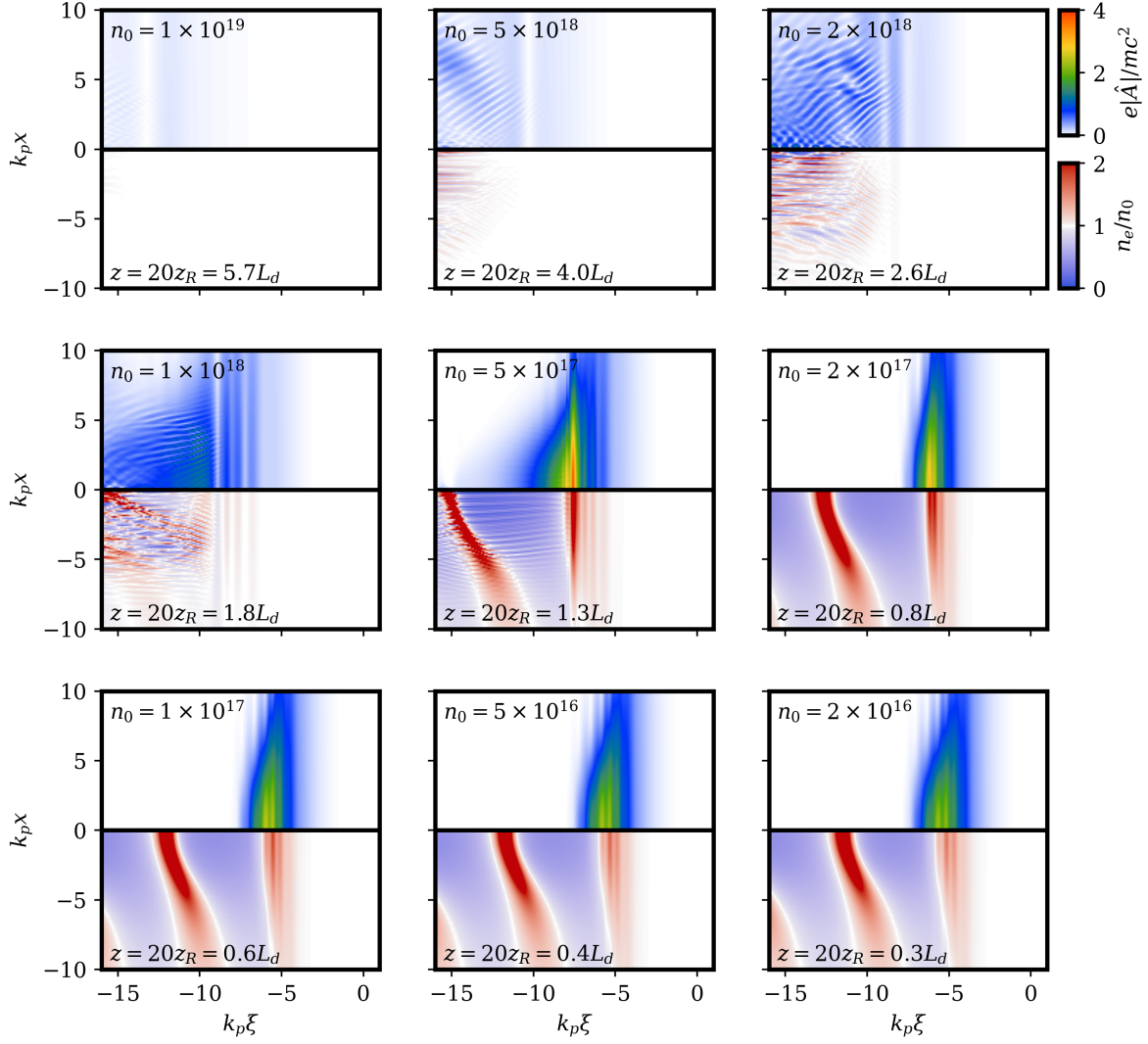


Figure 3.2: Snapshots of the laser envelope and electron charge density as in Fig. 3.1, but at fixed $z = 20z_R$. The cases with $n_0 \geq 1 \times 10^{18} \text{ cm}^{-3}$ are fully depleted at this stage. On the other hand, the three lowest-density cases are nearly indistinguishable in structure. This suggests that $L_{\text{guide}} \propto z_R$ at these lower densities.

value of $k_p \xi$ where $E_z = 0$ in all simulations indicate that the wake velocity satisfies the scaling $\beta_w - 1 \propto \left(\frac{\omega_p}{\omega_0}\right)^2$ at all densities, consistent with the model of Lu et al. On the other hand, Fig. 3.3(b) shows these cases at a fixed $z = 20z_R$. Here it is clear that for $n_0 \leq 5 \times 10^{17} \text{ cm}^{-3}$, the wake amplitude is roughly the same for the lower densities. This is consistent with the

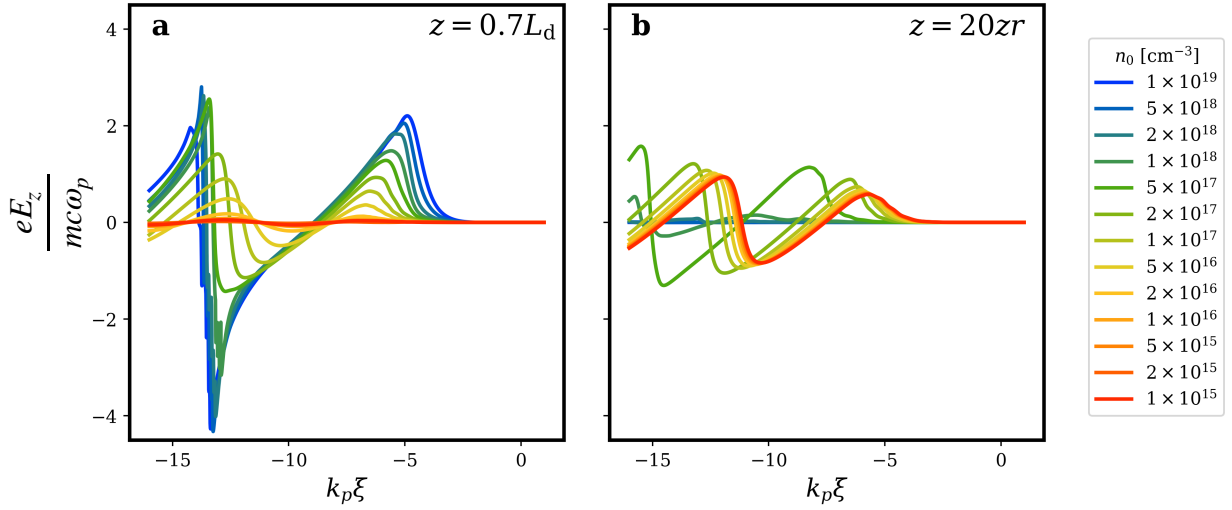


Figure 3.3: On-axis longitudinal electric field in normalized units for cases ranging from $n_0 = 1 \times 10^{19} \text{ cm}^{-3}$ down to $1 \times 10^{15} \text{ cm}^{-3}$. (a) At $z = 0.7L_d$, it is clear that the normalized wake amplitude decreases with density. The constancy of the position of the zero-crossing indicates that $\beta_w - 1 \propto \left(\frac{\omega_p}{\omega_0}\right)^2$ as in Lu et al. (b) At $z = 20z_R$, the amplitude of the cases with $n_0 \leq 5 \times 10^{17} \text{ cm}^{-3}$ is nearly identical. This is consistent with the scaling $L_{\text{guide}} \propto z_R$ at these lower densities.

scaling $L_{\text{guide}} \propto z_R$ and the bound $L_{\text{guide}} \geq 20z_R$ for these lower densities.

Finally, Fig. 3.4 shows the maximum value of $e|\hat{E}|/mc\omega_0$ as a function of the propagation distance scaled by L_d in (a) and by z_R in (b). This quantity is plotted instead of $e|\hat{A}|/mc\omega_p$ because the latter peaks at late time with a value much larger than a_0 . (The quantities are related by Eq. 2.57: $\hat{E} = ik_0\hat{A} + \partial_\xi\hat{A}$.) This effect results from photon deceleration associated with pump depletion [132, 133] and makes the physics of guiding less evident. Figure 3.4(a) shows that for densities $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$, $L_{\text{guide}} \approx 1.5L_d$. As mentioned above, we define L_{guide} for simplicity as the length required for $\max e|\hat{E}|/mc\omega_0$ to decrease to half of its initial value.¹ For lower densities, L_{guide} is clearly less than L_d and follows a weaker scaling than

¹Another relevant length scale is L_{drive} , the length scale over which the laser drives the nonlinear wake. For the cases in this study, L_{drive} is approximately the length required for the laser to attain $\max e|\hat{E}|/mc\omega_0 \approx 2$

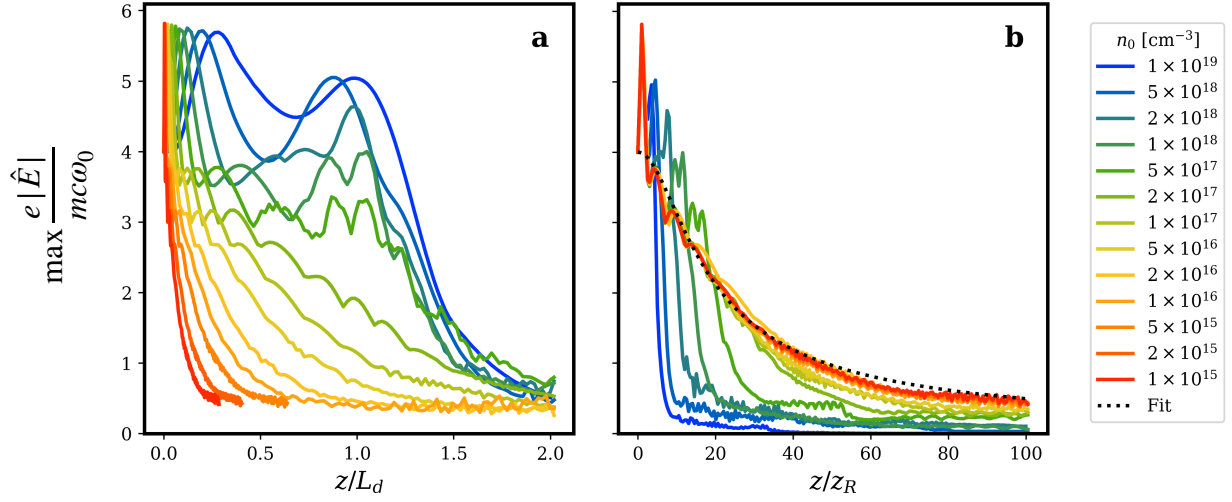


Figure 3.4: Evolution of the maximum of $e|\hat{E}|/mc\omega_0$ for cases ranging from $n_0 = 1 \times 10^{19} \text{ cm}^{-3}$ down to $1 \times 10^{15} \text{ cm}^{-3}$. (a) When plotted against z/L_d , it is clear that the cases with $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$ have $L_{\text{guide}} \approx 1.5L_d$ and exhibit similar dynamics on this length scale. The cases with $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$ do not propagate as far as L_d follow a different scaling. (b) When plotted against z/z_R , we see that $L_{\text{guide}} \approx 25z_R$ for $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$. The fit is the expected evolution under vacuum diffraction with a pulse with an effective Rayleigh range $z_{R,\text{eff}} = 12.5z_R$.

$L_{\text{guide}} \propto L_d$. Figure 3.4(b) shows that for densities $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$, $L_{\text{guide}} \approx 25z_R$. By inspection, we found that the data is reasonably well-described by an empirical fit of the form

$$\frac{e|\hat{E}|}{mc\omega_0} \approx \frac{a_0}{\sqrt{1 + (z/z_{R,\text{eff}})^2}} \quad (3.1)$$

for the cases with $\omega_0/\omega_p \gtrsim 75$ where the effective Rayleigh range is $z_{R,\text{eff}} = 12.5z_R$. In this case, $L_{\text{guide}} = 2z_{R,\text{eff}}$. The fit is overlaid in Fig. 3.4(b). This is the same form as for a laser pulse diffracting in vacuum with $z_R \rightarrow z_{R,\text{eff}}$.

Here we summarize the results from this subsection in dimensionless units. These QPAD simulations suggest that for the parameters $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 4$, the

so that $L_{\text{guide}} \approx L_{\text{drive}}$. The two length scales could be different for a different a_0 .

guiding length can be estimated as

$$L_{\text{guide}} \approx \begin{cases} 1.5L_d, & \omega_0/\omega_p \lesssim 50 \\ 25z_R, & \omega_0/\omega_p \gtrsim 75 \end{cases} \quad (3.2)$$

where there is a gradual transition for $50 \lesssim \omega_0/\omega_p \lesssim 75$.

3.2.3 Particle Acceleration

The results in the previous section showed that the scaling $L_{\text{guide}} \leq L_d$ and the scaling shifts from $L_{\text{guide}} \propto L_d$ shifts to $L_{\text{guide}} \propto z_R$ for $\omega_0/\omega_p \gtrsim 75$. These same scalings would be expected to apply to the energy gain and acceleration length of accelerating electrons because the acceleration length should satisfy $L_{\text{acc}} \propto L_{\text{guide}}$ when $L_{\text{guide}} < L_d$. To study this, a slice from the beam of test particles initially near the rear of the first plasma bubble is selected. The region with $k_p \xi \in [-11, -10.9]$ and $k_p r \in [0, 1]$ is shown by the white box in Fig. 3.5. This method works because the structure of the wake at early time is nearly identical for all the cases. Electrons initially located at values of $k_p \xi$ slightly less than this are lost early on because of their interaction with the electron sheath in the rear of the bubble. Electrons ahead of this value experience a lower initial accelerating gradient, but are also accelerated for a longer distance because it takes a longer distance for the rear of the bubble to cross them. A more sophisticated analysis was also performed by using a separate ξ value corresponding to the slice with the highest energy measured during simulation in regions where the beam current was still close to its initial value (i.e., where electrons were not lost). This more sophisticated analysis yielded similar results. Results from the less sophisticated analysis are presented for simplicity.

Figure 3.6(a) shows the evolution of the maximum Lorentz factor (which is almost the same for every particle in the slice) within this slice as a function of z/L_d for each of the 13 densities tested. For $n_0 \geq 1 \times 10^{18} \text{ cm}^{-3}$, the particles are accelerated for approximately $L_{\text{acc}} = L_d$, after which they lose energy because of dephasing. The energy gain for $n_0 \geq 2 \times 10^{17} \text{ cm}^{-3}$

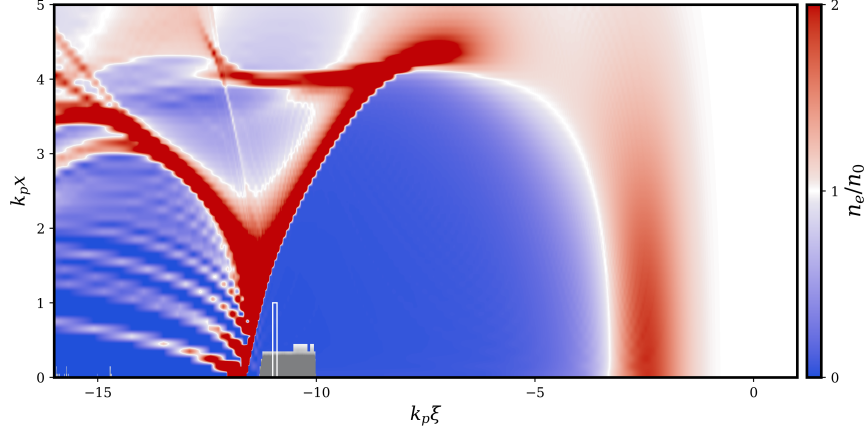


Figure 3.5: Illustration of the beam (gray) and the slice of the beam selected (white box) for examining the energy gain of each simulation. The same slice can be used because the structure of the wake at early time is nearly identical. A more sophisticated analysis using the ξ slice of maximum energy gain for each simulation yielded similar results. The simulation shown is the $n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$ case at $z = 0.5z_R$.

is found empirically to be $\Delta\gamma \approx 1.2\Delta\gamma_{\text{Lu}}$. On the other hand, for $n_0 \leq 5 \times 10^{16} \text{ cm}^{-3}$, the acceleration length and energy gain are significantly lower than the predictions from Lu et al. For each case in Fig. 3.6(a), the data is plotted up to the point at which $\Lambda = \Lambda_0/2$ for this slice. The current within this slice as a function of propagation distance is shown in Figure 3.6(b). In most cases, the current is constant and equal to its initial value until it abruptly transitions to 0. This occurs when the electron sheath at the rear of the bubble crosses the beam. This occurs in spite of the particles moving forward with respect to the etching laser position because of the competing effect of the contracting wake wavelength as the wake transitions from nonlinear ($k_p\lambda_w \approx 2\sqrt{a_{\text{max}}(z)}$) to linear ($k_p\lambda_w = 2\pi$). For some cases, the transition is less abrupt or exhibits oscillatory motion. This occurs when the rear sheath does not cross the particles during the transition to the linear wake. A more reliable indicator of when the beam is no longer useful such as the spot size or normalized emittance could potentially be used, but this one suffices for this study.

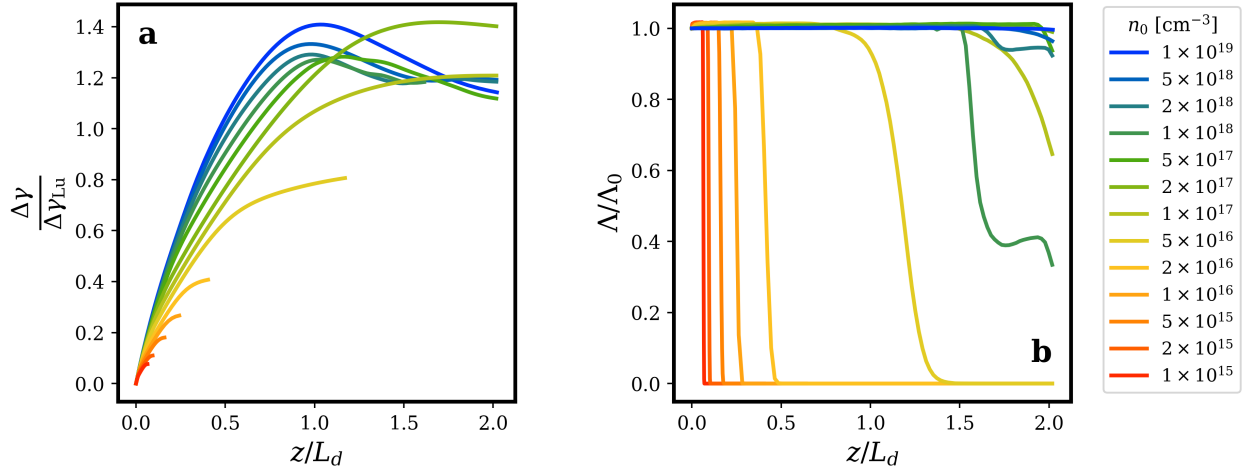


Figure 3.6: (a) Evolution of the maximum Lorentz factor for particles in the region $k_p \xi \in [-11, -10.9]$ and $k_p r \in [0, 1]$. The energy gain for $n_0 \geq 2 \times 10^{17} \text{ cm}^{-3}$ is approximately $\Delta\gamma = 1.2\Delta\gamma_{\text{Lu}}$. For $n_0 \leq 5 \times 10^{16} \text{ cm}^{-3}$, the acceleration length and energy gain are significantly lower than the predictions from Lu et al. (b) Beam current for electrons in this region. The data in (a) are plotted up to the lowest distance at which $\Lambda \leq \Lambda_0/2$.

From the data in Fig. 3.6, the energy gain $\Delta E \equiv \max_z \Delta\gamma(z)mc^2$ and acceleration length $L_{\text{acc}} \equiv \arg \max_z \Delta\gamma(z)$ can be extracted for each density. These are plotted against the density in Fig. 3.7. Figure 3.7(a) shows that the Lu et al. scaling $\Delta\gamma = \frac{2}{3} \left(\frac{\omega_0}{\omega_p} \right)^2 a_0$ is accurate for $n_0 \geq 5 \times 10^{16} \text{ cm}^{-3}$. The validity of the scaling law for $\Delta\gamma$ extends beyond where $L_{\text{guide}} < L_d$ because of the competing effect of the wake wavelength described above. For densities below this, there is a clear breakpoint in the data for which a weaker power-law scaling holds. By inspection, the expression $\Delta\gamma = 25k_p z_R$ models the data well for $n_0 \leq 2 \times 10^{16} \text{ cm}^{-3}$. One could obtain a better fit through regression, but this simple model makes good physical sense. In Fig. 3.7(b), one can see the same trend for the acceleration length L_{acc} . In this case, the Lu et al. scaling is almost exact for $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$. For the density range $5 \times 10^{16} \text{ cm}^{-3} \leq n_0 \leq 2 \times 10^{17} \text{ cm}^{-3}$, the Lu et al. scaling is less accurate but is still within a factor of 2. For $n_0 \leq 2 \times 10^{16} \text{ cm}^{-3}$, there is clearly a different scaling. By inspection, the

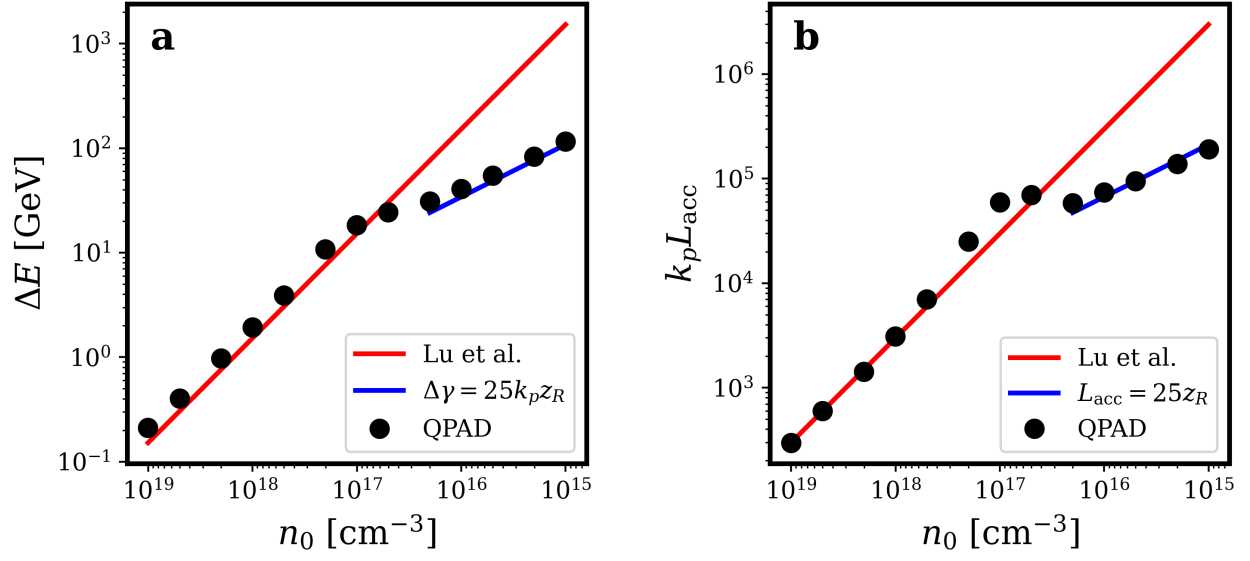


Figure 3.7: The maximum energy gain $\Delta E \equiv \max_z \Delta\gamma(z)mc^2$ and acceleration length $L_{\text{acc}} \equiv \arg \max_z \Delta\gamma(z)$ are extracted from the data in Fig. 3.6 and respectively plotted in (a) and (b). (a): the energy gain is accurately modeled by Lu et al. for $n_0 \geq 5 \times 10^{16} \text{ cm}^{-3}$. $\Delta\gamma = 25k_p z_R$ models the data well for $n_0 \leq 2 \times 10^{16} \text{ cm}^{-3}$. (b): the Lu et al. model is almost exact for the acceleration length for $n_0 \geq 5 \times 10^{17} \text{ cm}^{-3}$, within a factor of 2 down to $n_0 = 5 \times 10^{16} \text{ cm}^{-3}$. The acceleration length is more accurately described by the model $L_{\text{acc}} = 25z_R$ for $n_0 \leq 2 \times 10^{16} \text{ cm}^{-3}$.

scaling $L_{\text{acc}} = 25z_R$ fits the data well. The measured values for ΔE and $k_p L_{\text{acc}}$ in Fig. 3.6 are tabulated in dimensionless units in Table 3.2

Here we summarize the results from this subsection in dimensionless units. These QPAD simulations suggest that for the parameters $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 4$, the change in the Lorentz factor (which is the normalized energy gain) and acceleration length may be estimated as

$$\Delta\gamma \approx \begin{cases} \frac{2}{3} \left(\frac{\omega_0}{\omega_p} \right)^2 a_0, & \omega_0/\omega_p \lesssim 150 \\ 25k_p z_R, & \omega_0/\omega_p \gtrsim 250 \end{cases} \quad (3.3)$$

n_0 [cm ⁻³]	ΔE [GeV]	L_{acc} [cm]
1×10^{19}	0.21	0.050
5×10^{18}	0.40	0.14
2×10^{18}	0.98	0.53
1×10^{18}	1.9	1.6
5×10^{17}	3.9	5.3
2×10^{17}	11	29
1×10^{17}	18	100
5×10^{16}	24	170
2×10^{16}	30	220
1×10^{16}	40	390
5×10^{15}	54	710
2×10^{15}	83	1600
1×10^{15}	120	3200

Table 3.2: Tabulation of the simulated values for the energy gain and acceleration length as plotted in Fig. 3.7. Other simulation parameters can be cross-referenced with Table 3.1.

and

$$L_{\text{acc}} \approx \begin{cases} L_d, & \omega_0/\omega_p \lesssim 150 \\ 25z_R, & \omega_0/\omega_p \gtrsim 250. \end{cases} \quad (3.4)$$

where both quantities exhibit a gradual transition for $150 \lesssim \omega_0/\omega_p \lesssim 250$.

3.3 Discussion and conclusions

In this study, the scaling laws for the acceleration length and energy gain in LWFA were tested by varying the dimensionless frequency ratio ω_0/ω_p for fixed dimensionless laser parameters of $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 4$. For a reference wavelength of 1000 nm, the cases correspond to plasma densities ranging from 1×10^{19} cm⁻³ to 1×10^{15} cm⁻³. By examining the evolution of the laser and wake, it was found that the laser guiding length L_{guide} (defined here as the z for which $\max e|\hat{E}|/mc\omega_0$ decreases to half of its initial value) satisfies $L_{\text{guide}} \approx 1.5L_d$ for $\omega_0/\omega_p \lesssim 50$, consistent with the assumptions of Lu et al (Ref. [1]).

However, for $\omega_0/\omega_p \gtrsim 75$, the simulations suggest that $L_{\text{guide}} < L_d$ and $L_{\text{guide}} \approx 25z_R$, a less favorable scaling. The transition to this scaling also manifests in the properties of electron acceleration. We find that the scaling law of Lu et al. $L_{\text{acc}} \approx L_d$ holds for $\omega_0/\omega_p \lesssim 150$, while $L_{\text{acc}} = 25z_R$ better fits the data for $\omega_0/\omega_p \gtrsim 250$. Similarly, the change in the Lorentz factor follows the scaling law of Lu et al. $\Delta\gamma \approx \frac{2}{3} \left(\frac{\omega_0}{\omega_p}\right)^2 a_0$ for $\omega_0/\omega_p \lesssim 150$, while $\Delta\gamma \approx 25k_p z_R$ better fits the data for $\omega_0/\omega_p \gtrsim 250$. The thresholds are different for laser guiding and for electron acceleration because of subtle details in the evolution of the wake and laser beyond depletion.

Here we comment on the accuracy of the simulations. In principle, one would expect that the required resolution $k_p \Delta_s$ scales as the Rayleigh range z_R . Using a fixed value increased the cost of the simulations but simplified the analysis by removing the need for numerical convergence testing in this parameter. The $k_p \Delta_\xi = 0.01$ parameter may under-resolve significant redshifting in some of the lower-density simulations. Reference [131] used the condition $k_0 \Delta_\xi \leq 2.45$ to set the longitudinal resolution. For the lowest-density simulation in this study, this would correspond to a resolution of $k_p \Delta_\xi \approx 0.002$, lower than the simulated value of $k_p \Delta_\xi = 0.01$. The resolution required to resolve redshifting can be estimated as follows. The maximum change in laser wavenumber can be estimated as $\frac{\Delta k}{k_0} \approx -\frac{1}{2} \left(\frac{k_p}{k_0}\right)^2 k_p L_{\text{guide}}$ [68]. Since $\mathbf{E}_\perp = e^{ik_0\xi} \hat{A} \hat{\mathbf{e}}$, this results in an oscillation in the laser envelope with wavelength $\lambda_{\text{osc}} = \frac{2\pi}{|\Delta k|}$. Using the estimate, we obtain $k_p \lambda_{\text{osc}} \approx 4\pi \left(\frac{k_0}{k_p}\right) (k_p L_{\text{guide}})^{-1}$. Taking $L_{\text{guide}} \approx 25z_R$ and $k_p w_0 = 4$, we obtain $k_p \lambda_{\text{osc}} = 0.06$ independently of k_0 and k_p . This wavelength is marginally resolved with the simulated resolution $k_p \Delta_\xi = 0.01$. Nonetheless, more simulations are needed at higher longitudinal resolution to confirm the results in this chapter for the cases with $k_0/k_p \gtrsim 250$.

The accuracy of QPAD for modeling the ponderomotive force in the presence of significantly redshifted laser light is also an area for investigation. Remarkable agreement between OSIRIS and QPAD was found in Ref. [8] in the presence of significantly redshifted laser light. The reasons for the accuracy are not completely clear. The redshifted light can be resolved by

making Δ_ξ sufficiently small, but the accuracy of the expression for $\mathbf{J}$ in terms of $\hat{A}$ is not obvious. Explaining the degree to which they agree is an area for future investigation. Another future direction is to validate QPAD at higher ratios of ω_0/ω_p against fully electromagnetic PIC simulations. This was done at $\omega_0/\omega_p = 25$ in Ref. [8]. These simulations become impractical when ω_0/ω_p is increased in the lab frame. However, the interaction can in principle be modeled with fully electromagnetic PIC by simulating the interaction in a Lorentz-boosted frame. This is an area for future investigation.

Aside from validating the results, a natural next step to extend the research would be to repeat the study for other pulse shapes and amplitudes to confirm that the behavior of $L_{\text{guide}} \propto z_R$ is satisfied beyond a threshold ω_0/ω_p . If so, then knowledge of the value of ω_0/ω_p at the transition and the value of the coefficient β for which $L_{\text{guide}} = \beta z_R$ could be valuable. One could attempt to find a more general parameterization for L_{guide} as a function of the dimensionless parameters a_0 , $k_p w_0$, $k_p \tau_0$, and k_0/k_p . One could also determine how the scaling depends on the degree of beam loading or explore varying the focal position of the laser within the plasma.

Based on the preliminary conclusions of this study, there are several consequences. In the near-term, the 5×10^{16} and $2 \times 10^{16} \text{ cm}^{-3}$ cases may be on the experimental horizon for the ELI laser system. It may then be possible to design an LWFA stage which can accelerate electrons using controlled self-injection to the order of 20-30 GeV over $\sim 2 \text{ m}$ for these laser and plasma parameters. (Self injection is not modeled by QPAD, but the simulations give an estimate for the energy gain that would be obtained by particles injected toward the beginning of the LWFA stage.) While this would be scientifically exciting, these simulations suggest that the energies would not be in the 30 – 80 GeV range predicted by the Lu et al. scaling for these laser and plasma parameters without additional improvements. In the short term, this motivates both computational and experimental work to optimize the laser profile to improve guiding. Improving the guiding by a factor of 2 could enable single-stage acceleration of electrons to the range of 50 GeV in the near term at laser facilities with 3–10 PW laser

systems. Guiding can potentially be improved by customization of the usual parameters for a single pulse ($a_0, w_0, \tau_{\text{rise}}, \tau_{\text{fall}}$) as well as full spatiotemporal pulse shaping, e.g., a ξ -dependent spot size as in Ref. [7]. As mentioned in the introduction, Ref. [1] proposed scaling the laser amplitude as $a_0 \propto (k_0/k_p)^{1/10}$ to improve guiding as the density is decreased. Methods to improve the guiding using a precursor laser pulse as in Ref. [134] could also be explored. Critically, these concepts appear not to have been explored at the relevant values of ω_0/ω_p in the range 150 – 250. For example, the precursor study of Ref. [134] was carried out at $\omega_0/\omega_p \approx 34$. One could also add modern machine-learning methods into this approach to optimize the laser guiding length or electron energy gain.

The results also have long-term consequences. The simulations suggest the hypothesis that for an initial laser profile that scales with k_p^{-1} as in this study, the same phenomenon will be observed that $L_{\text{guide}} \propto L_d$ for densities above some threshold density and $L_{\text{guide}} \propto z_R$ below this value; these scalings would also be expected to apply to L_{acc} and $\Delta\gamma$ at a potentially different threshold density as seen here. If this is true, then there are several important ramifications. First, the laser-to-beam efficiency for a fully loaded wake will be constant for densities above the threshold, but decrease for densities below this. This is because the maximum charge that can be loaded in the wake scales as $Q \propto k_p^{-1}$, as shown in Chapter 2. The laser energy scales as $E_L \propto k_0^2 k_p^{-3}$ for fixed $k_p w_0$ and $\omega_p \tau$. Therefore, if $\Delta\gamma \propto (k_0/k_p)^2$, then the laser-to-beam efficiency given by $\eta = Q\Delta\gamma/E_L$ is constant in k_0 and k_p . On the other hand, for the scaling $\Delta\gamma \propto k_p z_R$ for a fixed $k_p w_0$ (so that $k_p z_R \propto k_0/k_p$), the laser-to-beam efficiency scales as $\eta \propto k_p/k_0$. Thus, the efficiency decreases with the scaling. To maximize efficiency, it is then necessary operate in the higher-density regime for which $L_{\text{acc}} \propto L_d$. One can attempt to push down the density at which the transition occurs by tailoring the pulse to maximize guiding. However, when using a uniform plasma, these effects ultimately place a constraint on the parameter space of plasma densities that is viable in the context of linear collider designs.

We close by noting that the results in this study also add contextual motivation to study

LWFA driven by flying focus laser pulses. In Ref. [8], it was pointed out that LWFA driven by a flying focus with an extended acceleration length could be compared to a LWFA driven by a conventional pulse (as in these simulations) with the same acceleration length but at a lower plasma density. But because of the diminishing returns in the energy gain and efficiency implied by these results, the flying focus becomes comparatively more advantageous as the acceleration length is increased. The argument presented in Ref. [8], which quantified the comparison by assuming $\Delta\gamma \approx \Delta\gamma_{\text{Lu}}$ for all densities, is strengthened by these results.

CHAPTER 4

Linear Propagation of Laser Pulses in Cold, Uniform Plasma

In most applications of laser-plasma interactions, the evolution of the laser is highly nonlinear in the sense that Maxwell's equations are nonlinear when coupled to the charge and current density. However, linear laser propagation, which occurs in the limit $a_0 \ll 1$, remains important to understand. In the context of this dissertation, laser-plasma interactions with flying focus pulses involve a combination of linear and nonlinear physics; the linear physics plays a large role in choosing the design parameters and describes how the laser propagates for the parts of the laser far from focus. More generally, the concepts which emerge from studying linear laser propagation, including phase and group velocity, beam solutions of Maxwell's equations, laser guiding, and pulse compression, form a basic picture for understanding nonlinear interactions. Such concepts retain descriptive power and have been fruitfully applied to build empirical models describing nonlinear processes.

Although linear laser propagation is treated extensively in the literature, the evolution of the laser envelope's transverse profile and longitudinal profile are typically considered separately (e.g. in Chapters 9 and 16 of Siegman's *Lasers* [135] and Chapters 16 and 18 of Zangwill's *Modern Electrodynamics* [136]). Quantitatively, standard texts such as these deal separately with one-dimensional pulses of the form $E_{\perp}(t, z, \mathbf{x}_{\perp}) = B(t, z)e^{i(k_0 z - \omega_0 t)}$ and continuous-wave beams of the form $E_{\perp}(t, z, \mathbf{x}_{\perp}) = C(z, \mathbf{x}_{\perp})e^{i(k_0 z - \omega_0 t)}$, the Gaussian beam being an example of $C(z, \mathbf{x}_{\perp})$ in the latter. (See equations 9.28 and 16.2 in Siegman and 16.126 and 18.84 in Zangwill.) For most applications, this is sufficient because either the

transverse or longitudinal evolution dominates. However, in the context of spatiotemporally structured laser pulses, the evolution of the full envelope $\hat{E}(t, z, \mathbf{x}_\perp)$ must be considered because the longitudinal and transverse structure are coupled. I did find several references considering the full structure, but these lacked a straightforward development or exposition. The calculations in this chapter address what I perceive to be a gap in the literature.

The most valuable result from this chapter is probably a clear description of the evolution of laser pulses initialized in time as $\mathbf{E}_\perp(0, z, \mathbf{x}_\perp) = B(z)C(z, \mathbf{x}_\perp)e^{ik_0z}\hat{\mathbf{e}}$ (generalized from Eq. 2.23). This form, which captures the essential features of high-power laser pulses produced using conventional optical methods, is commonly used to initialize simulations of laser-plasma interactions. Thus, understanding its properties is important for simulators. This understanding is also valuable for theorists and experimentalists. As a case in point, understanding these details for conventional pulses was the key to formulating the theory of Arbitrarily Structured Laser (ASTRL) pulses, presented in Chapter 5, and their use in LWFA in Chapter 6. A particular question about this class of laser pulses drove me to develop these calculations: what is the relationship between a continuous-wave Gaussian beam and a pulse initialized as a Gaussian beam multiplied by a longitudinal envelope $B(z)$ when accounting for linear dispersion? In cold, uniform plasma, answer is that the full evolution is approximately $\mathbf{E}_\perp(t, z, \mathbf{x}_\perp) \approx B(z - v_g t)C(z, \mathbf{x}_\perp)e^{i(k_0z - \omega_0 t)}\hat{\mathbf{e}}$, where $\omega_0 \equiv \omega(k_0\hat{\mathbf{z}})$ and $v_g = \hat{\mathbf{z}} \cdot \nabla_{\mathbf{k}_0}\omega(k_0\hat{\mathbf{z}})\hat{\mathbf{e}}$ come from the dispersion relation for electromagnetic waves in cold, uniform plasma: $\omega(\mathbf{k}) = \omega_p^2 + c^2k^2$. I sense that this was already widely known at an intuitive level. These calculations show why and quantify the approximation.

Methodologically, the approach here is an integral formulation. It mostly follows and generalizes calculations in Zangwill. The case of a linear dielectric with an arbitrary dispersion relation is first considered. (An additional benefit of these calculations is an illustration of the power of dispersion relations for solving initial-value problems; in my education in plasma physics, dispersion relations were almost only presented as a way of understanding monochromatic waves.) Applications to the cases of vacuum and cold, uniform plasma for

separable laser pulses are then presented. Self-contained differential formulations for these special cases are presented in Chapters 5 and 6. The integral and differential formulations are complementary. The differential approach is more succinct (which is the main reason it was presented in the published work corresponding to Chapters 5 and 6) and is better-suited to finite-difference methods. The integral approach is more natural for consideration of arbitrary dispersion relations and provides explicit solutions. Clarity is my goal in the exposition.

4.1 Integral solutions to linear partial differential equations using dispersion relations

Suppose we are given a linear partial differential equation $\hat{L}\psi = 0$, where $\hat{L}$ is a linear differential operator and $\psi(t, \mathbf{x})$ is a complex-valued scalar function, along with an initial condition $\psi(0, \mathbf{x}) = \psi_0(\mathbf{x})$. For example, for electromagnetic waves propagating linearly in cold, uniform plasma, we have $\hat{L} = \partial_t^2 - c^2 \nabla^2 + \omega_p^2$ and ψ is any component of $\mathbf{E}$, $\mathbf{B}$, or $\mathbf{A}$.

How can we derive the evolution for $\psi(t, \mathbf{x})$ in terms of the initial data $\psi(0, \mathbf{x})$? The dispersion relation is one approach. Suppose that we know the dispersion relation $\omega(\mathbf{k})$, which has the property that plane waves

$$u_{\mathbf{k}}(t, \mathbf{x}) \equiv \exp \left[i \left(\mathbf{k} \cdot \mathbf{x} - \omega(\mathbf{k})t \right) \right] \quad (4.1)$$

satisfy $\hat{L}u_{\mathbf{k}} = 0$ for each $\mathbf{k}$. Also, let $\tilde{\psi}_0(\mathbf{k})$ be the Fourier transform of the initial condition $\psi_0(\mathbf{x})$, so that

$$\tilde{\psi}_0(\mathbf{k}) = \frac{1}{2\pi} \int \psi_0(\mathbf{x}) \exp [-i\mathbf{k} \cdot \mathbf{x}] d\mathbf{x} \quad (4.2)$$

and

$$\psi_0(\mathbf{x}) = \int \tilde{\psi}_0(\mathbf{k}) \exp [i\mathbf{k} \cdot \mathbf{x}] d\mathbf{k} \quad (4.3)$$

Then from Eq. 4.1 and the linearity of $\hat{L}$, a solution to the initial value problem may be constructed as

$$\psi(t, \mathbf{x}) = \int \tilde{\psi}_0(\mathbf{k}) \exp \left[i \left(\mathbf{k} \cdot \mathbf{x} - \omega(\mathbf{k})t \right) \right] d\mathbf{k} \quad (4.4)$$

Is this *the* solution? This depends on whether the PDE admits unique solutions. This is equivalent to whether there is a multi-valued dispersion relation, i.e., multiple values of $\omega(\mathbf{k})$ for which $u_{\mathbf{k}}(t, \mathbf{x})$ in Eq. 4.1 is a solution of $\hat{L}\psi$ for a given $\mathbf{k}$. In this case, data for N initial time derivatives of ψ are required to express the unique solution for an N -valued dispersion relation. This is not a totally arcane issue, because the dispersion relation for electromagnetic waves in cold plasma is double-valued: $\omega_{\pm}(\mathbf{k}) = \pm\sqrt{c^2k^2 + \omega_p^2}$. In practice, Eq. 4.4 is the unique solution if the initial condition is composed by only one branch of the dispersion relation. In the context of light in plasma, this means only forward-going or backward-going waves. This assumption is true for the cases considered in this chapter. One could generalize the approach to relax this assumption.

The restriction of ψ to be a scalar function avoids a complication in Eq. 4.4: if ψ is promoted to a vector, then polarization must be added to the plane waves in Eq. 4.1. As in the case of a multi-valued dispersion relation, a more general superposition than Eq. 4.4 can be expressed, in this case by taking an additional integral over polarization. This complicates matching the initial condition to the superposition of plane waves unless the initial condition is composed of plane waves of only a single polarization. This is also true when ψ is not promoted to a vector but is the component of a vector: Eq. 4.4 remains the unique solution if the initial condition for the full vector solution is only composed of plane waves of a single polarization. This is assumed as an approximation throughout this chapter when Eq. 4.4 is applied to construct solutions of $\mathbf{E}(t, \mathbf{x})$ for laser pulses.

With these two caveats, Eq. 4.4 illustrates an important point: the dispersion relation for a linear partial differential equation is a natural tool for the solution of initial value problems.

4.1.1 Formulation in terms of the envelope

Suppose that $\tilde{\psi}_0(\mathbf{k})$ is strongly peaked at some $\mathbf{k}_0$ in Fourier space. In this case, it is natural to redefine the initial data $\psi_0(\mathbf{x})$ in terms of an envelope $A_0(\mathbf{x})$ defined according to

$$\psi_0(\mathbf{x}) = \exp[i\mathbf{k}_0 \cdot \mathbf{x}] A_0(\mathbf{x}) \quad (4.5)$$

To relate the solution $\psi(t, \mathbf{x})$ to $A_0(\mathbf{x})$, it is necessary to relate $\tilde{\psi}_0(\mathbf{k})$ to $A_0(\mathbf{x})$ so that Eq. 4.4 can be applied. This comes from Eq. 4.2:

$$\tilde{\psi}_0(\mathbf{k}) = \frac{1}{2\pi} \int A_0(\mathbf{x}) \exp[i\mathbf{k}_0 \cdot \mathbf{x}] \exp[-i\mathbf{k} \cdot \mathbf{x}] d\mathbf{x} \quad (4.6)$$

$$= \tilde{A}_0(\mathbf{k} - \mathbf{k}_0) \quad (4.7)$$

where $\tilde{A}_0(\mathbf{k})$ is the Fourier transform of $A_0(\mathbf{x})$. Inserting this into Eq. 4.4 and substituting $\mathbf{u} \equiv \mathbf{k} - \mathbf{k}_0$, we obtain:

$$\psi(t, \mathbf{x}) = \exp \left[i \left(\mathbf{k}_0 \cdot \mathbf{x} - \omega(\mathbf{k}_0)t \right) \right] \int \tilde{A}_0(\mathbf{u}) \exp \left[i \left(\mathbf{u} \cdot \mathbf{x} - [\omega(\mathbf{u} + \mathbf{k}_0) - \omega(\mathbf{k}_0)] t \right) \right] d\mathbf{u} \quad (4.8)$$

Therefore, we can exactly express the evolution of $\psi(t, \mathbf{x})$ in terms of an evolving envelope $A(t, \mathbf{x})$:

$$\psi(t, \mathbf{x}) = \exp \left[i \left(\mathbf{k}_0 \cdot \mathbf{x} - \omega(\mathbf{k}_0)t \right) \right] A(t, \mathbf{x}) \quad (4.9)$$

$$A(t, \mathbf{x}) \equiv \int \tilde{A}_0(\mathbf{u}) \exp \left[i \left(\mathbf{u} \cdot \mathbf{x} - [\omega(\mathbf{u} + \mathbf{k}_0) - \omega(\mathbf{k}_0)] t \right) \right] d\mathbf{u} \quad (4.10)$$

It is then natural to Taylor-expand $\omega(\mathbf{u} + \mathbf{k}_0)$ about $\mathbf{u} = \mathbf{0}$ (i.e., $\mathbf{k} = \mathbf{k}_0$):

$$\omega(\mathbf{u} + \mathbf{k}_0) \approx \omega(\mathbf{k}_0) + u_a \partial_{k_a} \omega(\mathbf{k}_0) + \frac{1}{2!} u_a u_b \partial_{k_a} \partial_{k_b} \omega(\mathbf{k}_0) + \dots \quad (4.11)$$

where a summation is implied by repeated indices. With this Taylor expansion, the envelope may be exactly expressed as:

$$A(t, \mathbf{x}) = \int \tilde{A}_0(\mathbf{u}) \exp \left[i \left(\mathbf{u} \cdot [\mathbf{x} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0)t] - \frac{1}{2!} u_a u_b \partial_{k_a} \partial_{k_b} \omega(\mathbf{k}_0)t - \dots \right) \right] d\mathbf{u} \quad (4.12)$$

This provides a sequence of increasingly accurate approximations to $A(t, \mathbf{x})$, and hence $\psi(t, \mathbf{x})$.

4.1.2 Phase and group velocity

From Eq. 4.9, the full solution $\psi(t, \mathbf{x})$ may be expressed as

$$\psi(t, \mathbf{x}) = \exp \left[i \mathbf{k}_0 \cdot (\mathbf{x} - \mathbf{v}_\phi t) \right] A(t, \mathbf{x}) \quad (4.13)$$

where $\mathbf{v}_\phi$ is the phase velocity defined as

$$\mathbf{v}_\phi \equiv \frac{\omega(\mathbf{k}_0)}{k_0} \hat{\mathbf{k}}_0 \quad (4.14)$$

Because $A(t, \mathbf{x})$ remains peaked in Fourier space at $\mathbf{u} = \mathbf{0}$, the total phase of $\psi(t, \mathbf{x})$ is approximately $i \mathbf{k}_0 \cdot (\mathbf{x} - \mathbf{v}_\phi t)$ plus a slowly varying offset from $A(t, \mathbf{x})$. Hence, the phase fronts approximately travel at the phase velocity.

If the Taylor series in Eq. 4.12 is truncated at first order, one obtains the approximation

$$A(t, \mathbf{x}) \approx \int \tilde{A}_0(\mathbf{u}) \exp \left[i \mathbf{u} \cdot (\mathbf{x} - \nabla_{\mathbf{k}} \omega(\mathbf{k}_0) t) \right] d\mathbf{u} \quad (4.15)$$

$$= A_0(\mathbf{x} - \mathbf{v}_g t) \quad (4.16)$$

where the group velocity $\mathbf{v}_g$ is defined as

$$v_g \equiv \nabla_{\mathbf{u}} \omega(\mathbf{k}_0) \quad (4.17)$$

Thus, to first order in the Taylor expansion for $\omega(\mathbf{k})$, the envelope maintains its shape but moves with at a constant velocity v_g .

4.1.3 The choice of t versus z as the propagation variable

All of the above was worked out with the initial condition specified at $t = 0$. This leads to t serving as the propagation variable. This was done for ease of notation. All of the above calculations can be repeated using one of the spatial variables, most commonly z , to prescribe the initial condition and to serve as the propagation variable. In the context of optics and plasma physics, the latter is usually more natural because of the common occurrence of discrete

transitions in space (e.g., a vacuum-plasma interface). Given a dispersion relation $\omega(\mathbf{k})$, the theory can be reformulated with z as the propagation variable by inverting to obtain $k_z(\omega, \mathbf{k}_\perp)$ and Taylor-expanding this in the integral. Mathematically, there is no difference between the two approaches aside from relabeling of variables and additional bookkeeping if a notational distinction between temporal and spatial variables is maintained. In this chapter, t is used as the propagation variable for simplicity. In several cases, the result when using z as the propagation variable is noted.

4.2 Linear propagation of laser pulses in vacuum

In the next two sections, we examine approximations for the envelope in Equation 4.12 using the dispersion relation for electromagnetic waves in cold, uniform plasma:

$$\omega(\mathbf{k}) = \sqrt{k_p^2 + k^2} \quad (4.18)$$

where dimensionless units are used so that $c = 1$ and hence $\omega_p = k_p$. We assume that the laser pulse is composed of plane waves with polarizations all in the same plane so that the scalar equation Eq. 4.9 is approximately true in vector form:

$$\mathbf{E}_\perp(t, \mathbf{x}) \approx \exp \left[i \left(\mathbf{k}_0 \cdot \mathbf{x} - \omega(\mathbf{k}_0)t \right) \right] A(t, \mathbf{x}) \hat{\mathbf{e}} + \text{c.c.} \quad (4.19)$$

and Eq. 4.12 is unmodified. The derivatives of the above dispersion relation can be computed and substituted into the Taylor series in Eq. 4.12. The derivatives up to third order may be computed as

$$\partial_{k_a} \omega(\mathbf{k}) = \frac{k_a}{\omega(\mathbf{k})} \quad (4.20)$$

$$\partial_{k_a} \partial_{k_b} \omega(\mathbf{k}) = \frac{\delta_{ab}}{\omega(\mathbf{k})} - \frac{k_a k_b}{\omega(\mathbf{k})^3} \quad (4.21)$$

$$\partial_{k_a} \partial_{k_b} \partial_{k_c} \omega(\mathbf{k}) = 3 \frac{k_a k_b k_c}{\omega(\mathbf{k})^5} - \frac{\delta_{ab} k_c}{\omega(\mathbf{k})^3} - \frac{\delta_{ac} k_b}{\omega(\mathbf{k})^3} - \frac{\delta_{bc} k_a}{\omega(\mathbf{k})^3} \quad (4.22)$$

Using these equations to evaluate the derivatives of $\omega(\mathbf{k})$ at the central wavevector $\mathbf{k}_0 = k_0 \hat{\mathbf{z}}$, we obtain the following nonzero terms:

$$\partial_{k_z} \omega(\mathbf{k}_0) = \frac{k_0}{\omega_0} \quad (4.23)$$

$$\partial_{k_z}^2 \omega(\mathbf{k}_0) = \frac{1}{\omega_0} - \frac{k_0^2}{\omega_0^3} = \frac{k_p^2}{\omega_0^3} \quad (4.24)$$

$$\partial_{k_{x,y}}^2 \omega(\mathbf{k}_0) = \frac{1}{\omega_0} \quad (4.25)$$

$$\partial_{k_z}^3 \omega(\mathbf{k}_0) = \frac{3k_0^3}{\omega_0^5} - \frac{3k_0}{\omega_0^3} = -\frac{3k_0 k_p^2}{\omega_0^5} \quad (4.26)$$

$$\partial_{k_{x,y}}^2 \partial_{k_z} \omega(\mathbf{k}_0) = -\frac{k_0}{\omega_0^3} \quad (4.27)$$

The $\partial_{k_{x,y}}$ terms mean that the equation holds for both ∂_{k_x} and ∂_{k_y} . In the limit that $k_0/k_p \gg 1$, these are approximately given by

$$\partial_{k_z} \omega(\mathbf{k}_0) \approx 1 - \frac{1}{2} \frac{k_p^2}{k_0^2} \quad (4.28)$$

$$\partial_{k_z}^2 \omega(\mathbf{k}_0) \approx \frac{k_p^2}{k_0^3} \quad (4.29)$$

$$\partial_{k_{x,y}}^2 \omega(\mathbf{k}_0) \approx \frac{1}{k_0} \quad (4.30)$$

$$\partial_{k_z}^3 \omega(\mathbf{k}_0) \approx -\frac{3k_p^2}{\omega_0^4} \quad (4.31)$$

$$\partial_{k_{x,y}}^2 \partial_{k_z} \omega(\mathbf{k}_0) \approx -\frac{1}{k_0^2} \quad (4.32)$$

We will now consider the vacuum case, for which $k_p = 0$.

4.2.1 First-order propagation

To first order in vacuum, $\mathbf{v}_g = \hat{\mathbf{z}}$. Therefore we obtain the approximation to first order

$$A(t, \mathbf{x}) \approx A_0(z - t, \mathbf{x}_\perp). \quad (4.33)$$

Suppose that the envelope is initially separable in its longitudinal and transverse spatial dependence:

$$A_0(\mathbf{x}) = B_0(z)C_0(\mathbf{x}_\perp) \quad (4.34)$$

Then to first order, the envelope is approximated as

$$A(t, \mathbf{x}) \approx B_0(z - t)C_0(\mathbf{x}_\perp) \quad (4.35)$$

Hence, to first order, the longitudinal envelope propagates in the direction of $\mathbf{k}_0$ at the speed of light and the transverse profile does not change. The initial condition's separability in longitudinal and transverse spatial dependence leads to a solution with this same property.

4.2.2 Second-order propagation: diffraction and paraxial beams

From Equations 4.24 and 4.25, we have $\partial_{k_z}^2 \omega(\mathbf{k}_0) = 0$ and $\partial_{k_x}^2 \omega(\mathbf{k}_0) = \partial_{k_y}^2 \omega(\mathbf{k}_0) = k_0^{-1}$ in vacuum. All other second-order derivatives vanish. Substituting these derivatives into the Taylor expansion, the second-order approximation to the envelope is given by

$$A(t, \mathbf{x}) \approx \int \tilde{A}_0(\mathbf{u}) \exp \left[iu_z(z - t) + i\mathbf{u}_\perp \cdot \mathbf{x}_\perp - \left(\frac{it}{2k_0} \right) u_\perp^2 \right] d\mathbf{u} \quad (4.36)$$

The integrand is separable in z and $\mathbf{x}_\perp$. Thus, for the separable initial condition in Eq. 4.34, the envelope remains separable:

$$A(t, \mathbf{x}) \approx B_0(z - t)C(t, \mathbf{x}_\perp) \quad (4.37)$$

where $C(t, \mathbf{x}_\perp)$ is given by

$$C(t, \mathbf{x}_\perp) = \int \tilde{C}_0(\mathbf{u}_\perp) \exp \left[i\mathbf{u}_\perp \cdot \mathbf{x}_\perp - \left(\frac{it}{2k_0} \right) u_\perp^2 \right] d\mathbf{u}_\perp \quad (4.38)$$

and $\tilde{C}_0(\mathbf{x}_\perp)$ is the Fourier transform of $C_0(\mathbf{x}_\perp)$. This is the Fresnel diffraction integral under the paraxial approximation and with t as the propagation variable instead of z . For example, suppose that the initial function $C_0(\mathbf{x}_\perp)$ is a gaussian at time $t = t_0$:

$$C_0(\mathbf{x}_\perp) = \exp \left[-\frac{x_\perp^2}{w_0^2} \right] \quad (4.39)$$

One may then compute

$$\tilde{C}_0(\mathbf{u}_\perp) = \frac{1}{2\pi} (\pi w_0^2)^{d_\perp/2} \exp \left[-\frac{w_0^2 u_\perp^2}{4} \right] \quad (4.40)$$

where $d_\perp$ is the transverse dimensionality (2 in 3D). Inserting this into Eq. 4.38, one obtains a temporal Gaussian beam:

$$C(t, \mathbf{x}_\perp) = \left[\frac{1}{q(t)} \right]^{d_\perp/2} \exp \left[-\frac{ik_0 x_\perp^2}{2q(t)} \right] \quad (4.41)$$

where $q(t) \equiv t - t_0 - iz_R$ is the complex beam parameter, t_0 is the focal time, and $z_R \equiv \frac{1}{2}k_0 w_0^2$ is the Rayleigh range.

To cast this discussion in terms slightly more familiar within the literature, it is helpful to repeat the calculation with z as the propagation variable. In this case, one obtains the following solution for the separable initial condition in Eq. 4.34 with $z \rightarrow t$:

$$A(t, \mathbf{x}) \approx B_0(z - t)C(z, \mathbf{x}_\perp) \quad (4.42)$$

where, for the initially Gaussian transverse profile in Eq. 4.39 at initial position $z = z_0$, $C(z, \mathbf{x}_\perp)$ is given by

$$C(z, \mathbf{x}_\perp) = \left[\frac{iz_R}{q(z)} \right]^{d_\perp/2} \exp \left[-\frac{ik_0 x_\perp^2}{2q(z)} \right] \quad (4.43)$$

and $q(z) \equiv z - z_0 + iz_R$ is the complex beam parameter and z_0 is the focal point. For the Gaussian transverse profile, Eqs. 4.37 and 4.42 are approximately the same if $\tau_0 \ll z_R$, where τ_0 is the duration of the longitudinal envelope $B_0(z - t)$.

What is the timescale over which the second-order approximation supercedes the first-order approximation? One can read off that this is the timescale $t = z_R$ from Eq. 4.37 with a temporal Gaussian beam for the transverse profile. One can also see this directly from the integral in Eq. 4.36. The contribution from each term in the Taylor expansion becomes significant when the term is of order unity because of its appearance as a complex phase. This leads to timescales $\tau_{x_a}, \tau_{x_a x_b}, \dots$ which provide an estimate for when each term leads to significant contributions to the evolution of the envelope. Here we work out this estimate from the $u_\perp^2$ term, which is the timescale $\tau_{xx} = \tau_{yy}$. For an initially Gaussian transverse profile, the maximal value of $u_\perp^2$ that contributes to the integral can be estimated from Eq. 4.40 from

the condition

$$\frac{w_0^2 u_{\perp, \max}^2}{4} \approx 1 \quad (4.44)$$

Hence, the timescale τ_{xx} over which the $\partial_{k_x, y}^2 \omega$ terms lead to corrections from the first-order approximation can be estimated from the condition $2\tau_{xx}/k_0 w_0^2 = 1$. This leads to the estimate

$$\tau_{xx} = \frac{1}{2} k_0 w_0^2 \quad (4.45)$$

which indeed is $\tau_{xx} = z_R$. This same method will be applied to estimate length scales over which the higher-order terms in the expansion contribute in the following sections.

To summarize: in vacuum, the second-order approximation for a laser pulse initially separable with respect to the longitudinal and transverse coordinates takes effect over a timescale $\tau_{xx} = z_R$. Under this approximation, the longitudinal envelope still propagates in the direction of $\mathbf{k}_0$ at the speed of light, while the transverse profile exhibits diffraction.

4.2.3 Third-order propagation: lowest-order spatiotemporal coupling

From Equations 4.26 and 4.27, we have $\partial_{k_z}^3 \omega(\mathbf{k}_0) = 0$ and $\partial_{k_x}^2 \partial_{k_z} \omega(\mathbf{k}_0) = \partial_{k_y}^2 \partial_{k_z} \omega(\mathbf{k}_0) = k_0^{-2}$ in vacuum. All other third-order derivatives vanish. Substituting these derivatives into Eq. 4.12 and accounting for the duplicate terms in the Taylor expansion, the third-order approximation to the envelope is given by

$$A(t, \mathbf{x}) \approx \int \tilde{A}_0(\mathbf{u}) \exp \left[i u_z (z - t) + i \mathbf{u}_{\perp} \cdot \mathbf{x}_{\perp} - \left(\frac{it}{2k_0} \right) u_{\perp}^2 + \left(\frac{it}{2k_0^2} \right) u_{\perp}^2 u_z \right] d\mathbf{u} \quad (4.46)$$

At this order, the integral is no longer analytically tractable (though it can be integrated numerically, as we will see below). Qualitatively, the most notable difference from the second-order approximation is that, because of the third-order $u_{\perp}^2 u_z$ term, the solution is no longer separable in its dependence on z and $\mathbf{x}_{\perp}$. This can be described as the emergence of spatiotemporal coupling on the timescale over which this term becomes of order unity. The same method as in the previous section can be used to estimate this timescale. We assume

the initial longitudinal profile is a Gaussian

$$B_0(z) = \exp \left[-\frac{z^2}{\tau_0^2} \right] \quad (4.47)$$

so that the Fourier transform is

$$\tilde{B}_0(u_z) = \frac{\tau_0}{2\sqrt{\pi}} \exp \left[-\frac{\tau_0^2 u_z^2}{4} \right] \quad (4.48)$$

and thus

$$\frac{\tau_0^2 u_{z,\max}^2}{4} \approx 1. \quad (4.49)$$

Using this and the same estimate for $u_{\perp,\max}$ in Eq. 4.44, we obtain the estimate

$$\tau_{xxz} \equiv \frac{1}{4} \tau_0 k_0^2 w_0^2 \quad (4.50)$$

for the timescale over which spatiotemporal coupling emerges in an initially separable pulse in vacuum. This quantity was referred to as L_{stc} in Ref. [7] and as $L_{s\zeta}$ in Ref. [8] (where in both cases the estimate was a factor of 2 larger). Note that $\tau_{xxz} = \frac{1}{2} N z_R$, where $N = \omega_0 \tau_0$ is approximately the number of cycles in the laser pulse.

4.2.4 Comparison between theory and simulation

Figure 4.1 shows comparisons between theory and simulation that illustrate the key points. Figure 4.1(a) shows $|A|$ for a laser pulse initialized by Eq. 4.37 with $C(z, \mathbf{x}_{\perp})$ given by the spatial Gaussian beam in Eq. 4.43. A sinusoid with finite extent is used for the B_0 function: $B_0(z) = \cos^2(\frac{\pi}{2} \frac{z}{\tau_{\text{rise}}}) I_{[-\tau_{\text{rise}}, \tau_{\text{rise}}]}(z)$, where $I[a, b]$ is the indicator function on the interval $[a, b]$ (Eq. 2.1). These choices are all made for comparison with an OSIRIS simulation mode in which the laser pulse is initialized in space at a fixed time with a finite extent for the laser duration and a spatially dependent Gaussian beam. The duration and spot-size are $\omega_0 \tau_{\text{rise}} = 20$ and $k_0 w_0 = 40$. The laser pulse is initialized at $ct = -5z_R$ at a distance $5z_R$ from its focal point and its profile is calculated at $ct = +5z_R$; this is done so that the amplitude and transverse extent of the pulse are similar at the beginning and end, simplifying visualization.

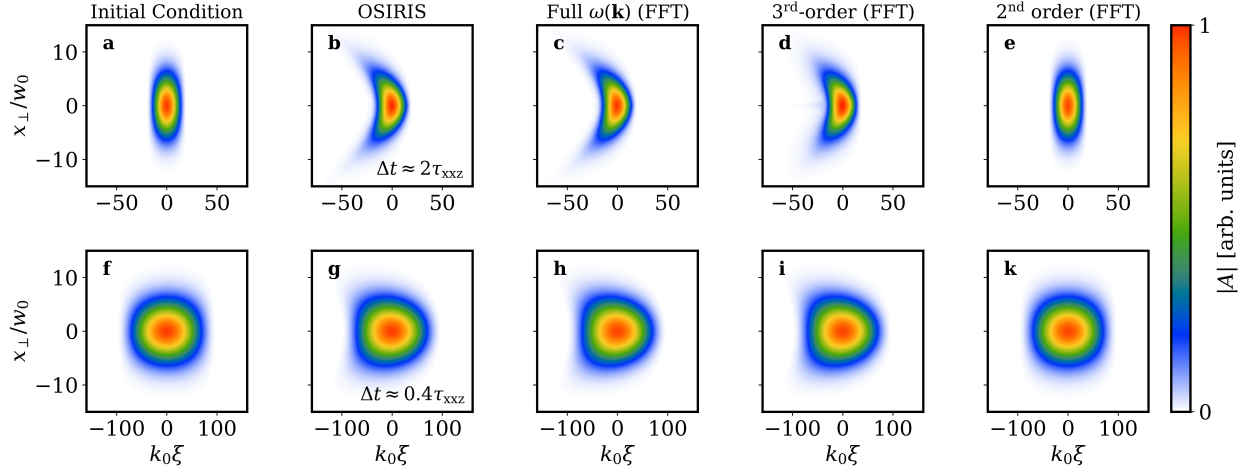


Figure 4.1: Comparison between theory and simulation and illustration of the effect of propagation beyond τ_{xxz} . (a) $|A|$ for the initial condition at $ct = -5z_R$. (b-e) Solutions at $ct = 5z_R$ calculated by simulation in OSIRIS, integrating using the full $\omega(\mathbf{k})$ (Eq. 4.10), integrating the third-order approximation (Eq. 4.46) and integrating using the second-order approximation (Eq. 4.36). Significant curvature in the true solution is consistent with the estimate $\Delta t = 2\tau_{xxz}$. (f-k) Analogous plots for a laser pulse with a duration 5 times larger, so that $\Delta t = 0.4\tau_{xxz}$. In this case, the pulse structure remains approximately separable, consistent with theory.

Figure 4.1(b) shows the result of simulating this profile in OSIRIS in quasi-3D geometry [78] with the solver introduced in Ref. [61] to mitigate numerical dispersion. Bowing of the pulse envelope is evident: the pulse has developed spatiotemporal structure which is not separable in the longitudinal and transverse coordinates. This can be understood from the τ_{xxz} parameter. Estimating $\tau_0 \approx \frac{1}{2}\tau_{\text{rise}}$ for the formula, we may estimate $\tau_{xxz} \approx 5z_R$ so that $\Delta t = 2\tau_{xxz}$. This is consistent with the observation of bowing of the envelope.

Next, Fig. 4.1(c) shows the result of evaluating Eq. 4.10 using the full dispersion relation, i.e., no truncation of the Taylor series, using a fully 3-dimensional FFT. The pulse structure is indistinguishable from the simulated profile obtained from OSIRIS. This agreement corroborates the theoretical formalism. Figure 4.1(d) shows the result of evaluating the third-order

approximation, Eq. 4.46. There are slight differences from the exact solution, especially on the axis, but the qualitative feature of bowing is captured at this order. Figure 4.1(e) shows the result of evaluating the second-order approximation, Eq. 4.36. As expected from Eq. 4.37, the pulse structure remains separable under this approximation and hence $|A|$ has the exact same structure as the initial condition.

Finally, Figs. 4.1(f-k) show the same quantities where the only difference is that the initial pulse duration is $\omega_0\tau_{\text{rise}} = 100$, a factor of 5 times larger. Hence, τ_{xxz} is 5 times smaller, so that $\Delta t \approx 0.4\tau_{xxz}$. In this case, though some curvature is seen in the OSIRIS, full FFT, and 3-order FFT cases, the pulse structure remains approximately separable. This is consistent with τ_{xxz} providing an estimate for the length scale over which curvature becomes prominent.

4.2.5 Geometric understanding of the length-scale for spatiotemporal coupling

The formula for τ_{xxz} (Eq. 4.50) can also be estimated geometrically. This was done in Ref. [7], which is reproduced in Chapter 5. Here a similar derivation is presented based on the results shown in Fig. 4.1. Fig. 4.2 depicts the geometry of the situation. Curvature occurs on the length scale over which $\Delta t \gtrsim \tau_0$. Light rays on-axis and off-axis at the out-of-focus spot size $w \approx w_0 f_0 / z_R$ both propagate a distance ct in a time t . The difference in the distance traveled along the axis can be estimated from the Pythagorean theorem as $c\Delta t \approx ct - \sqrt{c^2 t^2 - (2w_0 f_0 / z_R)^2}$. By design, $f_0 = ct/2$. Substituting this in and Taylor-expanding, one obtains the estimate $c\Delta t \approx 2t/k_0^2 w_0^2$. The timescale τ_{stc} over which spatiotemporal coupling develops can be estimated from the condition that $\Delta t \approx \tau_0$, which leads to the condition $t_{\text{stc}} \approx \frac{1}{2}\tau_0 k_0^2 w_0^2$. This estimate has the same scaling but is a factor of 2 different from the value of τ_{xxz} in Eq. 4.50. This example illustrates the utility of heuristic physical pictures.

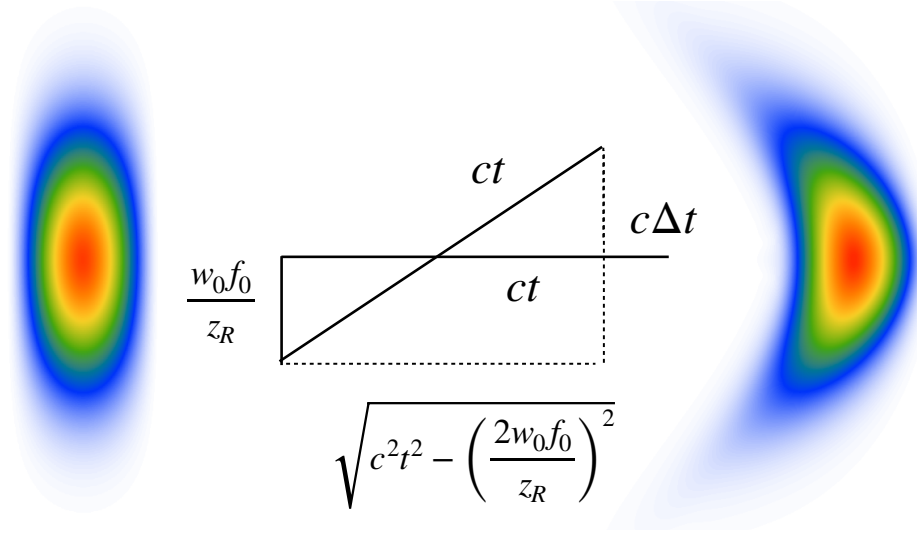


Figure 4.2: Geometry leading to the development of spatiotemporal coupling as evident in Fig. 4.1

4.3 Linear propagation of laser pulses in cold, uniform plasma

The above calculations are now repeated for the case of cold, uniform plasma. For simplicity, the calculations are done in the limit $k_0 \gg k_p$ so that the approximations in Eqs. 4.28-4.32 may be used, but the calculations could also be done without making this approximation.

4.3.1 First-order propagation

The results to first order in plasma are the same as in vacuum except for the different value of $\mathbf{v}_g \approx \left[1 - \frac{1}{2} \left(\frac{k_p}{k_0}\right)^2\right] \hat{\mathbf{z}}$. The envelope evolves as

$$A(t, \mathbf{x}) \approx A_0(z - v_g t, \mathbf{x}_\perp). \quad (4.51)$$

with v_g given above. Similarly, for the separable initial data in Eq. 4.34, we obtain the first-order solution

$$A(t, \mathbf{x}) \approx B_0(z - v_g t) C_0(\mathbf{x}_\perp). \quad (4.52)$$

Compared to vacuum, the only difference is that the pulse propagates with $v_g < c$.

4.3.2 Second-order propagation: group-velocity-delayed paraxial beams and dispersive longitudinal broadening

Including all the derivatives of $\omega(\mathbf{k})$ to second order, one obtains the approximation

$$A(t, \mathbf{x}) \approx \int \tilde{A}_0(\mathbf{u}) \exp \left[iu_z(z - v_g t) + i\mathbf{u}_\perp \cdot \mathbf{x}_\perp - \left(\frac{it}{2k_0} \right) u_\perp^2 - \left(\frac{ik_p^2}{2k_0^3} \right) u_z^2 \right] d\mathbf{u} \quad (4.53)$$

The differences from the vacuum case are (1) that $v_g \neq 1$ and (2) that there is a new term in the exponential proportional to itu_z^2 . Its effect on the pulse's longitudinal evolution is analogous to the effect of the term proportional to $itu_\perp^2$ on the pulse's transverse evolution. Again, it is helpful to consider the example of a pulse which is initially separable in its longitudinal and transverse dependence as in Eq. 4.34. In this case, the envelope is approximated to second order as

$$A(t, \mathbf{x}) = B(t, z)C(t, \mathbf{x}_\perp) \quad (4.54)$$

where

$$B(t, z) \approx \int \tilde{B}_0(u_z) \exp \left[iu_z(z - v_g t) - \left(\frac{ik_p^2}{2k_0^3} \right) u_z^2 \right] du_z \quad (4.55)$$

For example, if the initial longitudinal profile is a gaussian $B_0(z) = \exp(-z^2/\tau_0^2)$, then the integral results in a longitudinal profile which is mathematically the same as a Gaussian beam in two spatial dimensions:

$$B(t, z) = \left[\frac{\tau_0^2}{4} + \frac{ik_p^2 t}{2k_0^3} \right]^{-1/2} \exp \left[-\frac{(z - v_g t)^2}{4 \left[\frac{\tau_0^2}{4} + \frac{ik_p^2 t}{2k_0^3} \right]} \right] \quad (4.56)$$

We see that the longitudinal profile evolves with a length-scale corresponding to the transverse profile's Rayleigh range given by

$$\tau_{zz} \equiv \frac{1}{2} \left(\frac{k_0}{k_p} \right)^2 k_0 \tau_0^2 \quad (4.57)$$

For example, the pulse duration evolves analogously to the laser spot size as

$$\tau(t) = \tau_0 \left[1 + \left(\frac{t}{\tau_{zz}} \right)^2 \right]^{1/2} \quad (4.58)$$

The τ_{zz} timescale can be calculated as in the previous sections by considering when the phase contributed by the term proportional to itu_z^2 becomes of order unity. It is the natural timescale over which the pulse will dispersively broaden longitudinally. Note also that if $t \ll \tau_{zz}$, then $B(t, z) \approx B_0(z - v_g t)$, i.e. the envelope is described by the first-order solution. Aside from these differences, the details of transverse focusing are unmodified by the presence of the plasma: at second order, the transverse evolution is exactly the same as it would be in vacuum (i.e., $C(t, \mathbf{x}_\perp)$ is given by Eq. 4.38). Spatiotemporal coupling is not induced at this order.

4.3.3 Plasma-induced local frequency evolution

It is worth noting that dispersion in cold, uniform plasma gives rise to a variation in the local frequency within a laser pulse. For simplicity, we ignore the transverse evolution (i.e. we assume $w_0 \gg \tau_0$) and consider an envelope given by $A(t, \mathbf{x}) = B(z, t)$ with $B(z, t)$ given by Eq. 4.56. The total envelope phase $\phi = \arg A$ may be extracted as

$$\phi(z, t) = \frac{\xi_g^2 \left(\frac{k_p^2 t}{2k_0^3} \right)}{\left(\frac{\tau^2}{4} \right) + \left(\frac{k_p^2 t}{2k_0^3} \right)^2} \quad (4.59)$$

Therefore, the change in local wavenumber $\Delta k_z(z, t) \equiv \partial_z \phi(z, t)$ is given by:

$$\Delta k_z(z, t) = \frac{2(z - v_g t) \left(\frac{k_p^2 t}{2k_0^3} \right)}{\left(\frac{\tau^2}{4} \right) + \left(\frac{k_p^2 t}{2k_0^3} \right)^2} \quad (4.60)$$

Inspecting this equation, we see that a linear chirp develops which first increases and then decreases in time.

This dispersion-induced chirp can be understood from a single-photon picture. Suppose the pulse is composed of photons of frequency ω , each of which has a dressed velocity $v_g(\omega)$.

After a time t , the ξ coordinate of each photon is given by:

$$\xi_\omega(t) = [v_g(\omega) - 1] t \quad (4.61)$$

$$= -\frac{1}{2} \frac{k_p^2}{\omega^2} t \quad (4.62)$$

$$\approx -\frac{1}{2} \frac{k_p^2}{k_0^2} \left[1 - \frac{\Delta\omega}{k_0} \right] t \quad (4.63)$$

Inverting, we see that the same linear chirp is predicted in the limit of large time, i.e.

$\frac{1}{2} \frac{k_p^2}{k_0^2} t \gg \tau$:

$$\Delta\omega(\xi) = \frac{z - v_g t}{\frac{1}{2} \frac{k_p^2}{k_0^3} t} \quad (4.64)$$

This is consistent with the analytically derived chirp. This is another example which illustrates the utility of heuristic physical pictures.

4.3.4 Third-order propagation: vacuum-like spatiotemporal coupling and additional longitudinal broadening

Substituting all of Eqs. 4.23-4.27 into the envelope expansion, we obtain the following approximation at third order in cold, uniform plasma:

$$A(t, \mathbf{x}) \approx \int \tilde{A}_0(\mathbf{u}) \exp \left[i u_z (z - v_g t) + i \mathbf{u}_\perp \cdot \mathbf{x}_\perp - \left(\frac{it}{2k_0} \right) u_\perp^2 - \left(\frac{ik_p^2 t}{2k_0^3} \right) u_z^2 \right. \\ \left. + \left(\frac{it}{2k_0^2} \right) u_\perp^2 u_z + \left(\frac{ik_p^2 t}{2k_0^4} \right) u_z^3 \right] d\mathbf{u} \quad (4.65)$$

The $(it/2k_0^2)u_\perp^2 u_z$ term is unchanged from vacuum. Thus, as in the vacuum case, the integral is no longer tractable for a separable initial condition. All the discussion on the length scale for the emergence of spatiotemporal coupling and the estimate for the τ_{xxz} parameter apply as in the vacuum case. The term proportional to itu_z^3 is an additional correction to pulse-broadening.

4.3.5 Comparison between theory and simulation

As in the vacuum case, the theory in this section can be compared directly with simulation. Figure 4.3 shows simulations and calculations for the same initial conditions and propagation time as in Fig. 4.1 but with plasma instead of vacuum. The frequency ratio is $k_0/k_p = 10$. The differences in the results are attributable to the presence of plasma. The values of τ_{xxz} are unmodified from the vacuum values, while τ_{zz} may be approximated for upper row as $\tau_{zz} \approx 6.25z_R$ so that $\Delta t \approx 1.6\tau_{zz}$. Because Δt is larger than τ_{zz} , effects of longitudinal broadening due to the plasma are expected. This is indeed observed. The bow structure for the vacuum case in Fig. 4.1(b) becomes a triangular structure in the case with plasma, as seen in Fig. 4.3(b). Unlike in the vacuum case, the pulse slips backward with respect to the variable ξ because $v_g < 1$. As in the vacuum case, Fig. 4.3(c) shows that the FFT integration using the full $\omega(\mathbf{k})$ agrees well with OSIRIS, corroborating the theory. The third-order approximation shown in Fig. 4.3 qualitatively captures the triangular structure observed in the full solution. The second-order approximation in Fig. 4.3 does not capture the triangular structure because the solution remains separable, but does capture the longitudinal broadening.

For the laser in Fig. 4.3(f-k), τ_{zz} is 5^2 times smaller, so $\Delta t \approx 0.06\tau_{zz}$. Hence effects of plasma-induced dispersion are expected to be negligible. This is indeed observed in Fig. 4.3(f-k), which are nearly indistinguishable from the vacuum cases (Fig. 4.1(f-k)) except that the pulse recedes with respect to ξ because $v_g < 1$.

4.4 Recovery of Envelope PDEs from the Integral Formulation

One can generate a PDE for the evolution of $\hat{A}$ by taking taking derivatives of Equation 4.12. Applying ∂_t , and applying spatial derivatives to match the result, we obtain

$$\partial_t A + \partial_{k_a} \omega(\mathbf{k}_0) \partial_{x_a} A - \frac{i}{2!} \partial_{k_a} \partial_{k_b} \omega(\mathbf{k}_0) \partial_{x_a} \partial_{x_b} A + \dots = 0 \quad (4.66)$$

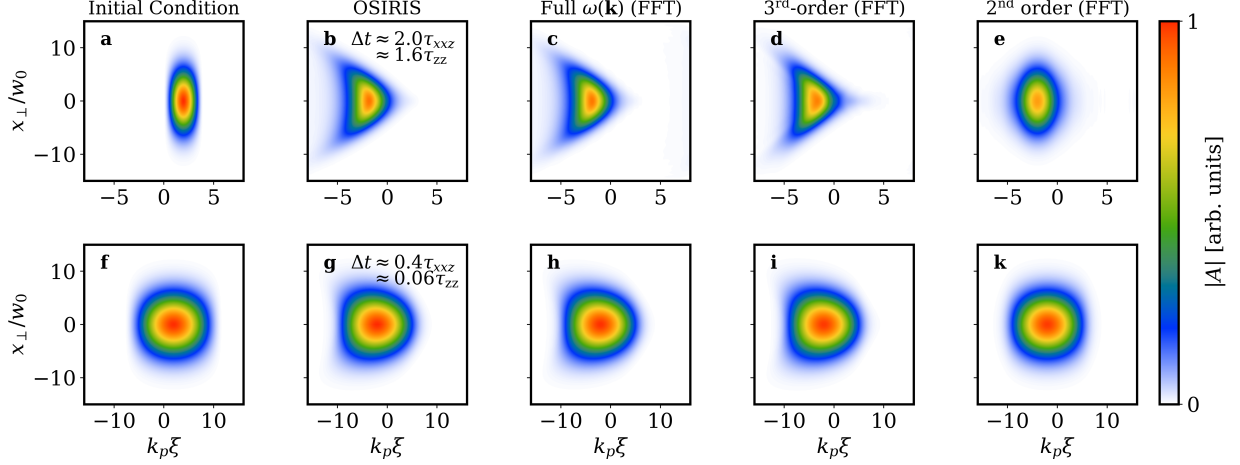


Figure 4.3: Demonstration of spatiotemporal coupling and additional longitudinal broadening induced by propagation in plasma in 3D. The initial condition and propagation length are the same as in the vacuum cases shown Fig. 4.1, but now a plasma is added such that $k_0/k_p = 10$. ($k_p = 0$ in Fig. 4.1). (a–e) The effects of plasma-induced dispersion manifest as broadening of the longitudinal envelope such that the bow structure in vacuum becomes a triangular structure in plasma. The agreement between OSIRIS and the FFT with the full $\omega(\mathbf{k})$ supports the theory, as in the vacuum case. (f–k) In this case plasma-induced dispersion is negligible; the evolution looks just as in vacuum except for the recession of the pulse with respect to ξ because $v_g < 1$.

This equation is exact, but can be truncated for an approximate equation. For example, for evolution in cold, uniform plasma, this equation at second order for $k_0/k_p \gg 1$ is

$$\partial_t A + \left(1 - \frac{1}{2} \frac{k_p^2}{k_0^2}\right) \partial_z A - \frac{i}{2\omega_0} \nabla_\perp^2 A - \left(\frac{i}{2} \frac{k_p^2}{k_0^3}\right) \partial_z^2 A \approx 0 \quad (4.67)$$

In practice, this differential approach is impractical because it assumes the accuracy of the linear solutions; but under this condition, we can fully write out the integral in Eq. 4.12. On the other hand, the differential approach is useful conceptually for comparison with the differential approach in nonlinear cases (see Eq. 2.43, Ref. [67]). In such cases, it may be possible to formulate a PDE without being able to write down analytic integral solutions.

4.5 Summary

In summary, this chapter contains an integral approach to understanding the linear propagation of laser pulses in plasma. The approach is based on material presented in Siegman's *Lasers* [135] and Zangwill's *Modern Electrodynamics* [136]. The value of this approach is its treatment of the full spatiotemporal evolution of the laser pulse. It provides a clear description of the evolution of laser pulses initialized as $\mathbf{E}_\perp(0, z, \mathbf{x}_\perp) = \hat{E}_\parallel(z)C(z, \mathbf{x}_\perp)e^{ik_0z}\hat{\mathbf{e}}$. It is shown that the full evolution is approximately $\mathbf{E}(t, z, \mathbf{x}_\perp) \approx \hat{E}_\parallel(z - v_g t)C(z, \mathbf{x}_\perp)e^{i(k_0z - \omega_0 t)}\hat{\mathbf{e}}$, where $\omega_0 \equiv \omega(k_0\hat{\mathbf{z}})$. The length scale over which the estimate remains accurate is easily derived from this framework. Higher-order effects can also naturally be quantified. Simulations in vacuum and plasma confirm and demonstrate the results of the theory.

CHAPTER 5

Arbitrarily Structured Laser Pulses

Spatiotemporal control refers to a class of optical techniques for structuring a laser pulse with coupled space-time dependent properties, including moving focal points, dynamic spot sizes, and evolving orbital angular momenta. This chapter fleshes out the concept of arbitrarily structured laser (ASTRL) pulses which generalizes these techniques. The ASTRL formalism employs a superposition of prescribed pulses to create a desired electromagnetic field structure. Several examples in this chapter illustrate the versatility of ASTRL pulses to address a broad range of laser-based applications, including laser wakefield acceleration, inertial confinement fusion, nanophotonics, and attosecond physics.

A historical background on the development of ASTRL pulses in the context of this dissertation is presented in Chapter 1. Most of this chapter comes from Ref. [7]. The latter is distributed under the terms of the Creative Commons CC BY 4.0 license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The only substantially new part is Sec. 5.4, discussing a specific class of closed-form ASTRL pulses discovered after publication of Ref. [7]. A few modifications were made in the other sections in reference to these solutions. The appendices of the paper were also moved into the body of the chapter. Most of the theory in this chapter relating to the structure of approximately separable laser-field solutions of the form in Eq. 5.6 and the corresponding length scale L_{stc} (Eq. 5.20) is redundant with the calculations in Chapter 4. For the sake of simplicity, I have not attempted to connect them. The main difference is that Chapter 4 takes an integral perspective while Chapters 5 and 6 take a differential perspective.

The notation in this chapter – foremost, the use of the variable A instead of $\hat{E}$ to describe the laser envelope (e.g., Eq. 5.3) – is slightly different from the notation used in the other chapters of this thesis. The variable A is a poor choice because it collides with the magnitude of the vector potential $\mathbf{A}$. However, I have decided to keep the original notation in order to maintain consistency with the original article, avoid introducing errors, and prevent the need to recreate the plots. I apologize to anyone whom this confuses.

5.1 Introduction

Light can exhibit structures in space, time, or coupled space-time. The realization that these structures can be shaped using optical techniques has led to a deeper understanding of the fundamental properties of light and advances in multiple laser-based applications. Spatial shaping involves the use of static optical elements, such as phase plates, deformable mirrors, or spatial light modulators, to form transverse structures with orbital angular momentum, Airy or Bessel profiles, or tailored speckle patterns [137, 138, 139, 140]. Temporal shaping typically employs Fourier methods, in which frequency components are separated by diffraction gratings and separately modified, to synthesize pulses with nearly arbitrary time-dependent waveforms and polarizations [141, 142].

Spatiotemporal pulse shaping invokes some combination of these techniques. Examples go back as far as smoothing by spectral dispersion (1989) [143] and pulse front tilt (1996) [144]. More modern examples include spatiotemporal optical vortices (STOVs) [145, 146, 147], diffraction-free light sheets [148, 149, 150], and flying focus concepts [3, 151, 152, 153]. Flying focus pulses, in particular, feature intensity peaks with programmable trajectories that can travel distances far greater than a Rayleigh range while maintaining a near-constant profile. The trajectory control afforded by these pulses has been proposed as a means to overcome limitations of traditional pulses in a number of applications, including laser wakefield acceleration [37, 38], Raman amplification [66], and nonlinear Thomson scattering [71], as well

as for the exploration of fundamental physics, such as radiation reaction [72] and nonlinear Compton scattering [154].

Spatiotemporally structured pulses may be created using either a forward- or inverse-design approach. The forward-design approach requires identifying a specific solution to Maxwell’s equations. Once known, this solution can be synthesized using programmable phase modulation techniques [155, 156]. This is the approach taken to create light sheets, STOVs, and pulses with dynamic orbital angular momenta. However, this approach does not provide a method to systematically identify new solutions of Maxwell’s equations with features that are desirable for a particular application. In the inverse-design approach, one imagines an electromagnetic structure with desirable features and synthesizes this structure using a superposition of solutions to Maxwell’s equations with known properties. This is the approach presented here.

This chapter reproduces the concept and theoretical formalism for generating laser pulses with arbitrary structure introduced in Ref. [7]. These arbitrarily structured laser (ASTRL) pulses are synthesized using a superposition of traditional laser pulses with controlled and varying properties. The ASTRL formalism generalizes the creation of pulses with evolving focal points (flying foci), spot size, orbital angular momentum, and polarization, and can also describe pulses with exotic topological structure such as STOVs. An example of each of these is presented. The flexibility of the ASTRL concept opens new and previously unimagined possibilities for structured light with the potential to improve a wide range of laser-based applications, while its simplicity facilitates implementation into simulations.

5.2 Theoretical formalism

The formulation of the ASTRL concept begins with the vacuum wave equation for the electric field of a laser pulse:

$$\left[\nabla^2 - \frac{1}{c^2} \partial_t^2 \right] \mathbf{E}(\mathbf{x}, t) = 0. \quad (5.1)$$

ASTRL pulses are constructed by superposing solutions to Eq. 5.1 with varying properties. The electric field of an ASTRL pulse can be expressed either as a sum or an integral over these solutions:

$$\mathbf{E}(\mathbf{x}, t) = \int d\eta \mathbf{E}_\eta(\mathbf{x}, t), \quad (5.2)$$

where η parameterizes the varying properties of the solutions. For example, the amplitude, spot size, focal point, pulse duration, or relative delay of each $\mathbf{E}_\eta$ can all depend on η . Equation 5.2 is the most general representation of an ASTRL pulse in vacuum.

Analytic solutions for $\mathbf{E}_\eta$ simplify the process of designing an ASTRL pulse for a specific application. Paraxial solutions, in particular, are expressed in terms of familiar quantities that can be parameterized in terms of η . To derive such solutions, consider a laser pulse propagating in the $\hat{z}$ -direction with constant polarization $\hat{\mathbf{e}}$. The transverse electric field can be expressed as an envelope $A(\mathbf{x}_\perp, z, t)$ (not to be confused with the vector potential) modulated by a carrier

$$\mathbf{E}_\perp(\mathbf{x}_\perp, z, t) = \frac{1}{2}A(\mathbf{x}_\perp, z, t)e^{i(k_0 z - \omega_0 t)}\hat{\mathbf{e}} + \text{c.c.}, \quad (5.3)$$

where k_0 is the central wavenumber and $\omega_0 = ck_0$. Substituting Eq. 5.3 into Eq. 5.1 and performing the Galilean change of variables $(\xi, s) = (z - ct, z)$ provides

$$[\nabla_\perp^2 + \partial_s^2 + 2ik_0\partial_s + 2\partial_\xi\partial_s] A(\mathbf{x}_\perp, s, \xi) = 0. \quad (5.4)$$

Upon applying the slowly-varying envelope approximation ($|\partial_\xi|, |\partial_s| \ll k_0$), the mixed derivative $(\partial_\xi\partial_s)$ and non-paraxial (∂_s^2) terms can be dropped and Eq. 5.4 reduces to

$$[\nabla_\perp^2 + 2ik_0\partial_s] A(\mathbf{x}_\perp, s, \xi) \approx 0. \quad (5.5)$$

Equation 5.5 is the paraxial wave equation and its operator is independent of ξ . This independence admits separable solutions, i.e., solutions with no spatiotemporal coupling, of the form

$$A(\mathbf{x}_\perp, s, \xi) \approx B(\xi)C(\mathbf{x}_\perp, s), \quad (5.6)$$

where $C(\mathbf{x}_\perp, s)$ satisfies $[\nabla_\perp^2 + 2ik_0\partial_s]C(\mathbf{x}_\perp, s) = 0$ and $B(\xi)$ is an arbitrary function. For example, $C(\mathbf{x}_\perp, s)$ may be any Laguerre-, Hermite-, or Ince-Gaussian mode and the longitudinal profile may be $B(\xi) = \exp[-(\xi/c\tau_0)^2]$. The validity of the approximations leading to Eq. 5.6 is addressed below in Sec. 5.7.

The transverse electric field of a coherent ASTRL pulse can be approximated by summing or integrating over paraxial solutions:

$$\mathbf{E}_\perp(\mathbf{x}_\perp, s, \xi) \approx \frac{1}{2}e^{ik_0\xi} \int d\eta B_\eta(\xi)C_\eta(\mathbf{x}_\perp, s)\hat{\mathbf{e}}(\eta) + \text{c.c.}, \quad (5.7)$$

where the properties of $B_\eta(\xi)$, $C_\eta(\mathbf{x}_\perp, s)$, and $\hat{\mathbf{e}}(\eta)$ may vary with respect to η . The integral over η may introduce spatiotemporal coupling, such that in general, Eq. 5.7 cannot be written as a separable function. Thus, $\mathbf{E}_\perp$ may possess a high degree of spatiotemporal coupling despite the absence of coupling in Eq. 5.6.

The η -dependence of $B_\eta(\xi)$ may encode a delay $\Delta(\eta)$, duration $\tau_0(\eta)$, and weight $B_0(\eta)$ for each constituent pulse as

$$B_\eta(\xi) = B_0(\eta) \exp \left[- \left(\frac{\xi - \Delta(\eta)}{c\tau_0(\eta)} \right)^2 \right]. \quad (5.8)$$

The durations $\tau_0(\eta)$ set the fastest timescale over which the properties of the ASTRL pulse can vary and determine the number of pulses required for a discrete version of Eq. 5.7. Unless otherwise stated, τ_0 and B_0 will be independent of η for the examples presented here. In principle, each pulse can have a different phase (potentially random) contained in $B_0(\eta)$; for all the examples considered here, $B_0(\eta)$ is real so that the pulses add coherently.

A natural example for the η -dependence of $C_\eta(\mathbf{x}_\perp, s)$ is an η -indexed Gaussian-beam:

$$C_\eta(\mathbf{x}_\perp, s) = \left[\frac{z_R(\eta)}{q_\eta(s)} \right]^{d_\perp/2} \exp \left[- \frac{ik_0|\mathbf{x}_\perp - \mathbf{x}_{\perp 0}(\eta)|^2}{2q_\eta(s)} \right], \quad (5.9)$$

where $q_\eta(s) \equiv s - s_0(\eta) + iz_R(\eta)$ is the complex beam parameter and $d_\perp$ is the number of transverse dimensions. In Eq. 5.9, the longitudinal focal position $s_0(\eta)$ (in $q_\eta(s)$), transverse focal position $\mathbf{x}_{\perp 0}(\eta)$, and Rayleigh range $z_R(\eta)$ (and associated spot size $w_0(\eta) = \sqrt{2z_R(\eta)/k_0}$)

may all be expressed as functions of the integration variable η , which allows for extended focal ranges and evolving transverse structure. More generally, the mode numbers of the transverse profile can depend on η , e.g., the radial or orbital angular momentum (OAM) quantum numbers of a Laguerre-Gaussian mode. Finally, $\hat{\epsilon}(\eta)$ may encode an evolving polarization.

5.3 Design of ASTRL pulses

The utility and generality of the ASTRL concept will be demonstrated by using Eq. 5.7 to construct several examples of spatiotemporally structured laser pulses. To begin, consider the flying focus [151, 3]. Each time slice in a flying focus pulse has a different focal point along the propagation axis. The time and location at which each slice comes to focus can be controlled to produce an intensity peak that travels at any velocity, including backwards or faster than the speed of light. Within the ASTRL formalism, a flying focus pulse may be described using a Gaussian transverse profile (Eq. 5.9) with $\eta \in [-\frac{1}{2}, \frac{1}{2}]$ and $s_0(\eta) = \eta L$. This choice of $s_0(\eta)$ produces a focal range extending from $s = -L/2$ to $s = L/2$, where L can be much larger than z_R . At each s in the focal range, the focus will occur at a time determined by the condition $\xi = \Delta(\eta) = \Delta(s/L)$. The delay $\Delta(\eta)$ required for a given focal trajectory can be derived from geometric optics as in Ref. [37]. For example, motion of the focal point at a constant velocity $v_f = c\beta_f$ is attained with $\Delta(\eta) = (1 - 1/\beta_f)L\eta$.

Figure 5.1(a) displays the on-axis envelope of a superluminal ($\beta_f = 1.05$) flying focus pulse obtained by evaluating Eq. 5.7. The white dashed line shows the spacetime trajectory $z = v_f t$ and demonstrates that the laser envelope travels at the expected velocity over 10 Rayleigh ranges. A finite-difference time-domain (FDTD) simulation of the full set of Maxwell's equations produces nearly identical results [Fig. 5.1(b)]. The FDTD simulation used Eq. 5.7 to initialize the fields, but did not make the envelope approximations leading to Eq. 5.5. A slight discrepancy can be observed near $ct \approx 5z_R$ due to the inexactness of Eq. 5.6. The simulation was conducted using the code OSIRIS [2] with quasi-3D azimuthal

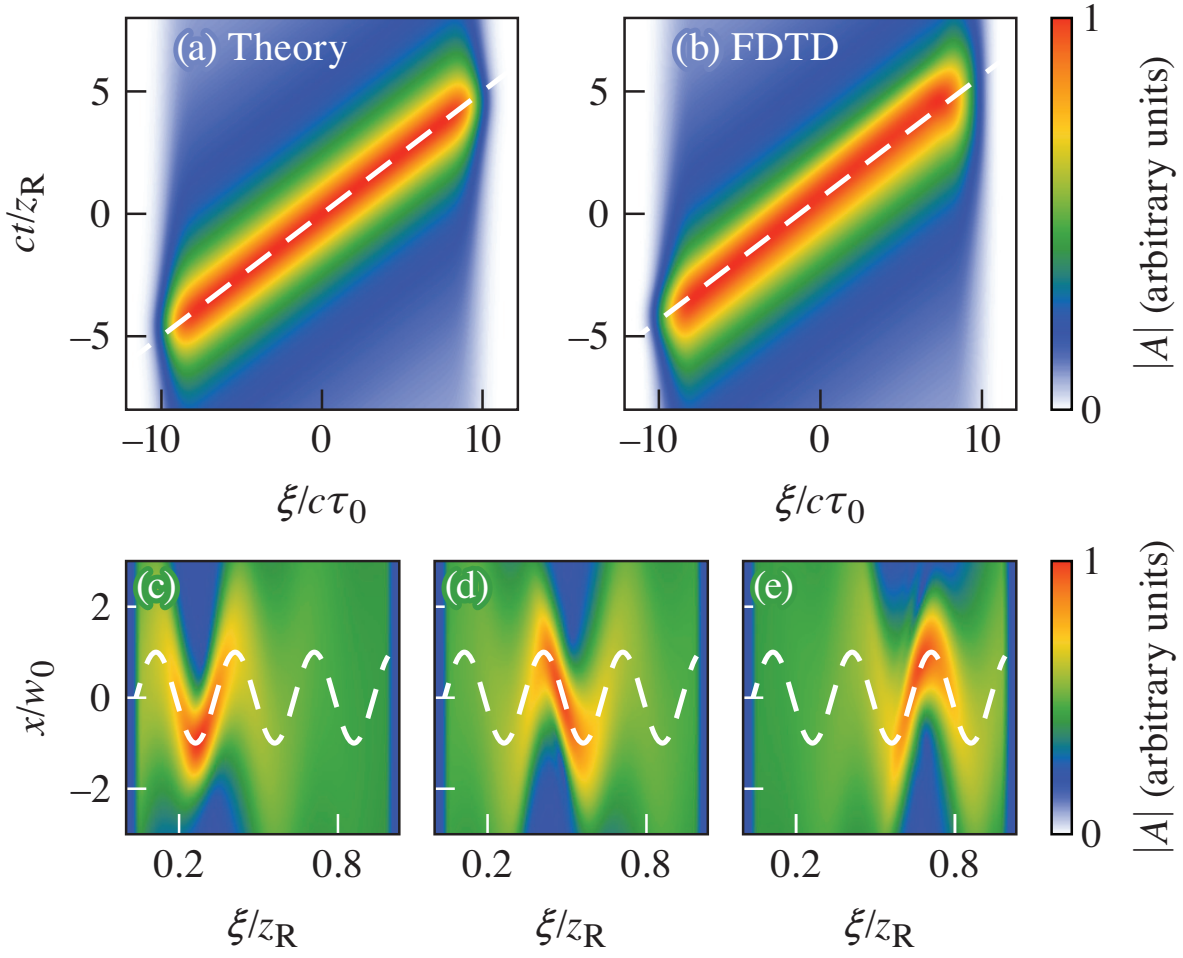


Figure 5.1: (a): On-axis evolution of the envelope of an ASTRL flying focus pulse found by evaluating Eq. 5.7 with $k_0 w_0 = 40$, $\omega_0 \tau_0 = 20$, and $\beta_f = 1.05$. The dashed line indicates the trajectory $z = v_f t$. (b) The envelope from a quasi-3D FDTD simulation for the same parameters. The two are in excellent agreement. (c) – (e): A generalized flying focus constructed using the ASTRL formalism in which the focal point oscillates transversely as it moves along the optical axis. The propagation was simulated in 2D using the FDTD method. The pulse was initialized at $ct/z_R = -10$ using Eq. 5.7 with $x_{\perp \text{osc}} = w_0$, $k_{\text{osc}} = 1/z_R$, $\beta_f = 1.05$, $L = 20z_R$, $k_0 w_0 = 40$, and $\omega_0 \tau_0 = 20$. From left to right, the frames show the pulse at $ct/z_R = -4$, 0, and 4.

decomposition [78] and the solver introduced in Ref. [61] to mitigate numerical dispersion.

The ASTRL concept can also be used to describe new types of generalized flying focus pulses. For instance, a pulse whose focal point oscillates in the transverse direction as it translates along the propagation axis could serve as a controllable wiggler for generating radiation from relativistic electron bunches or enhance direct laser acceleration of electrons through new parametric resonances. Such a pulse can be constructed by parameterizing both the transverse and longitudinal focal coordinates: $\mathbf{x}_{\perp 0}(\eta) = x_{\perp \text{osc}} \sin[k_{\text{osc}} s_0(\eta)] \hat{\mathbf{x}}$ and $s_0(\eta) = \eta L$, where, as before, $\eta \in [-\frac{1}{2}, \frac{1}{2}]$. Figures 5.1(c)–(e) display the envelope from 2D OSIRIS simulations of this configuration. The intensity peak undergoes the expected oscillatory motion. Here, C_η is given by Eq. 5.9 and $\Delta(\eta) = (1 - 1/\beta_f)L\eta$, i.e., the focal point moves with a constant axial velocity.

Aside from moving focal points, many applications can benefit from pulses that have a fixed focal point with prescribed spatiotemporal profiles or ξ -dependent polarization (Fig. 5.2). Traditional Fourier techniques for temporal pulse shaping, as in Ref. [142], produce pulses described by Eq. 5.6 with a customized $B(\xi)$. Similarly, previous polarization-structuring methods amount to evaluating Eq. 5.7 with a prescribed $\hat{\epsilon}(\eta)$ and B_η , but with a C_η that is independent of η . Figure 5.2(a) shows such an example. This particular ASTRL pulse is constructed from a discrete sum of seven pulses, indexed by $\eta \in \{0, 1, \dots, 6\}$, that share a common focal point $s_0(\eta) = 0$ and constant Rayleigh range $z_R(\eta)$. The delay and polarization of each pulse are given by $\Delta_\eta = -2c\tau_0\eta$ and $\hat{\epsilon}(\eta) = [\cos(\pi\eta/5), \sin(\pi\eta/5), 0]$ while C_η is given by Eq. 5.9. Evolving polarization structures as in this example have led to advances in nonlinear spectroscopy [157], quantum control [158, 159], nanophotonics [160], and a number of other areas [161, 162]. More exotic pulses with continuous polarization evolution can also be described using the ASTRL concept.

A time-dependent spot size can improve the performance of applications including laser wakefield acceleration [1, 163] and inertial confinement fusion [164, 165, 166]. In laser wakefield acceleration, a spot size that decreases in ξ (i.e., increases in time at a fixed z), can stabilize

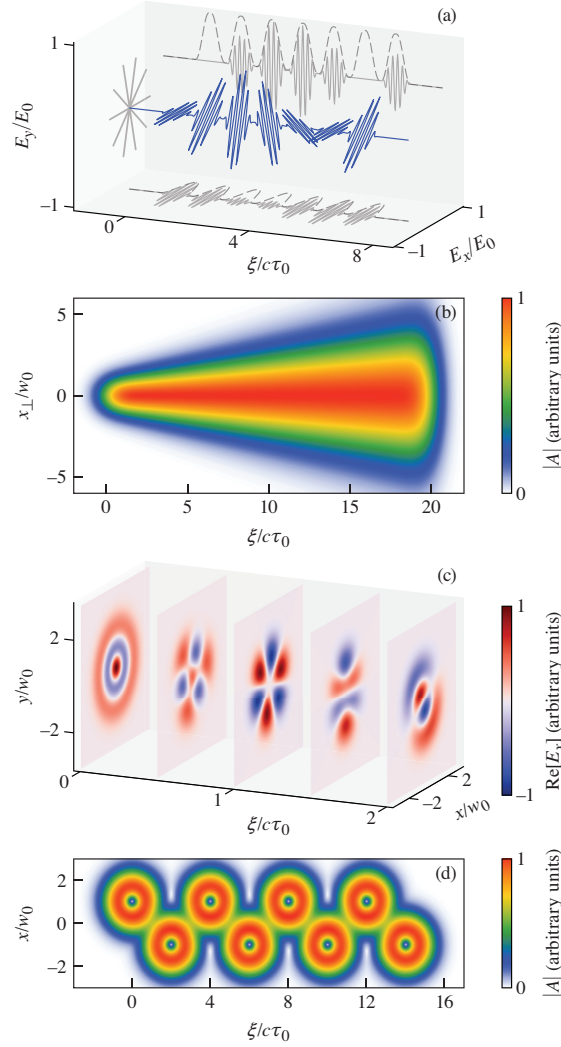


Figure 5.2: Examples of ASTRL pulses constructed from superpositions of laser pulses focused to the same point ($s_0(\eta) = 0$): (a) A pulse everywhere linearly polarized, but with a time-dependent polarization angle. The duration of each pulse in the superposition is $\omega_0\tau_0 = 20$. (b) A laser pulse with a spot size that continuously decreases in time with $k_0w_{0\min} = 40$, $w_{0\max} = 4w_{0\min}$, $\omega_0\tau_0 = 40$, and $\Delta_{\max} = 20c\tau_0$. (c) Time slices of the transverse electric field of a laser pulse whose angular momentum continuously varies in ξ . (d) A lattice of STOVs with $k_0w_0 = 40$ and $\omega_0\tau_0 = 20$. All examples are evaluated using Eq. 5.7 and shown at the focal point $s = 0$.

the propagation of an intense laser pulse in a plasma channel [163] while a spot size that increases in ξ can potentially increase the self-guiding distance (see Refs. [1, 134]). In inertial confinement fusion, a spot size that decreases continuously in ξ can mitigate cross-beam energy transfer (this effect has been called “zooming”) [164, 165, 166]. As demonstrated in Fig. 5.2(b), a pulse with a decreasing spot size in ξ can be described by Eqs. 5.7, 5.8, and 5.9 with a spot size parameterized as $w_0(\eta) = w_{0\min} + \eta(w_{0\max} - w_{0\min})$ and the delay $\Delta(\eta) = -\Delta_{\max}\eta$, where $\eta \in [0, 1]$.

The orbital angular momentum and radial mode numbers (ℓ and p , respectively) of Laguerre-Gaussian modes provide additional degrees of freedom to structure the time-dependence of the transverse profile. Figure 5.2(c) displays a sequence of time slices in an ASTRL pulse formed by superposing three pulses that partially overlap in time with $(\ell, p) = (0, 2), (3, 0), (1, 1)$. Each pulse is indexed by $\eta \in \{0, 1, 2\}$, has a delay $\Delta_\eta = -2c\tau_0\eta$, and a weight $B_0(\eta) = 1, 0.8, 1$. This construction produces a continuous transformation in angular momentum. In the first, third, and fifth frames, one pulse dominates, and the transverse profile is nearly a pure Laguerre-Gaussian mode; in the second and fourth frames, two pulses interfere to produce a hybrid mode. Investigations into structured OAM have only recently begun [155, 156, 167, 168, 169], but its potential for ultrafast probing of chiral systems has already drawn attention [169]. Many other applications based on OAM [170], including optical trapping and manipulation [171, 172, 173, 174], imaging [175, 176, 177], quantum optics [178, 179], nonlinear optics [180], and laser-plasma interactions [34, 181, 182, 183, 184, 185, 186], may also benefit from added control over its structure.

As a final example, the ASTRL concept can be used to construct STOVs. In fact, the vacuum STOV solution introduced in Ref. [145],

$$A(\mathbf{x}_\perp, s, \xi) = \left[\frac{\xi}{c\tau_0} \pm \frac{ix}{w(s)} \left(\frac{-q^*(s)}{q(s)} \right)^{1/2} \right] \frac{z_R}{q(s)} \exp \left[-\frac{ik_0 x_\perp^2}{2q(s)} \right] \exp \left[-\left(\frac{\xi}{c\tau_0} \right)^2 \right], \quad (5.10)$$

is a superposition of two separable solutions as in Eq. 5.6, where the C_η functions are Hermite-Gaussian modes of order $(0, 0)$ and $(1, 0)$. Here the s -dependence has been included,

$w(s) \equiv w_0 \sqrt{1 + (s/z_R)^2}$, and $q^*(s)$ is the complex conjugate of $q(s)$. The generalization of Eq. 5.10 to higher-order vortex topologies (see Ref. [147]) can also be expressed as a superposition of separable solutions. Figure 5.2(d) illustrates a novel STOV “lattice” assembled by superposing solutions in the form of Eq. 5.10. The solutions are indexed by $\eta \in \{0, 1, \dots, 7\}$ with delays $\Delta_\eta = -2c\tau_0\eta$, transverse displacements alternating between $x_0 = \pm w_0$, and polarities (the sign in Eq. 5.10) alternating between ± 1 .

5.4 Closed-form ASTRL pulses

In addition to superposing laser pulse solutions with varying properties as in Eq. 5.7, there is another way to generate approximate solutions for spatiotemporally structured laser fields from Eq. 5.5. Suppose $C_\alpha(\mathbf{x}_\perp, s)$ are solutions to the paraxial wave equation $[\nabla_\perp^2 + 2ik_0\partial_s]C_\alpha(\mathbf{x}_\perp, s) = 0$, where a multi-index α parameterizes properties of C_α – for example, for a Gaussian beam, $\alpha = (w_0, s_0, \mathbf{x}_{\perp 0})$ fully parametrizes the solution. The key idea is that because the differential operator in Equation 5.5 is independent of ξ , α itself may depend on ξ so long as the approximations leading to Eq. 5.5 still hold. More explicitly,

$$A(\mathbf{x}_\perp, s, \xi) \approx C_{\alpha(\xi)}(\mathbf{x}_\perp, s) \quad (5.11)$$

is an approximate solution of Eq. 5.4. For the solutions in Eq. 5.11, the transverse profile is a solution of the paraxial wave equation at a fixed ξ , but the transverse profile may vary continuously as a function of ξ . A special case of Eq. 5.11 was used to model a flying focus in Ref. [70]. As an example, consider the Gaussian beam $C_\eta(\mathbf{x}_\perp, s)$ from Eq. 5.9 with $\eta \rightarrow \alpha$. Letting $\alpha(\xi) = (w_0(\xi), s_0(\xi), \mathbf{x}_{\perp 0}(\xi))$ vary with ξ , we obtain the following approximate solution of Eq. 5.4:

$$A(\mathbf{x}_\perp, s, \xi) \approx a_0(\xi) \left[\frac{z_R(\xi)}{q(s, \xi)} \right]^{d_\perp/2} \exp \left[-\frac{ik_0 |\mathbf{x}_\perp - \mathbf{x}_{\perp 0}(\xi)|^2}{2q(s, \xi)} \right], \quad (5.12)$$

where the complex beam parameter $q(s, \xi) \equiv s - s_0(\xi) + iz_R(\xi)$ now depends explicitly on both ξ and s and the complex amplitude $a_0(\xi)$ only depends on ξ .

Like Eq. 5.7 with the choices $B_\eta(\xi)$ and $C_\eta(\mathbf{x}_\perp, s)$ given in Eq. 5.8 and Eq. 5.9, Eq. 5.12 can be used to generate the fields of a constant-velocity flying focus pulse. The focal range and velocity are determined by $a_0(\xi)$ and $s_0(\xi)$. It is illustrative to consider a constant amplitude profile $a_0(\xi) = a_0$, constant minimum spot size $w_0(\xi) = w_0$, constant transverse offset $\mathbf{x}_\perp(\xi) = \mathbf{0}$, and a linear parameterization of the focal length $s_0(\xi) = \alpha\xi$. For these choices, $|A|^2$ can be expressed as

$$|A(\mathbf{x}_\perp, s, \xi)|^2 = a_0^2 \left[\frac{w_0}{w(s, \xi)} \right]^{d_\perp} \exp \left[-\frac{2x_\perp^2}{w(s, \xi)^2} \right] \quad (5.13)$$

where

$$w(s, \xi) \equiv w_0 \left[\frac{(s - s_0(\xi))^2}{z_R^2} + 1 \right]^{1/2} \quad (5.14)$$

is the spot size of each slice. For a fixed s , $|A|^2$ is peaked on-axis at the value of ξ where $w(s, \xi)$ is minimal (i.e., the point of smallest spot size where the pulse is in focus). This occurs at a value $\xi = \xi_p$ satisfying $s = s_0(\xi_p)$, where the minimal spot size $w(s, \xi_p) = w_0$ is attained. In this case, $\xi_p(s) = \alpha^{-1}s$. Thus, the ξ coordinate of the intensity peak varies linearly with s . The intensity peak follows the spacetime trajectory $s = s_0(\xi)$, so in (z, t) coordinates, the trajectory is $z(t) = \alpha[z(t) - t]$. Solving for $z(t)$, we obtain $z(t) = [\alpha/(\alpha - 1)]t$, which tells us that the normalized velocity of the intensity peak in vacuum is $\beta_f = \alpha/(\alpha - 1)$. Equivalently, to choose α for a desired β_f , we have $\alpha = \beta_f/(\beta_f - 1)$. This parameter is the same as the ratio $s_0(\eta)/\Delta(\eta)$ in Section 5.3. Note that if $a_0(\xi)$ varies (and hence also the power $P(\xi)$), then $s = s_0(\xi_p)$ no longer holds. In this case, the intensity peak does not occur at the slice which is in focus.

What is the significance of the structured laser field in Eq. 5.11? Equation 5.11 and Eq. 5.7 are approximately equal for certain choices of the B and C functions. In this circumstance, Eq. 5.11 is the more natural choice to use for calculations (both theoretical and computational) because it is simpler. No integral needs to be evaluated. For example, in the case of a flying focus with constant velocity described by Eq. 5.7, at any ξ value, the pulses with peaks ξ_0 satisfying $|\xi - \xi_0| \lesssim c\tau_0$ dominate the integral. Thus, at any ξ value, there is a spread of

peak values of $\Delta\xi_0 \approx c\tau_0$. The spread in focal points contributing to the integral from the C_η factor is then given by $\Delta s_0 \approx \alpha\Delta\xi_0 = [\beta_f/(\beta_f - 1)]c\tau_0$, where $\alpha = \beta_f/(\beta_f - 1)$ comes from the previous paragraph. But if $|\Delta s_0| \ll z_R$, then all pulses contributing have approximately the same C_η function. (All the pulses are at a similar stage of evolution since C_η evolves on the scale of z_R .) Thus, under this condition, the longitudinal factor can be approximated as a delta function: $B(\xi - \Delta(\eta)) \approx B_0\delta(\xi - \Delta(\eta))$. But when using a delta function for B , one obtains exactly Eq. 5.11. Hence, when $\Delta s_0 = |\beta_f/(\beta_f - 1)|c\tau_0 \ll z_R$, the closed-form equation Eq. 5.11 approximates Eq. 5.7 for a constant-velocity flying focus. Another way of looking at this: for certain choices of the B and C functions, Eq. 5.7 is a convolution of Eq. 5.11. For example,

$$A(\mathbf{x}_\perp, s, \xi) \approx \int B(\xi - \xi_0)C_{\alpha(\xi_0)}(s, \mathbf{x}_\perp) d\xi_0 \quad (5.15)$$

is a convolution of Eq. 5.11, where $B(\xi - \xi_0)$ acts as a convolutional filter, and is of the same form as Eq. 5.7 with η reparameterized to a dummy variable ξ_0 . When the duration of the B function is sufficiently small, the convolution does not result in a substantial difference from the original function, Eq. 5.11.

As an example of the utility of Eq. 5.11 for analytic calculations, the duration of a constant-velocity flying focus pulse can be explicitly calculated from Eq. 5.12. Without loss of generality, we compute the duration at $s = 0$. From Eq. 5.13, the on-axis intensity profile in 3 spatial dimensions satisfies

$$|A(\mathbf{0}, 0, \xi)|^2 \propto \frac{1}{\alpha^2\xi^2 + z_R^2} \quad (5.16)$$

From this, the full-width at half-maximum of the intensity peak may be calculated exactly as

$$c\tau_{\text{fwhm}} = 2 \left| \frac{\beta_f - 1}{\beta_f} \right| z_R \quad (5.17)$$

This equation illustrates an important point: the pulse duration depends on the flying focus normalized velocity β_f , laser wavenumber k_0 , and laser spot size w_0 (the latter two implicitly through z_R). This is a constraint that conventional laser pulses do not possess, i.e., the

duration τ_0 can be freely chosen independently of k_0 and w_0 (as in Eq. 5.8). For flying focus pulses, this dependence of the duration on other laser parameters generally constrains their application. For example, a negative-velocity flying focus with $\beta_f = -1$ has been proposed for several applications [66, 68]. In this case, the full-width at half maximum of the intensity is given from Eq. 5.17 by $c\tau_{\text{fwhm}} = 4z_R$, which may be impractically large for standard laser and plasma parameters.

In the particular case of Eq. 5.17, another consequence of the dependence of τ_{fwhm} on β_f is that $\tau_{\text{fwhm}} \rightarrow 0$ as $\beta_f \rightarrow 1$. In practice, $\beta_f = 1$ is a degenerate, unrealizable case because $\alpha \rightarrow \infty$ in this limit. This pathology arises from the parameterization $s_0(\xi) = \alpha\xi$ with a constant amplitude $a_0(\xi) = a_0$. For a conventional laser pulse, $s_0(\xi)$ is constant and $a_0(\xi)$ varies; in this case, the duration is determined by $a_0(\xi)$ and can be arbitrary while the pulse propagates luminally in vacuum. However, in this case the pulse decays in amplitude over the length scale z_R . Note that one could potentially design a non-degenerate closed-form flying focus using Eq. 5.12 with $\beta_f = 1$ by allowing $a_0(\xi)$ to vary, since in this case the location of the intensity peak no longer occurs at the slice which is in focus (see above).

For any $\beta_f \neq 1$, the structure of the laser field will deviate from Eq. 5.11 over length scales longer than L_{stc} (Eq. 5.20). This means that Eq. 5.11 may not behave as expected when designing a flying focus with a focal range $L > L_{\text{stc}}$. This effect becomes prominent when $\beta_f \approx 1$ because $L_{\text{stc}} \propto \tau_{\text{fwhm}}$. Fig. 5.3 shows two simulations which illustrate the efficacy of Eq. 5.11 for designing structured light and clarify this point about the pulse duration. Flying focus pulses described by Eq. 5.12 are shown with $\beta_f = 1.05$ (Fig. 5.3(a-b)) and $\beta_f = 1.005$

(Fig. 5.3(c-d)). Both pulses are initialized using Eq. 5.12 at $t = 0$ with the amplitude profile

$$a_0(\xi) = \begin{cases} 0, & \xi < -\Delta\xi \\ \sin^2 \left[\frac{\pi}{2} \left(\frac{\xi + \Delta\xi}{4\Delta\xi} \right) \right], & \xi \in [-\Delta\xi, -\frac{3}{4}\Delta\xi] \\ 1, & \xi \in [-\frac{3}{4}\Delta\xi, \frac{3}{4}\Delta\xi] \\ \sin^2 \left[\frac{\pi}{2} \left(\frac{\xi - \Delta\xi}{4\Delta\xi} \right) \right], & \xi \in [\frac{3}{4}\Delta\xi, \Delta\xi] \\ 0, & \xi > \Delta\xi \end{cases}, \quad (5.18)$$

where $\Delta\xi \equiv \frac{L}{\alpha}$ is the spread in pulse duration leading to propagation through a distance $L = 10z_R$, and the focal point profile

$$s_0(\xi) = \alpha \left[\xi + \frac{1}{2}\Delta\xi \right] \quad (5.19)$$

where $\alpha = \beta_f/(\beta_f - 1)$ as before. Both pulses have $k_0 w_0 = 40$. In each case, the left plot (Figs. 5.3(a,c)) is the initial structure of the pulse initialized using Eq. 5.12 and the right (Figs. 5.3(b,d)) is the state after a time $\Delta t = L = 10z_R$. For the sake of illustration, the pulses are initialized well-within the focal trajectory: the peak originally occurs at $\xi = \Delta\xi/2$. The initial structure of the two pulses is the same aside from a scale factor of 10 in the longitudinal direction. The duration of the $\beta_f = 1.005$ case is $10\times$ shorter, in agreement with Eq. 5.17.

The state of the pulses at $ct = L = 10z_R$ is shown in Figs. 5.3(b,d). If both pulses evolved according to the approximate solution in Eq. 5.12, they would be symmetric with respect to the initial condition. For the case with $\beta_f = 1.05$ (Fig. 5.3(b)), this is approximately the case. On the other hand, while there is still a discernible focal point for the $\beta_f = 1.005$ case in Fig. 5.3(d), the structure is clearly distorted, the amplitude is reduced, and the ξ value in focus deviates substantially from the prescribed value of $k_0\xi = 20$.

The duration of the pulses may be estimated (not exactly because of the inconstancy of $a_0(\xi)$) from Eq. 5.17 as $\omega_0\tau_{\text{fwhm}} = 8$ for $\beta_f = 1.005$ and $\omega_0\tau_{\text{fwhm}} = 80$ for $\beta_f = 1.05$. For a rough estimate of when effects of spatiotemporal coupling are expected to rise, this duration

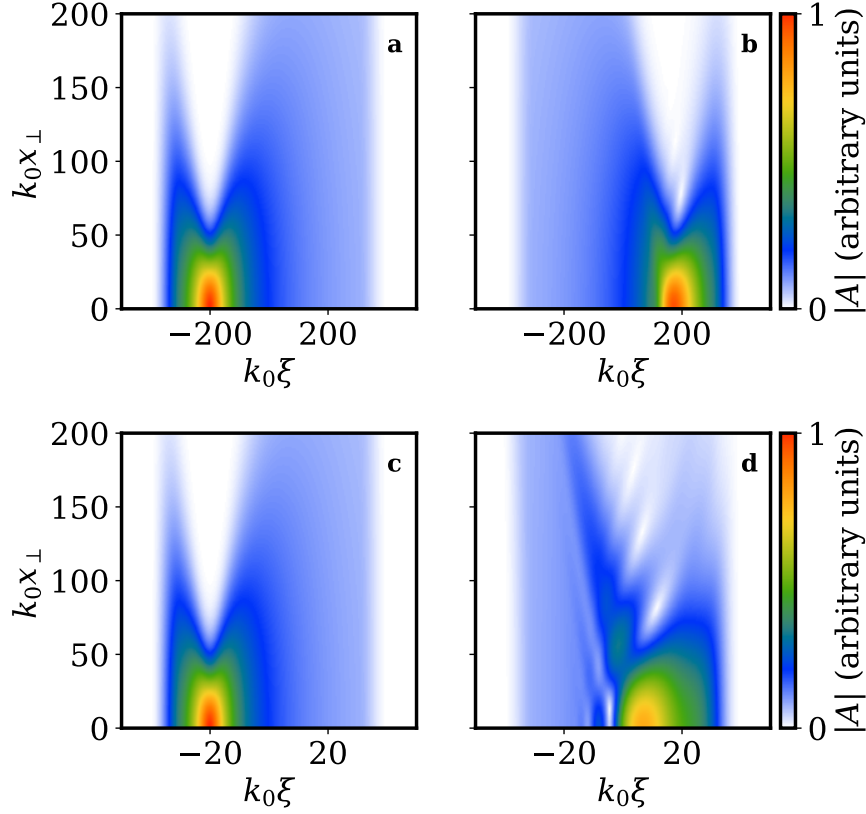


Figure 5.3: QPAD simulations illustrating the spatiotemporal structure of flying focus ASTRL pulses described by the closed-form Eq. 5.12, as well as differences in propagation resulting from the dependence of the duration on the normalized focal velocity β_f (Eq. 5.17). (a) A flying focus with $\beta_f = 1.05$ at $t = 0$. (b) Final state at $c\Delta t = 10z_R$. The state is mirror-symmetric (with minor deviations) because it is described accurately by Eq. 5.12. In this case, $L_{\text{stc}} = 80z_R \gg c\Delta t$. (c) A flying focus with the same focal range but with $\beta_f = 1.005$. The structure is the same as in (a) except for compression by a factor of $10\times$ in the longitudinal direction. (d) At $c\Delta t = 10z_R$, there is significant distortion from Eq. 5.17, consistent with the estimate $L_{\text{stc}} = 8z_R \lesssim c\Delta t$. Pulse parameters are in the text.

can be inserted into the formula for the spatiotemporal coupling length scale, Eq. 5.20. This yields $L_{\text{stc}} = 8z_R$ for $\beta_f = 1.005$ and $L_{\text{stc}} = 80z_R$ for $\beta_f = 1.05$. This estimate is consistent with the observation of major deviations from Eq. 5.12 for the $\beta_f = 1.005$ case and minor deviations for the $\beta_f = 1.05$ case for propagation over $c\Delta t = 10z_R$.

5.5 Experimental approaches to synthesizing ASTRL pulses

Optical configurations for producing specific subclasses of ASTRL pulses have already been proposed or experimentally realized [37, 146, 151, 187]. These configurations may be modified to produce new kinds of ASTRL pulses informed by Eq. 5.7, such as the oscillating flying focus in Fig. 1(c)-(e). In addition, experimental approaches for synthesis of arbitrary solutions of Maxwell's equations as in [155, 156] can be used to create solutions designed using the ASTRL approach. Here we propose a different approach, based on recent developments in divided pulse amplification, which enables synthesis of more general classes of ASTRL pulses.

Wavelength-multiplexed communication and power scaling have motivated decades of progress in methods for combining continuous-wave laser beams [188, 189, 190, 191]. Conceptually, N independent beams or N copies of one beam are individually amplified and then recombined into a single beam with N times the original power. Splitting and recombination are achieved using standard optics, such as diffraction gratings, fibers, and beamsplitters. More recently, these methods have been applied to ultrashort laser pulses as a path towards pulses beyond the petawatt class [192, 193, 194, 195, 196, 197, 198, 199, 200], with coherent combinations of up to 81 pulses [201]. In this method, one could also use random or engineered phases for each of the N copies.

Divided pulse amplification techniques can be extended to produce ASTRL pulses. In addition to being amplified, the separated pulses can be independently focused, delayed, polarization-rotated, or altered in other ways before being recombined. Figure 5.4 shows a schematic of the proposed experimental concept in which different optical assemblies

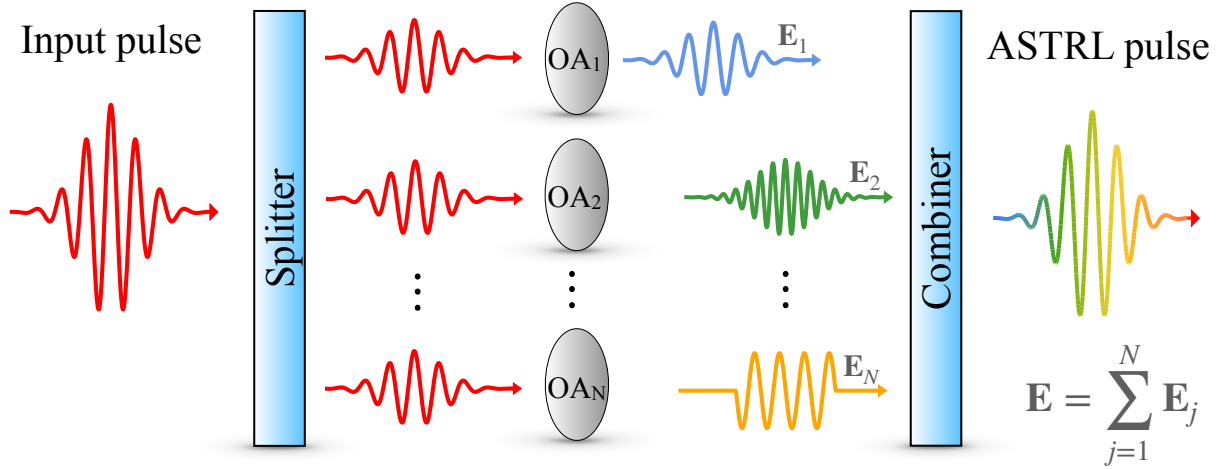


Figure 5.4: An experimental concept to produce new classes of ASTRL pulses. Each pulse is independently manipulated by a separate optical assembly, labeled “OA_j”, which can alter the amplitude, transverse profile, frequency, delay, or other properties. The pulses are then coherently combined to form an ASTRL pulse.

independently manipulate each of the divided pulses. As an example, optical assemblies composed of a spiral phase plate and delay stage could produce a pulse with continuously evolving OAM as in Fig. 2(c); alternatively, using a half-waveplate and delay stage could produce a pulse with continuously evolving polarization as in Fig. 2(a). Despite the potential challenges of alignment and timing, divided pulse manipulation could provide unprecedented control over the structure of the combined pulse and enable creation of the broad class of pulses described by Eq. 5.7. Note that Ref. [169] has used a special case of this concept to combine two pulses with different OAM; the combined pulse was used to drive high-harmonic attosecond pulses.

This approach to synthesizing ASTRL pulses is based on approximating the integral in Eq. 5.7 as a discrete sum of pulses. To illustrate this, consider the $\beta_f = 1.05$ flying focus example shown in Fig. 5.1. Figure 5.5 demonstrates that 20 pulses are more than sufficient to approximate the integral when $L = 10z_R$. More generally, the minimum number of laser

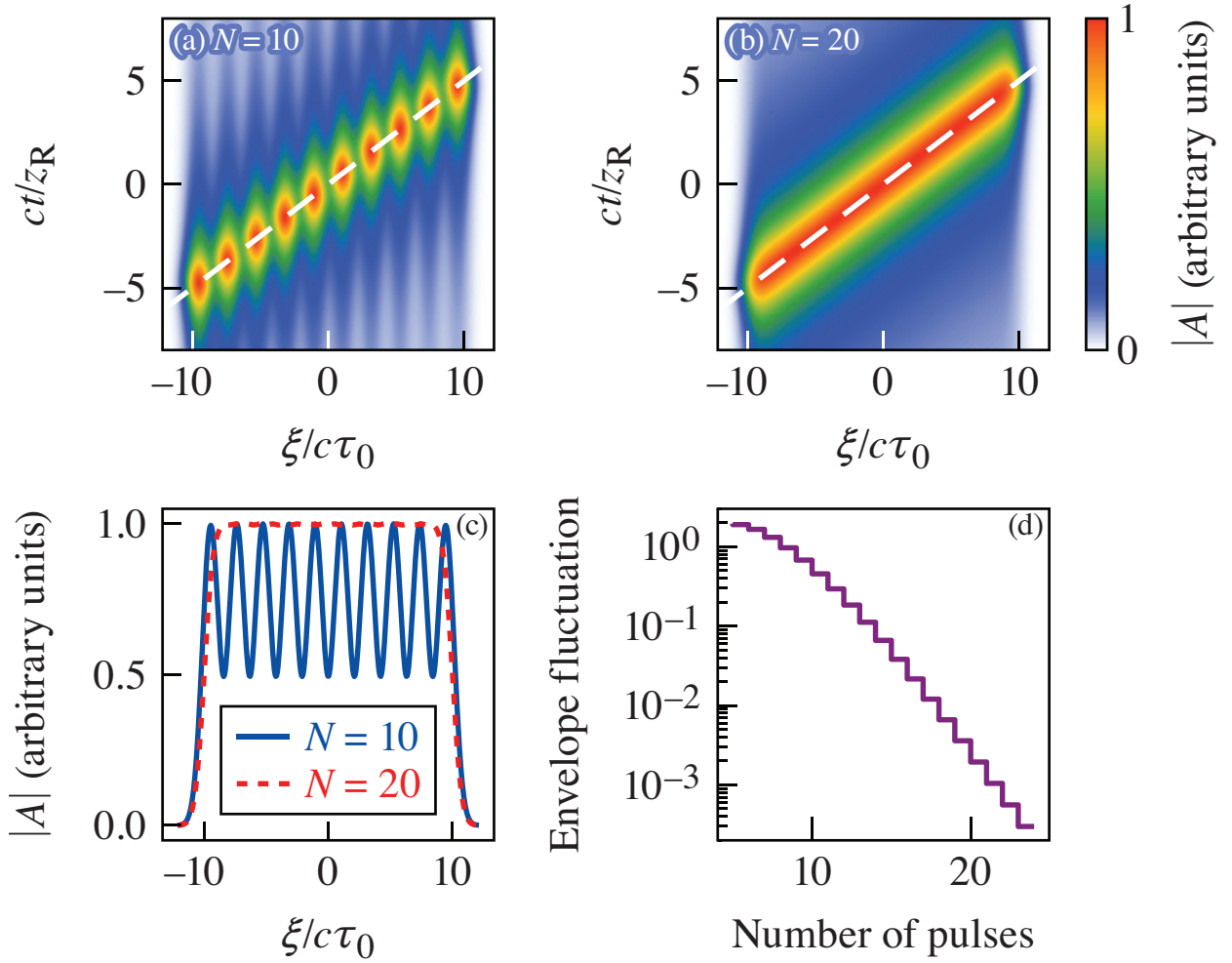


Figure 5.5: (a,b) Discrete approximations of Eq. 5.7 for the $\beta_f = 1.05$ flying focus pulse shown in Fig. 1 with $N = 10$ and $N = 20$ pulses, respectively; the continuum is well approximated by 20 pulses, or 2 pulses per z_R . (c) Peak on-axis amplitude for the $N = 10$ and 20 cases, showing oscillation for $N = 10$. (d) Relative amplitude oscillation, defined as $(\max |A| - \min |A|)/(\max |A| + \min |A|)$ over the interval $[-z_R, z_R]$. The relative oscillation decreases exponentially with N , which determines the number of pulses required for a given tolerance.

pulses required to approximate Eq. 5.7 can be estimated by ensuring that the sum-total duration of all constituent pulses is greater than the duration of the entire ASTRL pulse

T , i.e., $N \gtrsim T/\tau_0$. For a flying focus pulse, the effective duration of the intensity peak, $\tau_{\text{eff}} = (1 - \beta_f)z_R/v_f$ [153], satisfies $\tau_{\text{eff}}/T = z_R/L$, which implies that $N \gtrsim (1 - \beta_f)L/v_f\tau_0$ pulses are required. Note, however, that for certain applications, discrete separation of the constituent pulses as in the $N = 10$ case of Fig. 5.5 may also be desirable.

For the closed-form ASTRL pulse in Eq. 5.11, methods for arbitrary waveform generation (e.g., Refs. [156, 155]) can potentially be used. Alternatively, this form can be used to phenomenologically model a given experimental setup. For example, in Ref. [202], this closed-form ASTRL pulse was used to computationally model the fields of a flying focus formed by sending a chirped pulse through a chromatic lens (as in Ref. [3]) in the context of laser wakefield acceleration.

5.6 Accuracy of Equation 5.6

As demonstrated by the examples above, an ASTRL pulse may generally exhibit spatiotemporal coupling despite being assembled from pulses with no spatiotemporal coupling (i.e., Eq. 5.6). The separability of Eq. 5.6 was derived by reducing the full wave equation (Eq. 5.4) to the paraxial wave equation (Eq. 5.5). The validity of the approximations leading to Eq. 5.5 may be determined by evaluating derivatives of the approximate solution $A(\mathbf{x}_\perp, s, \xi) = z_R/q(s) \exp[-ik_0 x_\perp^2/2q(s)] \exp[-(\xi/c\tau_0)^2]$. At any location, $|\partial_s^2 A| \ll |2ik_0 \partial_s A|$ when $k_0 w_0 \gg 1$; this is the usual paraxial approximation, which fails only for tightly focused laser pulses. On the other hand, $|2\partial_\xi \partial_s A| \ll |2ik_0 \partial_s A|$ when $\omega_0 \tau_0 \gg 1$. This condition only fails for pulses with durations less than a few-cycles. These two conditions, i.e., $k_0 w_0 \gg 1$ and $\omega_0 \tau_0 \gg 1$, place constraints on C_η and B_η , respectively.

For an envelope that is initially separable at some $s = s_i$, spatiotemporal coupling will develop beyond a range $|s - s_i| \gtrsim L_{\text{stc}} \equiv \frac{1}{2}c\tau_0 k_0^2 w_0^2$ because of the $\partial_\xi \partial_z$ term in Eq. 5.4 (see Sec. 5.7). There are two common scenarios that can be considered. In the first scenario, every pulse comprising an ASTRL pulse is separable at its focus. Because each pulse only

contributes significantly within a Rayleigh range of its focus, the condition on the accuracy of Eq. 5.6 can be expressed as $L_{\text{stc}} \gtrsim z_R$ or, equivalently, $\omega_0 \tau_0 \gtrsim 1$. In the second scenario, which is more applicable to field initialization in a simulation, every pulse is separable at a specified distance L_η from focus. The condition on the accuracy of Eq. 5.6 is then given by $L_{\text{stc}} \gtrsim \max_\eta L_\eta$. For example, in Fig. 1(a-b) $\max_\eta L_\eta = \frac{1}{2}L_{\text{stc}}$, which contributes to the slight discrepancy at the end of the focal region. However, even if $L_{\text{stc}} \lesssim \max_\eta L_\eta$ and spatiotemporal coupling develops in the constituent pulses, the qualitative features of the desired electromagnetic structure would remain intact.

5.7 Length scale for emergence of spatiotemporal coupling

Over a long enough propagation path, the $2\partial_s\partial_\xi$ term omitted in Eq. 5.5 will lead to the development of spatiotemporal coupling in an initially separable envelope. The length scale over which this coupling develops can be estimated using geometric optics. Consider a pulse initialized by Eq. 5.6 at $s = -s_0$ with a focal point at $s = 0$. The distance from a radial location r to the focal point is given by $d(r) = \sqrt{r^2 + s_0^2}$. The separable form of Eq. 5.6 begins to break down when the delay between the center and outer radii of the pulse (at the same initial ξ) becomes comparable to its duration τ_0 . For an initial Gaussian transverse profile, the spot size at $s = -s_0$ is $w(-s_0) \approx w_0 s_0 / z_R = 2s_0 / k_0 w_0$, where w_0 is the spot size at focus. Setting $r = w$ provides the delay $\Delta \sim \sqrt{(2s_0 / k_0 w_0)^2 + s_0^2} - s_0 \sim 2s_0 / (k_0 w_0)^2$. Equating Δ to $c\tau_0$, we find that that Eq. 5.6 breaks down for propagation distances

$$s_0 \gtrsim L_{\text{stc}} \equiv \frac{1}{2}c\tau_0 k_0^2 w_0^2. \quad (5.20)$$

For most pulses of interest, $\omega_0 \tau_0 \gg 1$ so that $L_{\text{stc}} \gg z_R$, i.e., Eq. 5.6 remains valid over distances much longer than a Rayleigh range.

Equation 5.20 can also be derived using a multi-scale expansion of the operators in Eq. 5.4, where it is explicitly assumed that the mixed derivative term is a first-order correction. The ∂_s operator on the full solution A may be expanded as $\partial_s \approx \partial_{s_0} + \partial_{s_1}$, where $\partial_{s_0} \gg \partial_{s_1}$. To

lowest order, $k_0 \partial_{s0}$ balances $\nabla_\perp^2$ and one finds the scaling, $\partial_{s0} \sim z_R^{-1}$. At first order, $k_0 \partial_{s1}$ balances $\partial_{s0} \partial_\xi$, providing the scaling relation $k_0/L_{\text{stc}} \approx 1/c\tau_0 z_R$. Upon solving for L_{stc} , this reduces exactly to Eq. 5.20.

5.8 Separable envelope decomposition of arbitrary laser fields

Consider an electric field with characteristics of a monochromatic laser pulse, i.e., a power spectrum peaked at a single ω_0 such that $\omega_0 \tau_0 \gtrsim 1$, but otherwise arbitrary structure. It is possible to show that such an arbitrary field may be decomposed as a superposition of separable pulses through a generalization of Eq. 5.7. In particular, the envelope may be exactly decomposed as

$$A(\mathbf{x}_\perp, s, \xi) = C_\psi^{-1} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{da db}{a^2 |a|^{1/2}} C^{(a,b)}(\mathbf{x}_\perp, s) \psi \left(\frac{\xi - b}{a} \right) \quad (5.21)$$

where C_ψ^{-1} is a constant, ψ is a wavelet mother function, and

$$C^{(a,b)}(\mathbf{x}_\perp, s) = \frac{1}{|a|^{1/2}} \int_{\mathbb{R}} d\xi A(\mathbf{x}_\perp, s, \xi) \psi^* \left(\frac{\xi - b}{a} \right) \quad (5.22)$$

is the wavelet transform of $A(\mathbf{x}_\perp, s, \xi)$ with respect to ξ for fixed $\mathbf{x}_\perp$ and s [203]. For example, one may use the Ricker wavelet $\psi(g) = \psi_0 \exp[-g^2/2](1 - g^2)$. In Eq. 5.21, a and b are respectively the length and delay of each constituent pulse. The properties of each $C^{(a,b)}(\mathbf{x}_\perp, s)$, such as spot size and focal point, may depend on a and b .

Equation 5.21 is a superposition of separable pulses at each slice in s . Given initial data for the envelope $A(\mathbf{x}_\perp, 0, \xi)$ at some slice $s = 0$, the decomposition may be used to approximately evolve the full solution $A(\mathbf{x}_\perp, s, \xi)$ using Eq. 5.6 as

$$A(\mathbf{x}_\perp, s, \xi) \approx C_\psi^{-1} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{da db}{a^2 |a|^{1/2}} C_{\text{parax}}^{(a,b)}(\mathbf{x}_\perp, s) \psi \left(\frac{\xi - b}{a} \right) \quad (5.23)$$

where $C_{\text{parax}}^{(a,b)}(\mathbf{x}_\perp, s)$ are solutions to the paraxial wave equation $[2ik_0 \partial_s + \nabla_\perp^2] C_{\text{parax}}(\mathbf{x}_\perp, s) = 0$ with initial condition $C_{\text{parax}}(\mathbf{x}_\perp, 0) = C^{(a,b)}(\mathbf{x}_\perp, 0)$. The properties of $C_{\text{parax}}^{(a,b)}$ determine the

range of s over which Eq. 5.23 remains valid. For example, if $C_0^{(a,b)}(\mathbf{x}_\perp)$ resembles a Gaussian beam with spot size $w_0(a,b)$ for each (a,b) , then Equation 5.23 remains valid when

$$|s| \lesssim \min_{a,b} L_{\text{stc}}(a,b) = \min_{a,b} ak_0^2 w_0(a,b)^2, \quad (5.24)$$

where the minimum is taken over (a,b) with significant contribution to the integral in Eq. 5.21. Note that the specific choice of wavelet is not important, but the wavelet determines the conditions required for validity of Eq. 5.23.

Equation 5.23 generalizes Eq. 5.7 and shows that any fully coherent laser pulse solution may be approximately decomposed as a superposition of pulses with separable envelopes. In this sense, all monochromatic laser pulses can be constructed as ASTRL pulses. However, we restrict the scope of ASTRL pulses to include pulses that have been intentionally designed to exhibit a particular structure either through Eq. 5.2 or Eq. 5.7.

5.9 Discussion

In this chapter, the concept of Arbitrarily Structured Laser (ASTRL) pulses was introduced. In addition to facilitating the creation of new and previously unimagined electromagnetic structures, the ASTRL formalism simplifies the injection of structured light in simulations. Previous simulations of flying focus pulses required a Fresnel integral to propagate the field from the optic plane of a specific experimental design to the simulation plane. In contrast, the ASTRL description enables direct field initialization in the simulation plane, which simplifies implementation. Further, by decoupling the structured light-matter interaction physics from the details of a particular optical configuration, the ASTRL description expedites prototyping of new concepts. For example, the axiparabola-echelon pair used in Ref. [37] would not need to be redesigned for every simulation. An optical configuration capable of delivering the desired pulse can be designed or implemented in a simulation after a concept shows promise.

The generality of the ASTRL concept extends well beyond Eq. 5.7 and the examples

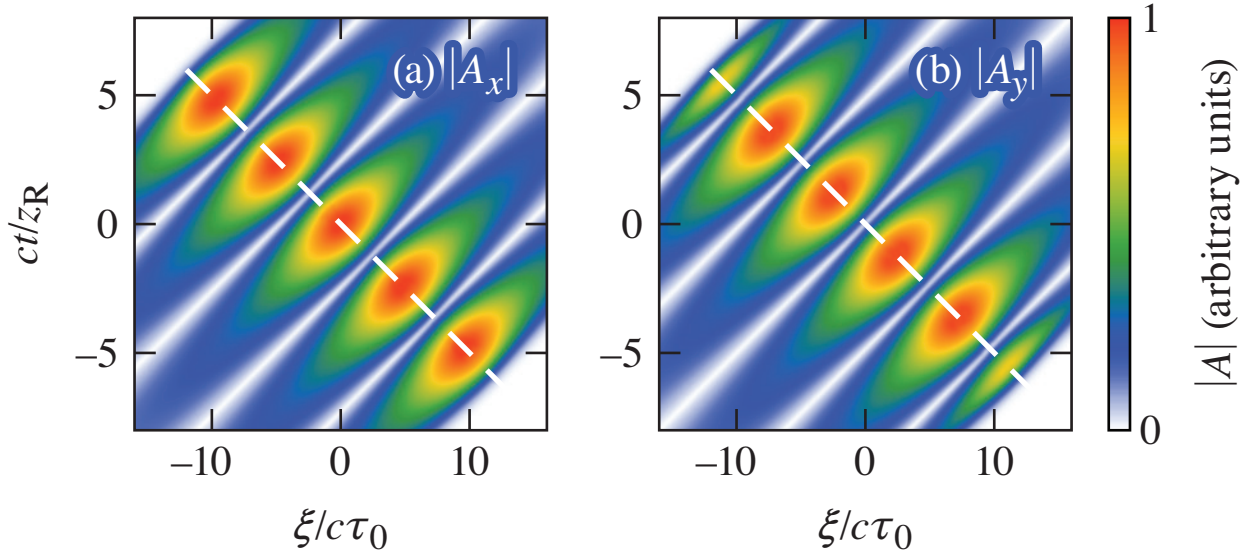


Figure 5.6: (a,b): On-axis evolution of the vector envelope for a negative-velocity flying focus with structured polarization. The components of the vector envelope $\mathbf{A}$ are defined by $A_j \equiv \hat{\mathbf{e}}_j \cdot \int d\eta B_\eta(\xi) C_\eta(\mathbf{x}_\perp, s) \hat{\mathbf{e}}(\eta)$. The parameters are identical to those in Fig. 5.1 except $\beta_f = -1$ and the polarization evolves continuously. Specifically, $\hat{\mathbf{e}}_x(\eta) = \cos(k_{\text{osc}}\eta L)$ and $\hat{\mathbf{e}}_y(\eta) = \sin(k_{\text{osc}}\eta L)$, where $k_{\text{osc}} = (3/2)z_R^{-1}$. The white dashed lines indicate the focal trajectory $z = -ct$. Both plots are evaluated using Eq. 5.7.

presented here. The time-dependence of any combination of parameters can be simultaneously structured. For instance, a negative-velocity flying focus can be combined with continuously evolving polarization (Fig. 5.6). Further, the ASTRL concept can be extended to accommodate multi-color pulses [204, 205], superpositions of pulses with vector polarization [206, 207, 208, 209], and pulses structured in the spatio-spectral domain. In fact, using a wavelet transform one can show that arbitrary monochromatic laser fields may be decomposed as a generalized version of Eq. 5.7 (see Sec. 5.8).

In summary, the ASTRL formalism provides a framework for constructing laser pulses with desired spatiotemporal structure. An ASTRL pulse is synthesized using superpositions of known solutions to Maxwell's equations, such as an ensemble of traditional laser pulses.

The flexibility of the ASTRL concept was illustrated with several examples, including pulses with moving focal points, dynamic polarization, evolving angular momentum, and nontrivial topological structure. In practice, ASTRL pulses can be synthesized using a number of experimental techniques, including the divided pulse manipulation scheme proposed here. New pulses based on the ASTRL concept may enable or enhance techniques in a range of scientific disciplines, including microscopy, nonlinear optics, quantum optics, and laser-plasma interactions.

CHAPTER 6

Laser Wakefield Acceleration with Arbitrarily Structured Laser Pulses

In the last chapter, the concept of Arbitrarily Structured Laser (ASTRL) pulses was introduced. In this chapter, we apply ASTRL pulses to laser wakefield acceleration (LWFA). LWFA may enable the next generation of TeV-scale lepton colliders. Reaching such energies will likely require multiple LWFA stages to overcome limitations on the energy gain achievable in a single stage. The use of stages, however, introduces challenges such as alignment, adiabatic matching between stages, and a lower average accelerating gradient. Here, we propose a discrete flying focus that can deliver higher energy gain in a single stage, thereby reducing the number of stages required for a target energy. A sequence of laser pulses with staggered focal points and delays drives a plasma wave in which an electron beam experiences a near-constant accelerating gradient over distances beyond those attainable with a conventional pulse. Results from simulations in this chapter demonstrate that a discrete flying focus with a total energy of 150 J can transfer 40 GeV per electron to a 50-pC beam in a single 30-cm stage, corresponding to 50 dephasing lengths. As discussed later, there is much room for optimization of this method to deliver higher energies within a single plasma stage. In addition, preliminary results demonstrating the efficacy of an ASTRL flying focus for extending the acceleration length of wakefields suitable for positron acceleration are presented.

Most of this chapter comes from Ref. [8]. This article is distributed under the terms of the Creative Commons CC BY 4.0 license, which permits unrestricted use, distribution, and

reproduction in any medium, provided the original work is properly cited. The main addition is Sec. 6.7, which explores the use of an ASTRL LWFA driver for positron acceleration.

6.1 Introduction

In laser wakefield acceleration (LWFA), a laser-driven plasma wave focuses and accelerates an electron beam [24]. Because plasma waves can sustain accelerating gradients that are orders of magnitude higher than those of conventional radio-frequency accelerators, LWFA has been considered for a next-generation TeV-scale lepton collider [210, 18]. However, the required TeV-scale beam energies are far higher than the GeV-scale energies that can be transferred to a beam in a single LWFA stage. This is due to two limitations: (i) dephasing [Fig. 6.1(a)], where the beam outruns the accelerating phase of the plasma wave, and (ii) pump depletion, where the laser pulse exhausts its energy by driving plasma waves. To overcome these limitations, the advanced accelerator community envisions chaining together multiple LWFA stages. For example, the Snowmass 2021 Design Frontier proposed an electron accelerator with an accelerating gradient of ~ 100 GeV/m, but only ~ 3 GeV of energy gain per 3-cm stage [18]. With these parameters, ~ 300 stages would be required to accelerate an electron beam to 1 TeV.

The use of stages for LWFA poses significant challenges, including beam transport between stages, adiabatic matching of the beam into each plasma section, and alignment and temporal synchronization of the beam and laser pulses [6, 211, 212, 19]. In addition, the meter-scale gaps required between each stage would reduce the average accelerating gradient by a factor of ~ 30 for the 3-cm stage case described above. These challenges, along with the overall collider complexity, can be mitigated by decreasing the number of stages N_{stages} . One approach is to increase the energy gain in each stage by operating at lower electron densities: $\Delta E_{\text{stage}} \propto n_0^{-1}$ so that $N_{\text{stages}} \propto n_0$. However, this approach decreases the accelerating gradient ($E_z \propto \sqrt{n_0}$) and hence increases the total length of plasma $L_T \propto n_0^{-1/2}$ [6]. Another approach would be

to use a density upramp to eliminate dephasing [112], but pump depletion would remain an issue.

Space-time structured laser pulses offer an alternative approach to enhancing the energy gain in a single LWFA stage [213, 37, 38, 45, 214, 215]. These “flying focus” concepts rely on a controllable-velocity intensity peak to keep the accelerating phase of the driven plasma wave collocated with an electron beam, thereby eliminating dephasing. This allows for operation at higher densities, enabling larger energy gains over the same length or identical energy gains over shorter lengths when compared to equal-energy conventional pulses. By providing higher energy gains, such concepts could reduce the number of required stages without increasing the length of the plasma. Flying-focus pulses have been experimentally demonstrated using two techniques: (i) chromatic focusing of a chirped laser pulse [3, 216] and (ii) reflection from an axiparabola–echelon pair, which independently focuses and delays radial annuli of a pulse [217]. This latter technique is now being considered as the basis for an experiment to achieve a 100-GeV-scale electron beam in a single meter-scale stage as part of the NSF OPAL project [46, 218].

Here, we introduce a concept for enhancing the energy gain in a single LWFA stage using a train of collinear laser pulses with different focal points [Fig. 6.1(b)]. The delays between the pulses in this discrete flying focus (DFF) are chosen so that each pulse comes into focus and drives the plasma wave a fixed distance ahead of the electron beam. Despite the subluminal group velocity of each pulse, the accelerating phase of the plasma wave advances at the vacuum speed of light. As a result, the accelerating phase remains collocated with a highly relativistic electron beam over the entire stage. In the linear regime (normalized vector potentials $a_0 \lesssim 0.5$), each pulse drives the plasma wave strongly over its Rayleigh range. In the nonlinear regime ($a_0 \gtrsim 2$), each pulse self-guides and strongly drives the plasma wave over its pump depletion length. Quasistatic particle-in-cell simulations in this regime show that a DFF can drive a plasma wave that accelerates a 50-pC electron beam by 40 GeV in a single 30-cm stage, corresponding to 50 dephasing lengths. The 40 GeV energy gain is not a

fundamental limitation of the DFF; we comment on how the energy can be increased at the end of the chapter.

The remainder of this chapter begins with theory describing propagation of a DFF pulse (Sec. 6.2). The theory extends the Arbitrarily Structured Laser (ASTRL) pulse concept [7] to account for plasma dispersion. The pulse and plasma parameters for an LWFA stage far longer than a dephasing length are designed using scaling laws in the linear and nonlinear regimes (Sec. 6.3). QPAD simulations [39, 81] based on these designs confirm that a plasma wave driven by a DFF can accelerate an externally injected electron beam over tens of dephasing lengths in both regimes (Sec. 6.3). Higher order effects of linear propagation not considered in the theory section are considered more carefully in Sec. 6.4. The accuracy of QPAD compared to fully electromagnetic PIC simulations is demonstrated and discussed in Sec. 6.6. Additional simulation details for the simulations in Sec. 6.3 are provided in Sec. 6.5. Simulations of the same scheme using a Laguerre-Gaussian driver for positron are shown in Sec. 6.7. Finally, a summary and possibilities for future directions, including various ways to optimize this configuration, are provide in Sec. 6.8.

6.2 Theory

The design of a DFF for LWFA begins with an extension of the formalism developed for ASTRL pulses [7] to account for linear dispersion in cold, uniform plasma. The propagation of a laser pulse through such a plasma is described by the wave equation

$$\left[\nabla^2 - \frac{1}{c^2} \partial_t^2 \right] \mathbf{A} = \frac{\omega_p^2}{c^2} \mathbf{A}, \quad (6.1)$$

where $\mathbf{A}$ is the vector potential of the pulse in the Coloumb gauge, $\nabla \cdot \mathbf{A} = 0$, and $\omega_p = (e^2 n_0 / m_e \epsilon_0)^{1/2}$ is the plasma frequency. For propagation along the $\hat{z}$ direction, the transverse vector potential can be expressed as a normalized vector envelope $\mathbf{a}(t, z, \mathbf{x}_\perp)$

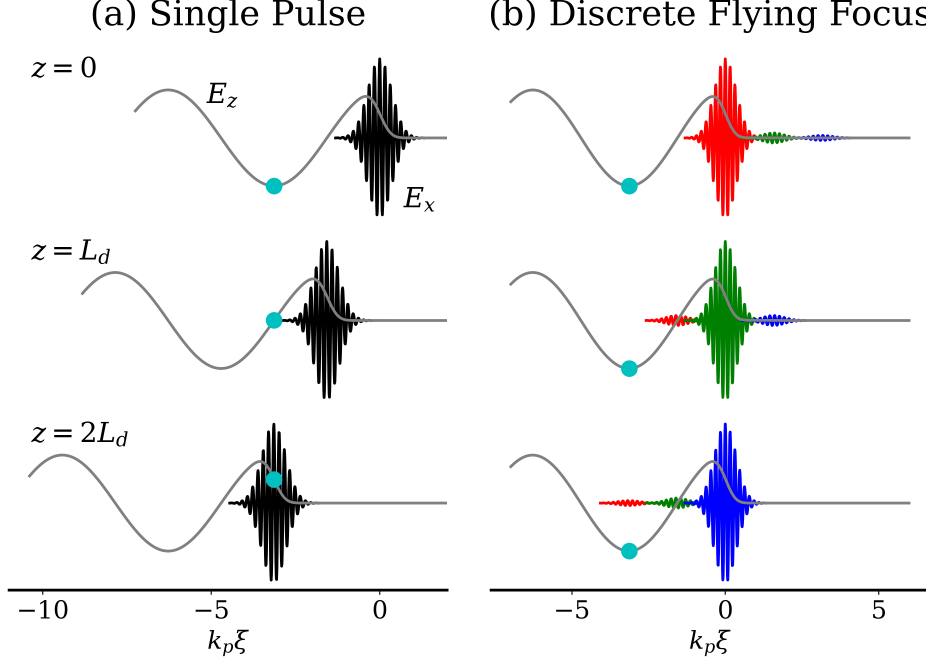


Figure 6.1: Comparison of LWFA with a conventional pulse and a discrete flying focus in the linear regime. (a) A single pulse traveling at its group velocity (black) drives a plasma wave (gray), whose longitudinal field E_z accelerates a highly relativistic electron (cyan). The pulse and longitudinal field slip backward in $\xi = z - ct$ as the pulse propagates from $z = 0$ to $2L_d$. As a result, the electron moves from an accelerating to a decelerating longitudinal field, which limits its energy gain—a process referred to as dephasing. (b) A sequence of pulses (red, green, blue) with appropriately selected delays and focal points drives a plasma wave (gray) whose longitudinal field remains stationary in ξ . This eliminates dephasing by ensuring that the electron (cyan) remains collocated with the maximum accelerating field over the entire distance. In both (a) and (b), the field of the laser pulse is calculated using Eq. (6.7), and the longitudinal field is calculated using linear theory [5, 6]. The specific parameters, chosen for illustration, are $k_0/k_p = 40$, $k_p\tau_0 = 0.5$, $f_j = jL_d$, and $\Delta_j = (1 - v_g/c)f_j = (\frac{\pi}{2})jk_p^{-1}$. The dephasing length is $k_pL_d = 1600\pi$, the single pulse in (a) has $k_pz_R = \infty$, and each pulse in (b) has $k_pz_R = 320$.

modulated by a carrier wave

$$\mathbf{A}_\perp(t, z, \mathbf{x}_\perp) = \frac{1}{2} \left(\frac{m_e c^2}{e} \right) \mathbf{a}(t, z, \mathbf{x}_\perp) e^{ik_0(z - v_\phi t)} + \text{c.c.}, \quad (6.2)$$

where k_0 is the central wavenumber and $v_\phi = \omega_0/k_0$ is the phase velocity, with $\omega_0^2 = \omega_p^2 + c^2 k_0^2$. Because the transverse components of the vector potential evolve independently, it is sufficient to consider the evolution of a single component: $\mathbf{a}(t, z, \mathbf{x}_\perp) = a(t, z, \mathbf{x}_\perp) \hat{\mathbf{e}}$ where $\hat{\mathbf{e}}$ is a polarization unit vector. Substituting Eq. (6.2) into Eq. (6.1) and transforming to the coordinates $(\zeta, s) \equiv (z - v_g t, z)$, where $v_g = c^2 k_0 / \omega_0 \approx [1 - \frac{1}{2}(\omega_0/\omega_p)^2] c$ is the linear group velocity, yields

$$\left[\partial_s^2 + 2\partial_s \partial_\zeta + \left(1 - \frac{v_g}{c} \right)^2 \partial_\zeta^2 + \nabla_\perp^2 + 2ik_0 \partial_s \right] a(\zeta, s, \mathbf{x}_\perp) = 0. \quad (6.3)$$

The first three terms in Eq. (6.3) are dropped since they are small compared to the $\nabla_\perp^2$ and $2ik_0 \partial_s$ terms (see Sec. 6.4). This reduces Eq. (6.3) to the steady-state paraxial wave equation

$$[\nabla_\perp^2 + 2ik_0 \partial_s] a(\zeta, s, \mathbf{x}_\perp) \approx 0. \quad (6.4)$$

The absence of ∂_ζ terms in Eq. (6.4) implies that each temporal slice of the laser pulse evolves independently. Thus, Eq. (6.4) admits separable solutions of the form

$$a(t, z, \mathbf{x}_\perp) = \mathcal{A}_0 B(z - v_g t) C(z, \mathbf{x}_\perp), \quad (6.5)$$

where the coordinates (t, z) have been reintroduced. Here, $\mathcal{A}_0$ is the amplitude, B is an arbitrary complex function, and C is a complex function satisfying the paraxial wave equation $[\nabla_\perp^2 + 2ik_0 \partial_z] C = 0$. This is the same form derived for vacuum propagation in Ref. [7] with v_g replacing the vacuum speed of light in B .

A laser pulse with an envelope described by Eq. (6.5) will excite a plasma wave with a relativistic phase velocity when propagating in an underdense plasma ($\omega_p \ll \omega_0$). In the linear regime ($a_0 \equiv \max|a| \lesssim 0.5$), the distance over which a plasma wave excited by the solution in Eq. (6.5) can accelerate a particle with $v_z \approx c$ is limited by diffraction and dephasing. Diffraction is associated with the transverse profile $C(z, \mathbf{x}_\perp)$, which determines the length

scale over which the pulse maintains a high intensity and strongly drives the plasma wave. Dephasing is associated with the temporal profile $B(z - v_g t)$, which determines the phase velocity of the plasma wave $v_w = v_g$. The wavelength of the wave is $\lambda_w = 2\pi k_p^{-1}$, where $k_p \equiv \omega_p/c$ is the plasma wavenumber. Thus, electrons with $v_z \approx c$ initially collocated with the maximum accelerating field of the wave will encounter a decelerating field when the laser pulse has slipped backward by $\Delta\xi = (\pi/2)k_p^{-1}$, where $\xi \equiv z - ct$. (Note that, in the linear regime, only a quarter of the full phase of the wave both accelerates and focuses an electron.) This occurs after a distance referred to as the dephasing length L_d , given by

$$k_p L_d \equiv \frac{\pi}{2} \left(1 - \frac{v_w}{c}\right)^{-1}. \quad (6.6)$$

For $\omega_0/\omega_p \gg 1$, $k_p L_d \approx \pi(\omega_0/\omega_p)^2$. Figure 6.1(a) illustrates dephasing in the linear regime for a single pulse with $C = 1$. The pulse (black) and accelerating field of the plasma wave (E_z , gray dotted line) slip backward with respect to the electron (cyan) such that it is no longer accelerated for $z > L_d$.

The key idea of the present work is that dephasing can be overcome by using a train of collinear pulses with different focal points. To demonstrate this, a linear superposition of N laser pulses of the form given by Eq. (6.5) is considered:

$$\mathbf{a}(t, z, \mathbf{x}_\perp) \approx \sum_{j=1}^N \frac{1}{2} \mathcal{A}_{0j} B_j(z - v_g t) C_j(z, \mathbf{x}_\perp) \times e^{ik_0(z - v_\phi t)} \hat{\mathbf{e}}_j + \text{c.c.} \quad (6.7)$$

The amplitude of each pulse $\mathcal{A}_{0j}$ is chosen to produce a desired amplitude a_0 for the overall structure. The profiles B_j and C_j are selected so that each pulse comes into focus when it has slipped backward into the value of ξ where the previous pulse was focused. Here, as in Ref. [7], a Gaussian temporal profile is used:

$$B_j(z - v_g t) = \exp \left[- \left(\frac{z - v_g t - \Delta_j}{c\tau_j} \right)^2 \right], \quad (6.8)$$

where τ_j is the duration and Δ_j is the initial peak position of each pulse. Similarly, the paraxial solution for a Gaussian beam is used for the transverse profile:

$$C_j(z, \mathbf{x}_\perp) = \left(\frac{z_{R,j}}{q_j(z)} \right) \exp \left[- \frac{ik_0 |\mathbf{x}_\perp|^2}{2q_j(z)} \right], \quad (6.9)$$

where $q_j(z) \equiv z - f_j + iz_{R,j}$ is the complex beam parameter, f_j is the focal point, $z_{R,j} = \frac{1}{2}k_0 w_{0j}^2$ is the Rayleigh range, and w_{0j} is the spot size of each pulse. For simplicity, the same amplitude $\mathcal{A}_{0j} = \mathcal{A}_0$, Rayleigh range $z_{R,j} = z_R$, and duration $\tau_j = \tau_0$ are used for each pulse.

With these choices, the overall amplitude of the envelope (Eq. (6.7)) peaks when each pulse comes into focus at $z_{\text{focus},j} = f_j$ and $t_{\text{focus},j} = (f_j - \Delta_j)/v_g$. The corresponding ξ value of the peak amplitude is then given by $\xi_{\text{focus},j} = z_{\text{focus},j} - ct_{\text{focus},j} = (c/v_g)\Delta_j + (1 - c/v_g)f_j$. Thus, each pulse will come into focus at the same value of $\xi_{\text{focus},j} = \xi_0$ if the Δ_j are chosen as

$$\Delta_j = \left(\frac{v_g}{c}\right) \xi_0 + \left(1 - \frac{v_g}{c}\right) f_j. \quad (6.10)$$

In the limit of many pulses ($N \gg 1$), this choice for Δ_j produces a continuous peak that moves through the plasma at the vacuum speed of light $v_f = c$ (or through vacuum at the superluminal velocity $v_f = c^2/v_g$). In practice, the phase velocity of the plasma wave v_w may be slightly different from v_f due to nonlinear effects, linear effects absent in Eq. (6.7) (e.g., higher-order dispersion or space-time couplings), and the dependence of the plasma wave structure on the entire pulse profile rather than its peak location. Nevertheless, Eq. (6.10) provides a starting point that can be tailored to compensate for these effects and achieve $v_w = c$.

Figure 6.1(b) illustrates how dephasing can be circumvented in the linear regime using a DFF with $N = 3$ pulses. The schematic shows the transverse vector potential of the three pulses [Eq. (6.7)] and the longitudinal electric field E_z of the plasma wave that they drive. As each pulse slips backward in ξ , it is replaced by its successor, which becomes the predominant driver of the plasma wave. With Δ_j and f_j related by Eq. (6.10), the ξ value of the peak accelerating field remains stationary and collocated with a highly relativistic electron (cyan) over multiple dephasing lengths ($z = 0$ to $2L_d$).

The total length that the DFF drives the plasma wave L_D is determined by the number of pulses N , the distance between successive focal points $\delta f_j = f_{j+1} - f_j$, and the distance over

which a single pulse drives the plasma wave L_p . For equally spaced focal points, $\delta f_j = \delta f$ and

$$L_D = (N - 1)\delta f + L_p. \quad (6.11)$$

A minimum number of pulses $N_{\min}$ is required to maintain a near-uniform accelerating field over the length L_D . This minimum can be estimated from the condition that the distance between successive focal points is less than the distance over which a single pulse drives the plasma wave: $\delta f < L_p$. A single pulse maintains a high amplitude and drives the plasma wave over a Rayleigh range z_R in the linear regime ($a_0 \lesssim 0.5$) and a pump depletion length L_{pd} in the nonlinear regime ($a_0 \gtrsim 2$) [1]. For a nonlinearly matched pulse with $k_p w_0 = 2\sqrt{a_0}$, the pump depletion length is given by $L_{pd} = (k_0/k_p)^2 c\tau_0$ [1]. Thus, the minimum number of pulses required for near-uniform acceleration may be estimated as

$$N_{\min} \approx \begin{cases} L_D/z_R, & a_0 \lesssim 0.5 \\ L_D/L_{pd}, & a_0 \gtrsim 2 \end{cases}. \quad (6.12)$$

In most cases of relevance, $z_R \ll L_{pd}$, such that a DFF would require many more pulses in the linear regime than the nonlinear regime.

A highly relativistic electron ($v_z \approx c$) that experiences a nearly uniform accelerating field over a distance L will gain an energy $m_e c^2 \Delta\gamma \approx -eE_z L$. In the linear and nonlinear regimes,

$$\Delta\gamma \approx \begin{cases} \frac{\pi}{8} k_p L a_0^2, & a_0 \lesssim 0.5 \\ \frac{1}{2} k_p L \sqrt{a_0}, & a_0 \gtrsim 2 \end{cases}. \quad (6.13)$$

With a conventional pulse, L is limited to a dephasing length: either $k_p L_d = \pi(k_0/k_p)^2$ in the linear regime or $k_p L_d = \frac{4}{3}(k_0/k_p)^2 \sqrt{a_0}$ in the nonlinear regime [1]. While the dephasing length can be increased by using a density gradient [219], the acceleration length would still be limited by pump depletion, which also scales as k_0^2/k_p^2 . Thus, for fixed a_0 and k_0 in the conventional case, increasing the energy gain in a single stage, $\Delta\gamma_C \propto n_0^{-1}$, requires operating at lower electron densities with weaker accelerating fields, $E_z \propto \sqrt{n_0}$, and longer stages, $L_d \propto n_0^{-3/2}$.

With the DFF, $L = L_D \approx N\delta f$, which is limited only by the number of pulses when δf is fixed. As a result, the single-stage energy gain can be increased without resorting to lower densities and weaker fields by using more pulses. For $\delta f = z_R$ in the linear regime and $\delta f = L_{pd}$ in the nonlinear regime,

$$\Delta\gamma_D \approx \begin{cases} \frac{\pi}{8} k_p z_R N a_0^2, & a_0 \lesssim 0.5 \\ \frac{k_0^2}{2k_p} c\tau_0 N \sqrt{a_0}, & a_0 \gtrsim 2 \end{cases}. \quad (6.14)$$

In this configuration, the energy gain increases linearly with the number of pulses N and has a more favorable scaling with the density compared to a conventional pulse: $\Delta\gamma_D \propto \sqrt{n_0}$ and $\propto 1/\sqrt{n_0}$ in the linear and nonlinear regimes, respectively. In practice, $k_p w_0 \lesssim 1$ in the linear regime for reasonable beam loading efficiency [112]. Furthermore, $c\tau_0 \approx \sqrt{a_0} k_p^{-1}$ in the nonlinear regime. Under these conditions, the energy gain scales as $\Delta\gamma_D \propto N/\sqrt{n_0}$ in the linear regime and $\Delta\gamma_D \propto N/n_0$ in the nonlinear regime.

By eliminating dephasing, LWFA driven by a DFF can operate at higher densities where the accelerating field is stronger: $E_z \propto \sqrt{n_0}$. This allows for (i) larger energy gains over the same length, (ii) equal energy gains over shorter lengths, or (iii) anything in between. Option (i) would minimize the number of stages required to reach a target energy. In a single stage of fixed length L , the ratio of energy gains attainable by a DFF and conventional pulse with the same a_0 and k_0 is given by $\Delta\gamma_D/\Delta\gamma_C = k_{pD}/k_p$ where $k_{pD} = (e^2 n_{0D}/m_e \varepsilon_0 c^2)^{1/2}$ and n_{0D} is the design density for LWFA driven by a DFF, which can be much higher than in LWFA with a conventional pulse [37]. For a conventional pulse, the value of k_p can be related to the dephasing length as

$$k_p = \begin{cases} k_0 \left(\frac{\pi}{k_0 L_d} \right)^{1/3} & a_0 \lesssim 0.5 \\ k_0 \left(\frac{4\sqrt{a_0}}{3k_0 L_d} \right)^{1/3} & a_0 \gtrsim 2 \end{cases}. \quad (6.15)$$

This leads to the useful expression [37]:

$$\frac{\Delta\gamma_D}{\Delta\gamma_C} \approx \left(\frac{k_{pD}}{k_0}\right) \begin{cases} \left(\frac{k_0 L}{\pi}\right)^{1/3} & a_0 \lesssim 0.5 \\ \left(\frac{3k_0 L}{4\sqrt{a_0}}\right)^{1/3} & a_0 \gtrsim 2 \end{cases}. \quad (6.16)$$

As an example, in a single stage with $L = 32$ cm, $n_{0D} = 3 \times 10^{18}$ cm $^{-3}$, $a_0 = 4$, and $\lambda = 2\pi c/\omega_0 = 800$ nm, $\Delta\gamma_D \approx 4\Delta\gamma_C$ —the energy gain attainable with a DFF is approximately $4\times$ higher than that attainable with a conventional pulse. The advantage of the DFF would be even larger at higher densities or with longer stages, albeit at the cost of more pulses. We note that while the ratio of energy gains increases as the density increases, the amount of charge that can be loaded into the wakefield decreases as $n_{0D}^{-1/2}$, so there is a balance between acceleration gradient and the amount of charge that can be loaded into the wakefield [1, 99, 118].

The number of pulses composing a DFF N also controls the amount of overlap between pulses (for fixed Δ_j and τ_j). When the pulses overlap significantly, the profile of the overall envelope [Eq. (6.7)] is sensitive to the relative phases of the pulses and may be different than that of the individual pulses. For example, the effective spot size w_{eff} may be larger than the spot size of each pulse w_0 . The pulses begin to overlap when their spacing is comparable to the pulse duration, i.e., $\Delta_{j+1} - \Delta_j \approx c\tau_0$. The number of pulses beyond which overlapping occurs may be estimated as T/τ_0 , where $T = \Delta_N - \Delta_1 \approx (1 - \frac{v_g}{c}) L/c$ is the total duration of the pulse, to obtain

$$N_{\text{overlap}} = \left(1 - \frac{v_g}{c}\right) \frac{L}{c\tau_0}. \quad (6.17)$$

In the nonlinear regime, $N_{\text{overlap}} \approx \frac{1}{2}N_{\text{min}}$. This shows that the pulses at least partially overlap in the nonlinear regime if the condition $N > N_{\text{min}}$ is satisfied, regardless of the pulse or plasma parameters. In the linear regime, $N_{\text{overlap}} \approx \frac{1}{4}(k_p^2 w_0^2 / \omega_0 \tau_0) N_{\text{min}}$. Here, the degree of overlap depends on the pulse and plasma parameters. In either case, overlap can be mitigated by choosing the polarization of successive pulses to be orthogonal, which effectively doubles N_{overlap} .

6.3 Simulations: electron acceleration

To demonstrate that a DFF can eliminate dephasing in LWFA, simulations were performed using the quasistatic, quasi-3D particle-in-cell (PIC) code QPAD [39]. QPAD models the evolution of the plasma electrons in response to the ponderomotive force of a laser pulse, as well as the electric and magnetic fields generated by the resulting charge separation. The induced electron current, in turn, feeds back onto the propagation of the pulse through the relativistic nonlinearities and ponderomotive modifications to the electron density profile. Further details regarding the algorithm and numerical parameters can be found in Sec. 6.5. Comparisons of QPAD to the full (i.e., non-quasistatic) PIC code OSIRIS [220] are presented in Sec. 6.6.

6.3.1 Linear Regime

Figure 6.2 shows that dephasing can be circumvented in the linear regime by using a DFF and provides insight into the DFF and plasma wave structure. A DFF with $N = 250$ and $a_0 = 0.1$ drives a plasma wave with $v_w = c$ over 20 dephasing lengths, $L = 20L_d$ (see Table 6.1 for other parameters). The absence of dephasing and the velocity $v_w = c$ are evident in Fig. 6.2(a) by the nearly identical structures and locations of the pulse and plasma wave in ξ , i.e., neither recedes in ξ over 20 dephasing lengths of propagation. This indicates that a highly relativistic electron ($v_z \approx c$) initially located in the maximum accelerating field will remain in that field over the entire length of the accelerator.

Figures 6.2(b) and (c) display the transverse structure of the DFF pulse and plasma wave. Both the DFF pulse and plasma wave maintain near-constant transverse profiles throughout propagation, which is critical to ensuring that an ultra-relativistic electron experiences a constant focusing force as it accelerates. The coherent superposition of the $N = N_{\min} \gg N_{\text{overlap}}$ pulses produces a continuous DFF envelope with an “X”-shaped structure, arising from the slice-dependent focal length. The temporal precursor of the DFF

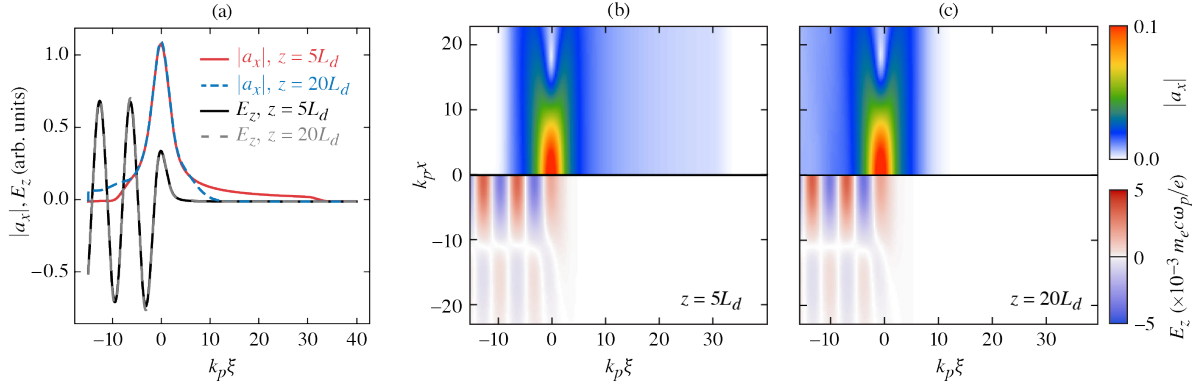


Figure 6.2: Evolution of a discrete flying focus and driven plasma wave in the linear regime ($a_0 = 0.1$). (a) The on-axis ($|\mathbf{x}_\perp| = 0$) envelope of the DFF pulse (red and blue) and longitudinal electric field of the plasma wave (black and gray) have nearly identical profiles in $\xi = z - ct$ after the pulse has propagated 5 (solid) and 20 (dashed) dephasing lengths. (b,c) The spatiotemporal structure of the DFF envelope (top) and longitudinal field (bottom) are also nearly identical at these propagation distances. From left to right in ξ , the variation in the transverse width of the DFF (i.e., the X structure) results from ξ slices that have already passed through focus, are at or near focus, and have yet to come into focus, respectively. This slice-dependent focusing is also the source of the pre and postcursors on the envelopes in (a). In this example, $N = 250 = N_{\min} \gg N_{\text{overlap}}$. Due to significant overlap between individual pulses, the spot size of the overall envelope $k_p w_{\text{eff}} = 9$ is larger than that of the individual pulses $k_p w_0 = 4$.

pulse, visible in Figs. 6.2(a) and (b), gradually recedes as each pulse comes into focus and is replaced by the “postcursor” visible in Figs. 6.2(a) and (c).

Figures 6.3(a)–(c) show how the structure of the accelerating field (E_z) depends on the number of pulses composing a DFF in the linear regime. With a single pulse [Fig. 6.3(a)], the on-axis ($|\mathbf{x}_\perp| = 0$) accelerating field slides backward in ξ and quickly drops in magnitude. These are direct results of the pulse driving a plasma wave with $v_w = v_g < c$ and quickly diffracting, respectively. The dashed line in the inset marks a trajectory traveling at v_g over

Laser parameters	Value	Normalized
λ_0	800 nm	$2\pi/25$
k_0	$7.8 \mu\text{m}^{-1}$	25
τ_0	16 fs	1.5
w_0	13 μm	4
z_R	640 μm	200
L_d	0.64 cm	2000
L_{pd}	0.32 cm	1000
Discrete flying focus parameters ($a_0 = 0.1$)		
N	—	40, 250
N_{min}	—	250
N_{overlap}	—	25
a_{0j}	—	0.1, 0.024
$f_{j+1} - f_j$	4 mm, 0.6 mm	1250, 200
$\Delta_{j+1} - \Delta_j$	3.2 μm , 0.5 μm	1, 0.16
Discrete flying focus parameters ($a_0 = 4$)		
N	—	20, 40
N_{min}	—	50
N_{overlap}	—	25
a_{0j}	—	4
$f_{j+1} - f_j$	8 mm, 4 mm	2500, 1250
$\Delta_{j+1} - \Delta_j$	6.4 μm , 3.2 μm	2, 1
Plasma parameters		
n_0	$3 \times 10^{18} \text{ cm}^{-3}$	—
k_p^{-1}	3.2 μm	1
L	16 cm	50,000

Table 6.1: Parameters for the simulations presented in Figs. 6.2 and 6.3. The normalized vector potential of each pulse a_{0j} is scaled to achieve the desired a_0 for the overall envelope. Note that for the profile choices defined in Eqs. (6.8) and (6.9), $a_{0j} = \mathcal{A}_{0j}$. The values for $f_{j+1} - f_j$ and $\Delta_{j+1} - \Delta_j$ are respectively listed for the $N > 1$ cases. Time and space are normalized by ω_p^{-1} and k_p^{-1} where indicated.

a single dephasing length. Thus, even if diffraction could be counteracted (for instance with a plasma waveguide), dephasing would still occur over a distance consistent with Eq. (6.6) ($k_p L_d \approx 2000$).

When the number of pulses is increased to $N = 40 < N_{\min}$, the accelerating field persists over the entire propagation distance of $25L_d$ [Fig. 6.3(b)]. However, the phase velocity of the plasma wave v_w evolves irregularly, and the longitudinal field exhibits discrete gaps due to the large separation between the focal points of each pulse. In this case, each pulse contributes to the drive over a long enough distance that spatiotemporal couplings develop in its profile. This invalidates the separable form of Eq. (6.5) (see Sec. 6.4) and contributes a temporal broadening that smooths the longitudinal field at larger distances in Fig. 6.3(b). Despite this smoothing, a constant accelerating field over the entire acceleration distance still requires $N \gtrsim N_{\min}$.

Figure 6.3(c) demonstrates that a DFF with $N = N_{\min} = 250$ pulses in the linear regime can excite a smooth and constant accelerating field with $v_w = c$. The parameters are the same as those in Fig. 6.2. A highly relativistic electron ($v_z \approx c$) initially collocated with the maximum accelerating field [blue stripe at $k_p \xi \approx -3$ in Fig. 6.3(c)] would travel along a vertical line in $\xi - z$ space and experience the maximum accelerating field over the entire 25 dephasing lengths.

6.3.2 Nonlinear Regime

While the linear regime provides important insights into the benefit of LWFA driven by a DFF, the nonlinear regime ($a_0 \gtrsim 2$) offers several key advantages. First, the accelerating field is much higher, allowing for significantly larger energy gains in a single stage. Second, the focusing fields are much stronger and linear in radius, which helps preserve the emittance. Finally, the DFF can be composed of far fewer pulses, which would reduce the complexity of an experimental implementation. This latter advantage arises because self-guiding allows each pulse in the DFF to strongly drive the plasma wave over its depletion length, rather

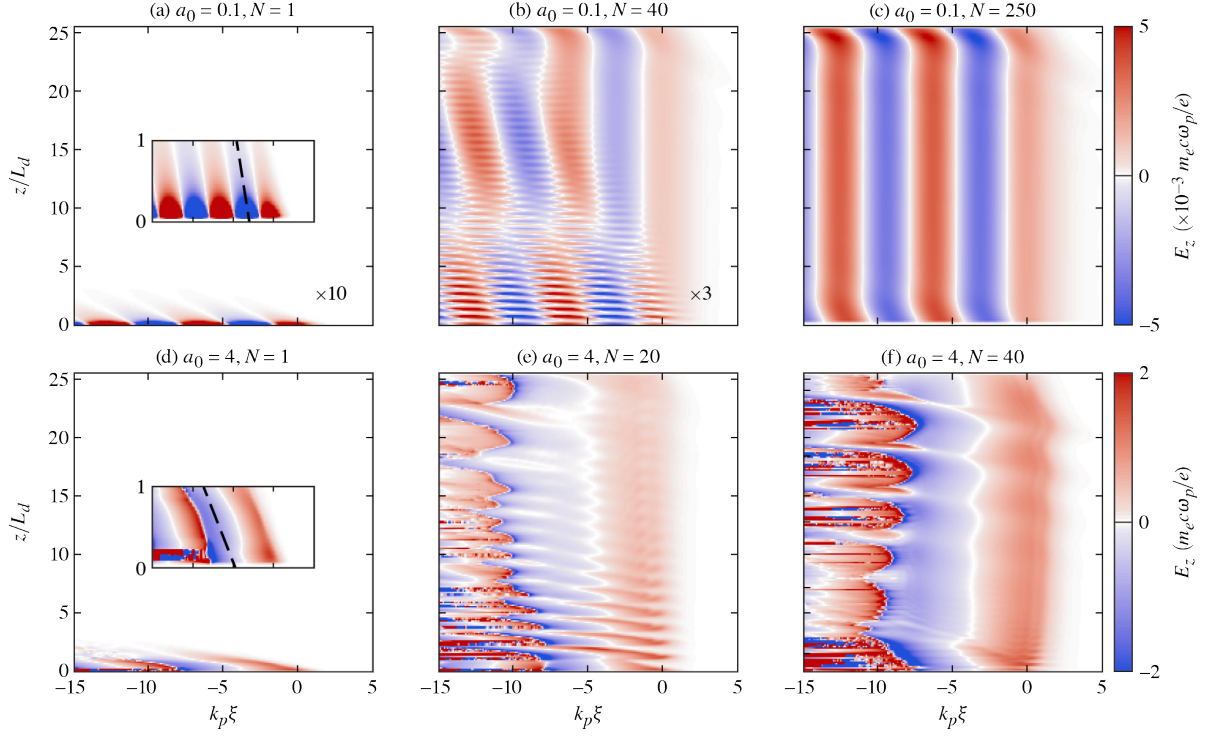


Figure 6.3: Dependence of the on-axis ($|\mathbf{x}_\perp| = 0$) longitudinal field E_z on the number of pulses driving the plasma wave in the (a,b,c) linear and (d,e,f) nonlinear regimes. (a) With a single pulse ($N = 1$) in the linear regime ($a_0 = 0.1$), the maximum accelerating field (dashed black line) recedes and drops in amplitude due to the subluminal group velocity and diffraction of the pulse. (b) For a DFF composed of $N = 40 < N_{\min} = 250$ pulses, the plasma wave persists over the entire propagation range of 25 dephasing lengths, but the large separation between pulses results in an irregular longitudinal field. (c) A DFF with $N = 250 = N_{\min}$ pulses drives a plasma wave with a smooth longitudinal field and a phase velocity $v_w = c$. (d) With a single pulse ($N = 1$) in the nonlinear regime ($a_0 = 4$), the maximum accelerating field (dashed black line) recedes and drops in amplitude due to a combination of the subluminal group velocity and depletion. (e,f) Similar to the linear regime, a DFF composed of too few pulses ($N = 20 < N_{\min} = 50$) produces an irregular field, while a DFF composed of a sufficient number of pulses ($N = 40 \approx N_{\min}$) produces a smooth field.

than being limited to the Rayleigh range as in the linear regime.

Figures 6.3(d)–(f) illustrate how the structure of the on-axis accelerating field depends on the number of pulses composing the DFF in the nonlinear regime. With a single pulse [Fig. 6.3(d)], the accelerating field slides backwards in ξ , as it did in the linear regime. However, in contrast to the linear regime, the maximum amplitude of the accelerating field is relatively constant over a single dephasing length (see inset). This is because the individual pulse parameters were chosen to be nonlinearly matched for self-guiding ($k_p w_0 = 2\sqrt{a_0}$), which prevents the drop in amplitude due to linear diffraction. Instead the amplitude drops over a depletion length $L_{\text{pd}} = L_d/2 \approx 10z_R$ as the pulse transfers its energy to the plasma wave.

The dashed line in the inset of Fig. 6.3(d) marks the trajectory of the accelerating field driven by a single pulse in the nonlinear regime. The accelerating field slips backward $2.5\times$ more rapidly than in the linear regime [see Fig. 6.3(a) inset]. This is consistent with Ref. [1], which predicts a $3\times$ higher rate of slippage due to etching, i.e., local depletion at the front of the pulse. Although the rate of slippage is higher, the nonlinear dephasing length is similar to the linear dephasing length ($k_p L_d \approx 2000$) because the wavelength of the plasma wave is longer in the nonlinear regime.

As in the linear regime, increasing the number of pulses in the nonlinear regime provides a smoother and more-constant accelerating field that persists over a longer distance [Figs. 6.3(e) and (f)]. However, the evolution of the accelerating field is now more complex. For $N = 20 < N_{\text{min}}$, the large separation between the pulses results in a longitudinal field with a modulated period [Fig. 6.3(e)]. In the nonlinear regime, the period of the plasma wave depends on the amplitude and spot size of the driving pulse, both of which oscillate as each pulse comes into focus and then depletes. At larger z , the modulations are partially smoothed due to the development of space-time couplings in the profile of each pulse, which increases their overlap.

Figure 6.3(f) demonstrates that a DFF with $N = 40 \approx N_{\text{min}}$ pulses in the nonlinear

regime can excite a smooth and constant accelerating field with $v_w = c$. Because the overall pulse is more uniform, the modulations in the accelerating field observed with $N = 20$ are mostly absent. In principle, the residual modulations could be mitigated by optimizing the parameters of each pulse, such as the relative delays, focal locations, durations, or spot sizes. This is a topic for future investigation. For the examples in Figs. 6.3(e) and (f), the polarization alternated between the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ directions from one pulse to the next. This reduces the interference between pulses and was found to produce a more stable accelerating structure.

With regards to the striations observable at smaller ξ in Figs. 3(d)–(f), QPAD is accurate within the first period of the plasma wave, but can deviate from full OSIRIS simulations in the nonlinear regime after the first period (see Sec. 6.6 for comparisons). This is a consequence of the quasistatic approximation, which assumes that the background electrons have a velocity less than v_w . These deviations, however, do not impact the accelerating field experienced by a separate population of highly relativistic electrons located in the first period.

6.3.3 Acceleration in the Nonlinear Regime

The plasma wave driven by a DFF in the nonlinear regime can accelerate an electron beam over many dephasing lengths while preserving the initial emittance of the beam. To demonstrate this, QPAD simulations were conducted with an axially uniform 48-pC, 1-GeV monoenergetic electron beam externally injected into a plasma wave driven by a DFF with $a_0 = 4$, $N = 80$, and $L = 50L_d = 32$ cm. These are the same DFF parameters used for the example in Sec. II and for Fig. 3(f), but with twice the length. *The length $L = 50L_d$ was chosen to accommodate demands on computational resources and does not represent a physical limitation of DFF-driven LWFA.*

Figure 6.4 shows that the DFF-driven plasma wave [Fig. 6.4(a)] accelerates the injected electrons from 1 GeV to an average energy of $m_e c^2 \bar{\gamma} = 41$ GeV [Fig. 6.4(b)]. This is $3\times$ higher than the energy achievable with a conventional pulse in a lower-density stage of the same

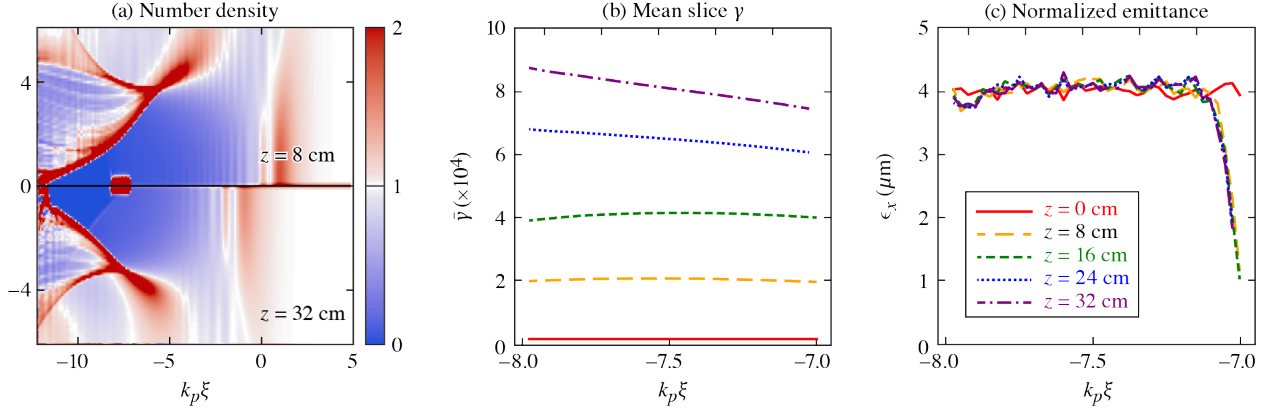


Figure 6.4: Acceleration of a 50 pC electron beam from 1 GeV to 41 GeV over 32 cm, or 50 dephasing lengths, using a discrete flying focus with tailored delays. (a) Electron density of the background plasma and electron beam normalized to n_0 . The beam profile and accelerating structure are preserved over the entire length of the accelerator. Note that the beam density saturates the color map. (b) Mean slice energy of the beam. At the end of the accelerator ($z = 32$ cm), the spectrum is nearly monoenergetic with an energy spread $\sigma_\gamma/\bar{\gamma} = 0.04$. (c) Normalized slice emittance. The emittance does not grow as the beam is accelerated.

length. Over the entire beam, the acceleration preserves the normalized slice emittance at its initial value of $\epsilon_x = 4 \mu\text{m}$ [Fig. 6.4(c)]. The average slice energy develops a slight positive chirp, which contributes to a final energy spread of $\sigma_\gamma/\bar{\gamma} = 0.04$. The slice energy spread, however, is significantly lower: $\overline{\sigma_\gamma(\xi)/\bar{\gamma}(\xi)} = 0.001$. The slight decrease in the spot size of the beam, observable in Fig. 6.4(a), is consistent with the theoretically predicted adiabatic evolution $\sigma_x(\gamma) \propto \gamma^{-1/4}$ [221].

For this example, the parameters of each pulse composing the DFF were tailored to prevent the injected beam from interacting with the rear sheath of the plasma wave. The delays Δ_j of the first two pulses were shifted back relative to Eq. (6.10) by $2k_p^{-1}$ and $1k_p^{-1}$, respectively. All subsequent pulses had delays given by $\Delta_j = [\frac{1}{2}(k_p/k_0)^2 - S] f_j$, where $f_j = 1250(j-1)k_p^{-1} \approx (5/4)(j-1)L_{\text{pd}}$. Here, $S = 1.25 \times 10^{-5}$ is a correction to the linear

group velocity factor $1 - v_g/c \approx \frac{1}{2}(k_p/k_0)^2 = 8 \times 10^{-4}$ that corresponds to an additional slippage of $1 \text{ } k_p^{-1}$ over $40L_d$ of propagation. These customized parameters were chosen to compensate a slight advance in the accelerating field observed in a separate QPAD simulation with uniformly spaced pulses. With uniform spacing, most of the beam was lost: the rear sheath of the plasma wave repeatedly swept over and defocused the beam. The advance of the accelerating field is visible for the analogous simulation with $L = 25L_d$ shown in Fig. 6.3(f).

The initial normalized emittance of the beam, $\epsilon_x = 4 \text{ } \mu\text{m}$, was chosen to satisfy the emittance matching condition for the initial spot size $k_p\sigma_x = 0.2$. This emittance value is relatively high in the context of linear collider designs. While the matched emittance can be decreased by using a smaller spot, the spot was limited in the simulations by computational demands on the transverse resolution $k_p\Delta x = 0.1$. To examine the emittance growth of a lower-emittance beam, QPAD simulations were also run with $L = 5L_d$ (i.e., $10\times$ shorter distance), a $10\times$ higher transverse resolution of $k_p\Delta x = 0.01$, a spot size of $\sigma_x = 0.02k_p^{-1} = 65 \text{ nm}$, and a matched normalized emittance of $\epsilon_x = 0.0126k_p^{-1} = 40 \text{ nm}$. These simulations also exhibited negligible emittance growth. This suggests that emittance growth would be negligible over a distance $L = 50L_d$ if the beam were initialized with a collider-quality emittance.

The DFF and electron beam parameters were chosen to demonstrate the concept and were not optimized for accelerated charge or efficiency. Nevertheless, it is still instructive to calculate the efficiency. The initial energy in the overall pulse was 150 J, while the total energy gain of the electron beam was 1.9 J, yielding an efficiency of 1.3%. This efficiency is low for two reasons. First, the beam needs to be in front of the rear sheath of the plasma wave because the bubble size oscillates, as seen in Fig. 6.3(f). Second, the blowout is incomplete, i.e., the background electrons have not been completely expelled in the first period of the plasma wave. If these issues can be resolved through optimization of the pulse train, the efficiency could be significantly higher. Equation 6 of Ref. [99] indicates that a total charge of $Q \approx 400 \text{ pC}$ with an optimized trapezoidal beam profile can be loaded in a fully blown out wave at $E_z = -m_e c \omega_p / e$, which would increase the efficiency to 10%.

6.4 Higher-Order Effects of Linear Laser Propagation

The laser profiles in Eqs. 6.5 and 6.7 were derived by dropping three terms from Eq. (6.3): the mixed derivative $\partial_s \partial_\zeta$, the dispersive term $(1 - v_g/c)^2 \partial_\zeta^2$, and the ∂_s^2 term. Each of these terms has a corresponding length scale $L_{s\zeta}$, $L_{\zeta\zeta}$, L_{ss} that determines when higher-order corrections to the approximate solution in Eq. (6.5) may be important. These length scales can be estimated for a single pulse following the multiscale expansion approach in Ref. [7]:

$$L_{s\zeta} = \frac{1}{2} c \tau_0 k_0^2 w_0^2 \quad (6.18)$$

$$L_{\zeta\zeta} = \frac{1}{2} \left(\frac{k_0}{k_p} \right)^2 k_0 (c \tau_0)^2 \quad (6.19)$$

$$L_{ss} = \frac{1}{2} k_0^3 w_0^4 \quad (6.20)$$

For example, the calculation for $L_{\zeta\zeta}$ is shown in detail here. We consider a pulse with the initial condition $a(\zeta, 0, \mathbf{x}_\perp) = \exp(-\mathbf{x}_\perp^2/w_0^2) \exp(-\zeta/c\tau_0)$ at $s = 0$. To zeroth order, we have $[\nabla_\perp^2 + 2ik_0 \partial_{s_0}]a = 0$. This determines the zeroth order solution $a(\zeta, s, \mathbf{x}_\perp) \approx \exp(-\zeta^2/(c\tau_0)^2) \left[\frac{z_R}{q(s)} \right] \exp \left[-\frac{ik_0 |\mathbf{x}_\perp|^2}{c\tau_0} \right]$ (i.e., Eq. 6.5). To first order, assuming that the $(1 - v_g/c) \partial_\zeta^2$ dominates the other first-order correction terms, we have $[(1 - v_g/c) \partial_\zeta^2 + 2ik_0 \partial_{s_1}]a = 0$. From the zeroth order solution, one may evaluate $\partial_\zeta^2 a \approx -\frac{2}{(c\tau_0)^2}$ at $(\zeta, s, \mathbf{x}_\perp) = (0, 0, \mathbf{0})$. Also, $1 - v_g/c \approx \frac{1}{2} \left(\frac{k_p}{k_0} \right)^2$. Hence, one obtains $|\partial_{s_1} a| \approx \frac{1}{2k_0} \left(\frac{k_p}{k_0} \right)^2 \frac{1}{(c\tau_0)^2}$. Finally, estimating $L_{s\zeta} \approx 1/|\partial_{s_1} a|$, one obtains the estimate in Eq. 6.19.

The $L_{s\zeta}$ and L_{ss} length scales are the same in vacuum and plasma. For a pulse initialized according to Eq. (6.5), $L_{s\zeta}$ is the length scale over which the envelope will develop non-separable dependencies on the temporal and transverse coordinates. Beyond the $L_{\zeta\zeta}$ length scale, group velocity dispersion causes substantial temporal broadening of the pulse, which reduces its amplitude. Finally, when L_{ss} is comparable to the Rayleigh range $z_R = \frac{1}{2} k_0 w_0^2$ (i.e., when $k_0^2 w_0^2 \approx 1$), non-paraxial corrections will cause the transverse profile to deviate from a solution to the paraxial wave equation.

These estimates apply to a single pulse. Due to interference, the evolution of the pulse

train is more complex. Nonetheless, the length scales for a single pulse provide estimates for when higher-order linear effects can contribute. When they do contribute, it may be necessary to tailor the pulse train to compensate. For example, the $L_{s\zeta}$ term results in a non-ideal radial group delay, where the amount of slippage depends on radius. This can be compensated by adjusting the Δ_j for each pulse. The correction can be estimated from geometric optics as $\Delta_j = \Delta_j^{(0)} + \Delta_j^{(1)}$, where $\Delta_j^{(0)}$ is the right-hand side of Eq. (6.10) and the first-order correction is given by

$$\Delta_j^{(1)} = \frac{f_j}{k_0^2 w_0^2}. \quad (6.21)$$

The simulation results in Section 6.3 can be interpreted in light of these considerations. The single-pulse length scales using the parameters in Table 6.1 are $k_p L_{s\zeta} = 7500$, $k_p L_{\zeta\zeta} = 2 \times 10^4$, and $L_{ss} = 10^4 z_R$. The correction in Eq. (6.21) is consistent with the evolution of the phase velocity in the case with $a_0 = 0.1$ and $N = 40$ shown in Fig. 6.3(b), where there is a large separation between the individual pulses. For the case with $a_0 = 0.1$ and $N = 250$ shown in Fig. 6.3(c), the phase fronts do not slip backward even though $L > L_{s\zeta}$ and $L > L_{\zeta\zeta}$. In this case, there is significant overlap between the pulses. The resulting interference produces an overall pulse with an effective spot size $k_p w_{\text{eff}} = 9$ and duration $\omega_p \tau_{\text{eff}} = 4$ that are larger than those of the individual pulses. Using these values in Eqs. (6.18) and (6.19), the length scales become $k_p L_{s\zeta} = 10^5$ and $k_p L_{\zeta\zeta} = 1.25 \times 10^5$ so that higher-order effects are negligible over the simulated length scale $k_p L = 5 \times 10^4$. The $a_0 = 4$ cases in Sec. 6.3 demonstrate that despite propagation beyond $L_{s\zeta}$ and $L_{\zeta\zeta}$ and some distortion to the profile in Eq. (6.7), it is possible to obtain good results without additional compensation to the delays Δ_j given by Eq. (6.10). However, additional tailoring could potentially improve the results.

6.5 Simulation details

QPAD is a quasistatic PIC code [39, 81]. Quasistatic PIC codes take advantage of several physics approximations to reduce computational expense: (i) the laser pulse or particle beam

driver does not evolve in the time it takes a slice of plasma particles to pass through it; (ii) all relevant waves have phase velocities near the speed of light; and (iii) in the case of a laser pulse, laser-cycle averaged (or “ponderomotive guiding center”) equations of motion can be used for the background electrons [26]. With these approximations, the full set of Maxwell’s equations are simplified to a set of Poisson-like equations for the potentials of the plasma wave and a nonlinear wave equation for the envelope of the laser pulse. The quasi-3D algorithm expands the fields, charge and current densities, and laser envelope into azimuthal modes. These quantities are calculated on grids in the coordinates (r, ξ) , and the expansion is truncated at a desired mode number. For the electron acceleration simulations presented in this work, only mode 0 was used.

In all cases considered in this study, a uniform plasma was simulated. The radius of the plasma R_p was large enough to fit the spot size $w_{\max}$ of the most out-of-focus laser pulse at $z = 0$. For the simulations with $k_p L = 50,000$ in Figs. 6.2 and 6.3, the most out-of-focus spot size was $w_{\max} \approx (L/z_R)w_0 = 1000k_p^{-1}$. These simulations used $R_p = 3300k_p^{-1}$. For the simulation with an externally injected beam in Fig. 6.4, the plasma radius was $R_p \approx 6600k_p^{-1} \approx 2$ cm. The initial Lorentz factor of the beam was $\gamma_0 = 2000$ so that $m_e c^2 \gamma_0 \approx 1$ GeV. The beam was cylindrically symmetric with a spot size $\sigma_x = 0.2k_p^{-1} = 650$ nm and a matched normalized emittance $\epsilon_x = k_p \sigma_x^2 (\gamma_0/2)^{1/2} = 1.26k_p^{-1} = 4 \mu\text{m}$ [221]. The current profile was constant in ξ with a length $k_p L_b = 1$. The resolution in all cases was $k_p \Delta z = 0.01$, $k_p \Delta x_{\perp} = 0.1$, and $\omega_p \Delta t = 10$. These values were chosen to ensure numerical convergence in the cases where $a_0 = 4$.

6.6 Comparison of nonlinear LWFA in QPAD and OSIRIS

In order to demonstrate the accuracy of QPAD, reduced-scale simulations were performed to compare QPAD and the fully electromagnetic PIC code OSIRIS [220] using identical laser pulse and plasma parameters. The OSIRIS simulations were run using the quasi-3D

azimuthal decomposition [222] and the field solver introduced in Ref. [223] to mitigate numerical dispersion. The transverse resolution is the same between the two simulations ($k_p \Delta x_\perp = 0.1$), while a finer longitudinal resolution of $k_p \Delta z = 0.008$ is used in OSIRIS. Remarkable agreement between the electron densities and the electric field envelopes of the laser pulses is evident in Fig. 6.5. The physical parameters correspond to the nonlinear ($a_0 = 4$) cases presented in Table I, but with $N = 1$ [Figs. 6.5(a) and (b)] or $N = 9$ [Figs. 6.5(c) and (d)].

Figures 6.5(a) and (b) show the electron densities and electric field envelopes of the laser pulses from (a) QPAD and (b) OSIRIS for $N = 1$ at $k_p z = 100 \ll k_p L_{pd}$. Within the first period of the plasma wave, the structure of the waves and envelopes agree. For $k_p \xi \lesssim -7$, the QPAD results deviate from the OSIRIS results: In the nonlinear regime, background electrons in the rear sheath of the wave can be accelerated to velocities close to v_w , which causes the quasistatic approximation to break down.

Figures 6.5(c) and (d) show the electron densities and electric field envelopes of DFF pulses from (c) QPAD and (d) OSIRIS for $N = 9$ at $k_p z = 9000$. A major difference from the $N = 1$ case seen in both the OSIRIS and QPAD simulations is the highly modulated and lengthened envelope of the pulses. As the individual pulses in the DFF deplete, they are locally redshifted. The lower group velocity of the redshifted light causes it to slide back into the plasma wave structure. This redshifted light can become trapped in the wave and modify its structure—for example, by preventing closure of the rear sheath, as seen in both Figs. 6.5(c) and (d). For longer stages, the redshifted light can cause multiple cycles of sheath opening and closing, as observed in Fig. 6.3(f). The local wavenumber of the light (not shown) drops from $k(\xi) = k_0 = 25k_p$ at $\xi = 0$ to $k(\xi) = 12k_p \approx k_0/2$ at $k_p \xi \approx -6$. To obtain good agreement between OSIRIS and QPAD, the longitudinal resolution $k_p \Delta \xi = 0.01$ in the QPAD simulations was chosen to resolve this redshifting.

Another difference between the $N = 1$ and $N = 9$ cases is the more complete blowout in the $N = 1$ case. The incomplete blowout driven by the DFF (on-axis density $\approx 0.1 n_0$)

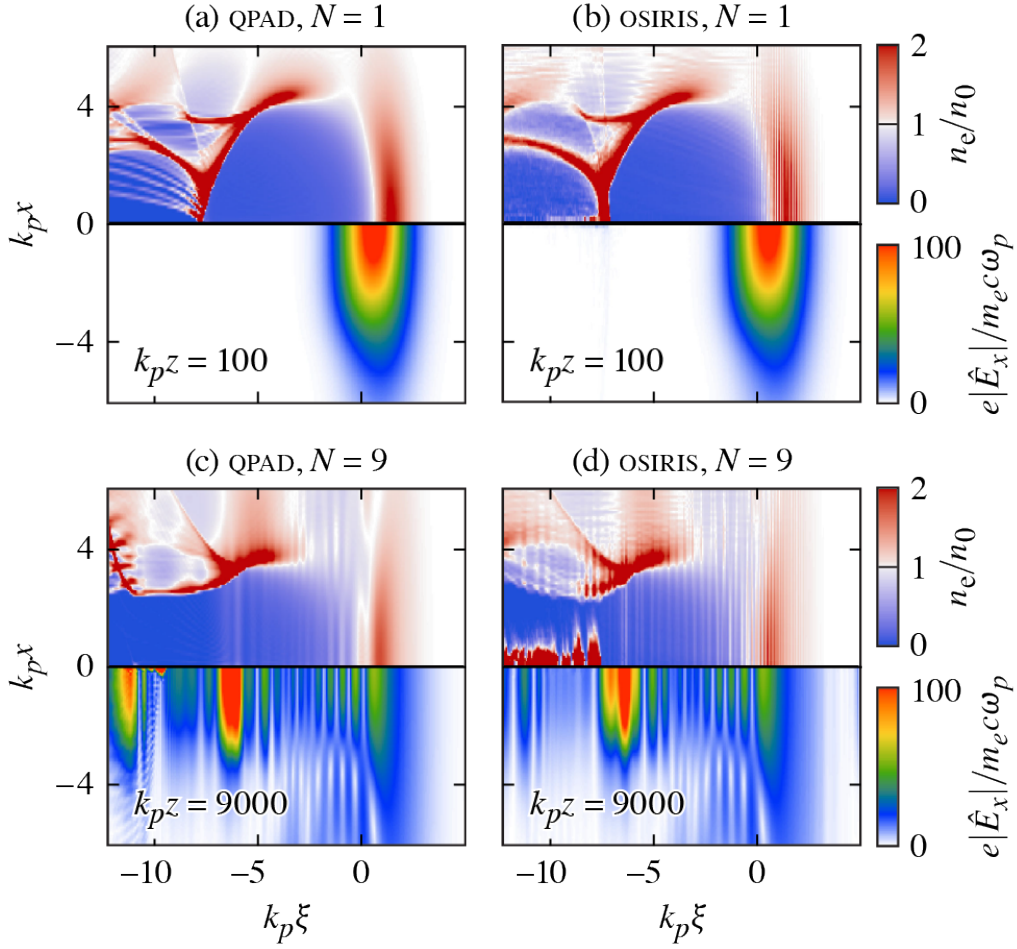


Figure 6.5: The electron densities and electric field envelopes from (left) QPAD and (right) OSIRIS simulations run with the same physical parameters. The results for a single pulse ($N = 1$) are shown in (a) and (b), while the results for a DFF with $N = 9$ pulses are shown in (c) and (d). The agreement between QPAD and OSIRIS is excellent. In (c) and (d), local depletion and redshifting causes the light to slip backward, where it is temporarily trapped by the plasma wave. This prevents the rear sheath from closing. Over longer distances, the sheath undergoes multiple cycles of opening and closing as the trapped light first accumulates and then escapes out the back. [See Figs. 6.3(e) and (f)]. The quasistatic model used in QPAD does not capture self-injection, which is observed in the OSIRIS simulations (d).

reduces the accelerating field and limits how much charge can be loaded in the plasma wave. The incomplete blowout is caused by a reduction in the DFF amplitude, which falls to about half its original amplitude. The amplitude reduction could potentially be addressed by further tailoring of the laser pulses.

The electron density from the $N = 9$ OSIRIS simulation features a 0.35-nC self-injected bunch [Fig. 6.5(d)]. Self-injection cannot be modeled within the quasistatic approximation [224]. Nevertheless, QPAD accurately models the laser pulse and plasma evolution in the region upstream of the self-injected bunch, where the dynamics are causally disconnected from the self-injected bunch. In this region, QPAD effectively cycle-averages the short-wavelength ($k \approx 2k_0$) oscillations in the plasma density observed in the OSIRIS simulations. Self-injection only occurs in the vicinity of the moving density spike at the rear of the sheath. Thus, self-injection does not alter the evolution of the externally injected beam in Fig. 6.4. However, the OSIRIS simulation suggests that the design presented in Fig. 6.4 could have dark current. This dark current could be mitigated by beam loading or tailoring the properties of the individual DFF pulses to stabilize the rear sheath.

Figures 6.5(c) and (d) also indicate that redshifted light would be spatially overlapped with an externally injected electron beam, as used in Fig 6.4. The coupling between this light and the electron beam (i.e., direct laser acceleration [225]) is not currently modeled in QPAD. In principle, direct laser acceleration could lead to an increase in the transverse momentum and emittance of the beam. In practice, the effect of direct laser acceleration on the beam is negligible: The beam betatron wavenumber k_β and the wavenumber of the redshifted light k_L are always far from the DLA resonance condition $k_\beta = k_L$. For the parameters used here, $k_L \approx 12k_p$ and $k_\beta(t) < k_\beta(0) = k_p(2\gamma_0)^{-1/2} = 0.016k_p$. To confirm that direct laser acceleration is negligible, the equations of motion for a subset of externally injected electrons were numerically solved using laser pulse and plasma wave fields derived from a phenomenological model, as in Ref. [62]. These calculations indicated that the emittance growth from the beam-laser interaction was negligible for the case in Fig. 6.4 with

$k_p L = 100,000$ when using a beam with a spot size $\sigma_{0x} = 0.02k_p^{-1} = 65$ nm and matched emittance $\epsilon_x = 0.0126k_p^{-1} = 40$ nm.

In Fig. 6.5, the electric field envelope for QPAD was calculated from the complex vector potential (amplitude and phase), while the electric field envelope for OSIRIS was extracted using a Hilbert transform.

6.7 Simulations: positron acceleration using an $\ell = 1$ Laguerre-Gaussian ASTRL driver

The examples in the previous section are not capable of accelerating positrons because the fields within the ion column are focusing only for negative charges. This happens because the laser intensity is peaked on-axis and expels electrons from the axis. Viera et al. introduced the idea of using a Laguerre-Gaussian pulse, which has an off-axis intensity peak, for acceleration of positrons in the nonlinear regime [34]. In this case, the ponderomotive force of the laser is directed radially inward for electrons with radii smaller than the intensity peak; for an appropriate laser amplitude, these electrons are compressed inward and form a filament on-axis which can provide focusing fields appropriate for positive charges. With a single pulse, the same limitations apply as in the case of electron acceleration: the laser pulse dephases and pump-depletes, constraining the length scale over which acceleration can occur. These issues are exacerbated by the weaker self-guiding of Laguerre-Gaussian modes. Using a discrete ASTRL flying focus with a Laguerre-Gaussian transverse profile for each mode can potentially address these issues. In this section, preliminary results are presented which suggest that this configuration can indeed create an accelerating structure for positrons which is maintained over many dephasing lengths.

6.7.1 Linear Regime

As before, we begin with the linear regime to provide intuition for the process. Figure 6.6 shows an example of a DFF with $(\ell, p) = (1, 0)$ Laguerre-Gaussian modes. The length scale is $L = 20,000k_p^{-1} \approx 10L_d$. The a_{0j} of each pulse is scaled so that $a_0 = 0.1$ at $k_p(s, \xi, \mathbf{x}_\perp) = (0, -16, 3)$. The parameters are otherwise the same as for the $N = 250$, $a_0 = 0.1$ case in Table 6.1 and Fig. 6.2, including the Δ_j and f_j sequences. The simulation uses modes up to $m = 1$, the only mode in which the laser envelope is nonzero. As desired, the laser amplitude peaks off-axis. Aside from this, the results are similar to those of the analogous simulation with a Gaussian transverse (i.e., $(\ell, p) = (0, 0)$ Laguerre-Gaussian) profile in Section 6.3: the linear wakefields are maintained over the range of propagation of the laser. In this case, the linear wakefields cannot accelerate positrons because they vanish on-axis.

6.7.2 Nonlinear Regime

Next, we consider the results in the nonlinear regime. Figure 6.7 shows a case with $N = 16$, $L = 20,000k_p^{-1} \approx 10L_d$, $(\ell, p) = (1, 0)$, and $a_0 = 20$ for each pulse. The parameters are otherwise the same as for the $a_0 = 4$, $N = 40$ case in Table 6.1. As in the cases with a Gaussian transverse profile, the structure of the laser and wake are much more complicated than in the linear regime. It is clear that at these two snapshots, the wake structure has not dephased and is suitable for the acceleration of positrons. The column of electrons along the axis results in electromagnetic fields of the right sign for acceleration and focusing of positrons along the axis.

Figure 6.8 shows the evolution of the on-axis accelerating field (E_z) and a transverse slice of the focusing field ($E_r - B_\phi$) at $k_p\xi = -20$. These plots demonstrate that for this simulation, the wake structure indeed persists over the 10 dephasing lengths of propagation. Recession of the wake is visible in the E_z lineout. The discreteness of the $N = 16$ pulses is

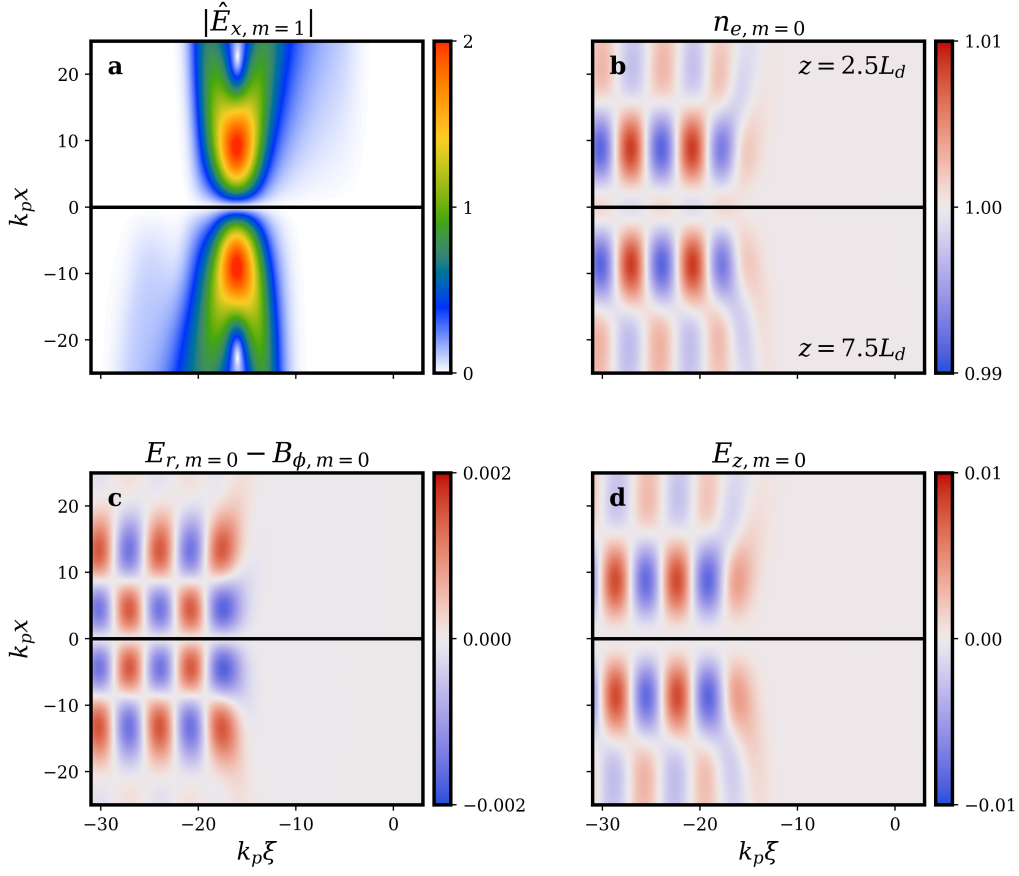


Figure 6.6: Linear wakefields driven by a flying focus described by an ASTRAL flying focus with $C_j(z, \mathbf{x}_\perp)$ given by an $(\ell, p) = (1, 0)$ Laguerre-Gaussian mode, $k_p w_0 = 4$ for each pulse, $k_p L = 20,000 \approx 10L_d$, and $N = 80$. Δ_j and f_j are chosen in the same way as for the electron acceleration cases. Top panels show the state at $z = 5000 \approx 2.5L_d$ and $z = 15,000 \approx 7.5L_d$. (a) Amplitude of the laser envelope for the $m = 1$ mode. The laser envelope peaks off-axis and retains the same structure throughout its propagation. The characteristic x-shape and precursor are visible, as with an $(\ell, p) = (0, 0)$ profile. (b-d) $m = 0$ mode of the electron number density, focusing field, and accelerating field. The structure is maintained over the 5 dephasing lengths for each quantity. However, this configuration cannot accelerate positrons because the focusing and accelerating fields vanish on the axis.

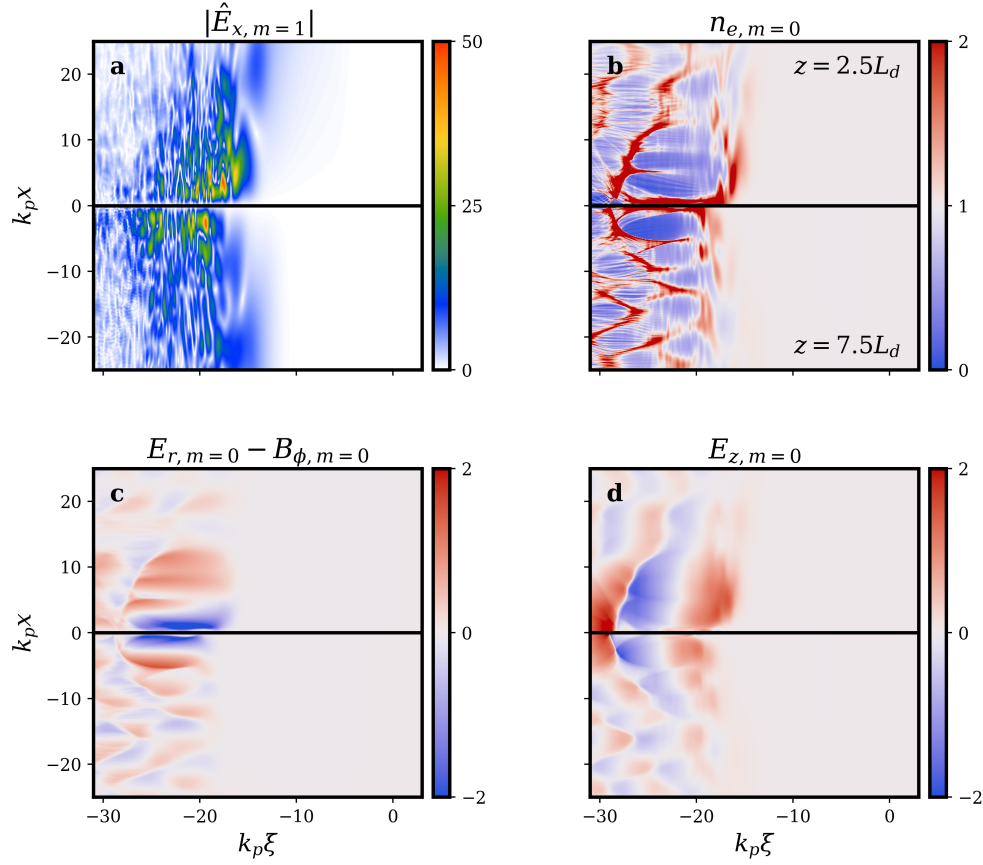


Figure 6.7: Nonlinear wakefields driven by a flying focus described by an ASTRL flying focus with an $(\ell, p) = (1, 0)$ Laguerre-Gaussian transverse profile for each pulse. Each pulse has $a_{0j} = 20$ and $k_p w_0 = 4$. Alternating polarization is used. Top panels show the state at $z = 5000 \approx 2.5L_d$ and $z = 15,000 \approx 7.5L_d$. (a) Amplitude of the laser envelope for the $m = 1$ mode. The evolution is significantly more complex than in the linear regime and appears turbulent. (b) Unlike in the linear regime, plasma electrons are pushed by the laser onto the axis. Multiple off-axis wakes are visible. (c-d) The blue region for the focusing field and red region of the accelerating field near $k_p \xi = -20$ are favorable for acceleration of positrons. This structure is maintained throughout the propagation of the laser.

evident. A more uniform wake could potentially be created by tailoring the laser pulses.

The value of $a_0 = 20$ was found through trial and error. A sufficiently large a_0 is required

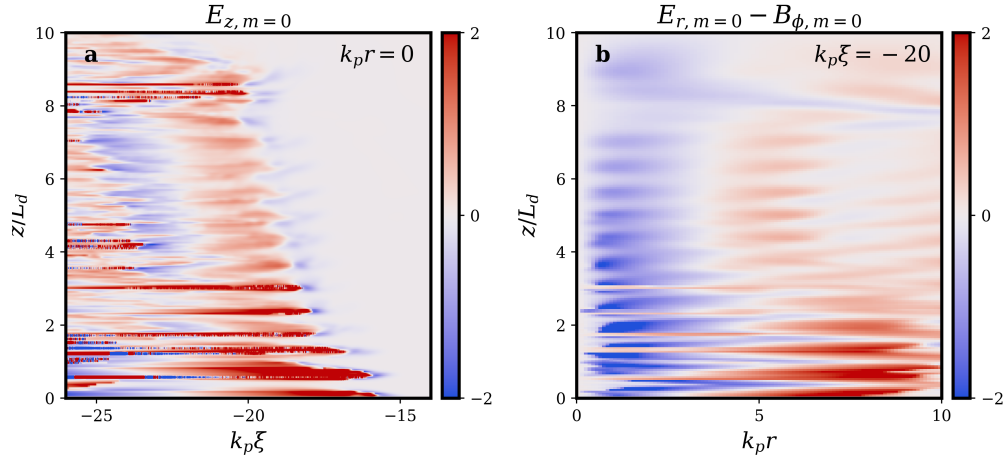


Figure 6.8: Time evolution of lineouts from the simulation in Fig. 6.7 for (a) the accelerating field on-axis ($k_p r = 0$) and (b) the focusing field at $k_p r = -20$. The accelerating and focusing fields persist over 10 dephasing lengths. Recession of the wake is visible in the E_z lineout. The discreteness of the $N = 16$ pulses is evident.

to produce the electron filament. It found that, unlike in the case of a Gaussian transverse profile, a sufficiently low number of pulses is required. In several cases in the nonlinear regime where the pulses overlapped substantially, it was observed that the laser peak was deflected substantially away from the axis, resulting in a negligible wake on the axis. Exploration of this physics and optimization of the pulse parameters is a natural area for future work.

Although the signs of the fields are seen to be favorable for positron acceleration in Figs. 6.7 and 6.8, the simulation cannot be run with a positron beam at the time of writing. This is because the numerical scheme breaks down within the electron filament for some timesteps. This is visible in 6.8(a) at several values of z . This phenomenon also occurs in the electron acceleration cases and is visible in Fig. 6.3. However, it was not a problem because in those cases the algorithm only breaks down behind the electron beam. In this case, the algorithm breaks down where the positrons would reside. An area for future development is to find numerical parameters that lead to convergence or develop a numerical scheme that addresses this. Estimating an average accelerating gradient of $eE_z/mc\omega_p = +1$, an order of

magnitude estimate of 10 GeV in 6 cm can be made for the energy gain positrons would experience in the fields of Fig. 6.8.

6.8 Discussion

A discrete flying focus can eliminate dephasing by driving a plasma wave with a phase velocity equal to the vacuum speed of light ($v_w = c$). This allows highly relativistic electrons to remain collocated with the maximum accelerating field of the plasma wave over a distance limited only by the number of pulses or total pulse energy. By removing the constraint of dephasing, a DFF-driven LWFA can operate at higher densities, providing larger energy gains in a single stage or equal energy gains in a shorter stage. The former would reduce the number of stages required to reach a target electron beam energy.

Quasi-3D PIC simulations with non-optimized parameters demonstrated that a DFF-driven plasma wave can accelerate a 50-pC electron beam from 1 GeV to 41 GeV over 50 dephasing lengths while preserving the beam quality. This is $3\times$ the maximum energy gain achievable with a conventional pulse in a lower-density stage of the same length. The acceleration length in the simulation was limited by computational resources and could, in principle, be extended to attain beam energies beyond 100 GeV in a single stage. With a longer stage or at higher density, the energy gain achievable by a DFF becomes even larger than that of a conventional pulse [Eq.(6.16)].

The simulation domain was filled with pre-ionized plasma. In practice, this would require a large amount of energy, which would reduce the overall efficiency. For the 41-GeV result, the energy required to pre-ionize the hydrogen gas would be $\pi R_p^2 L_p n_0 U_i \approx 2600$ J, where $R_p = 2.1$ cm is the maximum radius of the simulation and $U_i = 13.6$ eV is the ionization potential. In principle, the energy required for ionization could be reduced by a factor of 3 with no changes to the laser pulse if the pre-ionized plasma were structured into a cone. Alternatively, the energy required for ionization could be reduced by orders of magnitude if

the DFF could be efficiently coupled into a pre-ionized plasma with a radius of a few spot sizes $R_p \approx 3w_0$. This approach would likely necessitate customization of the pulse parameters and radial density profile to mitigate refraction. Another approach to reducing the required energy by orders of magnitude is to redesign the DFF to both ionize a gas and drive a plasma wave as in Ref. [12]. The spacing of the pulses would be chosen to account for linear dispersion in both the gas and plasma. This approach would also necessitate careful design of the pulse and plasma to mitigate refraction—in this case ionization-induced. Both of these approaches are areas for future work.

The DFF was designed using the formalism of ASTRL pulses [7]. The ASTRL formalism is sufficiently general that it encompasses continuous flying focus pulses, and therefore is inherently more flexible. The configuration considered here consists of multiple, collinear pulses with tunable delays, focal points, and polarizations that can be independently tuned. However, the ASTRL formalism also allows for tunable spot sizes, durations, and topological structure, such as transverse or longitudinal orbital angular momentum. This provides an extremely open parameter space for optimizing LWFA. In addition, unlike the axiparabola-echelon flying focus, the DFF used here features a near-constant spot size, which helps to mitigate evolution of the accelerating field. The cost for these benefits is increased experimental complexity due to the use of multiple pulses.

Future work could explore customization of individual pulse parameters in order to optimize plasma formation, energy gain, or beam quality. For instance, the techniques described in Refs. [226, 227] could be adapted to determine a pulse train that mitigates refraction, produces a more complete blowout, or suppresses the oscillations in the rear sheath of the plasma wave observed in Figs. 6.3(e) and (f). More generally, the duration, axial shape, spot size, or orbital angular momentum of each pulse could also be tuned using the ASTRL formalism, offering extensive possibilities. A train of incoherent pulses could also be used to relax requirements on the coherence of each pulse. The alternating polarization case, which resembles an incoherent sum, provides confidence that this approach could be effective.

Here, the DFF was used solely to produce a plasma wave with $v_w = c$, but the flexibility to structure the pulse train could be beneficial in a variety of LWFA scenarios. For instance, a DFF could drive a plasma wave with a slower phase velocity for muon acceleration [228], or a superluminal phase velocity to prevent wave breaking [229]. The first few pulses could be tailored to either facilitate self-injection or to suppress it when external injection is employed.

The use of a DFF to produce wakefields suitable for positron acceleration by using an $(\ell, p) = (1, 0)$ Laguerre-Gaussian transverse profile for each pulse was explored. Such modes can produce wakes with electron filaments on-axis so that positrons can be accelerated and focused. It was found that this method can indeed extend the acceleration length to at least ~ 10 dephasing lengths. This is a natural area for continued investigation. One of the challenges that would need to be addressed is the occasional failure of the numerical algorithm within the on-axis electron filament.

An experimental realization of a DFF pulse could take advantage of techniques for manipulating and assembling multiplexed laser pulses. For instance, the multiplexed pulses of a high-power fiber laser [230] could be independently delayed and focused before being coherently combined on a collinear path. Alternatively, if the central frequency of each pulse can vary, a broad bandwidth pulse could be dispersed into several pulses that could be independently delayed and focused. Note, however, that dividing the bandwidth in this manner would place a lower bound on the duration τ_j of each pulse. The possibility of using different central frequencies and assessing the limitations thereof requires further investigation. A schematic and further discussion of possible techniques can be found in the original work on ASTRAL, Ref. [7].

CHAPTER 7

Acceleration of the Particle-in-Cell Code OSIRIS with Graphics Processing Units

This chapter includes details on the implementation of GPU acceleration in the particle-in-cell code OSIRIS. Although GPU-enabled OSIRIS was not used for results presented in this dissertation outside this chapter, it can be used for future LWFA simulations and its concepts can be applied in other PIC codes such as WarpX.

Fully relativistic particle-in-cell (PIC) simulations are crucial for advancing knowledge of plasma physics. Modern GPU-based supercomputers offer the potential to perform PIC simulations of unprecedented scale, but require robust and feature-rich codes that can fully leverage their computational resources. In Ref. [9], this demand is addressed by adding GPU acceleration to the PIC code OSIRIS. An overview of the algorithm, which features a CUDA extension to the underlying Fortran architecture, is given. Detailed performance benchmarks for thermal plasmas are presented, which demonstrate excellent weak scaling on NERSC's Perlmutter supercomputer and high levels of absolute performance. The robustness of the code to model a variety of physical systems is demonstrated via simulations of Weibel filamentation and laser-wakefield acceleration run with dynamic load balancing. Finally, measurements and analysis of energy consumption are provided that indicate that the GPU algorithm is ~ 14 times faster and is ~ 7 times more energy-efficient than the optimized CPU algorithm on a node-to-node basis. The described development addresses the PIC simulation community's computational demands both by contributing a robust and performant GPU-accelerated PIC code and by providing insight on efficient use of GPU hardware.

This chapter is taken almost verbatim from Ref. [9]. The latter is distributed under the terms of the Creative Commons CC BY 4.0 license, which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited. The only substantial difference is Fig. 7.5 and an associated paragraph in Sec. 7.4, which discusses an idea to improve the performance of the code for simulations involving stationary plasmas in a moving window (as in laser wakefield acceleration). Other than this, self-referential statements (e.g., “this paper”) were translated to self-references to the chapter or to references to Ref. [9]. Finally, the original article was written heavily in the first-person plural; the text has been lightly edited to reflect the single-author nature of the dissertation, e.g. by rewording to remove most instances of the word “we”.

7.1 Introduction

Since the early 2000s, graphics processing units (GPUs) have emerged as the architecture of choice for scientific computing, owing to their increased computational throughput, parallelism, and energy efficiency relative to CPUs [231, 232, 233]. Modern supercomputers increasingly rely on GPUs for acceleration: at the time of writing, TOP500 rankings report that GPU hardware is used by 66 of the 100 fastest and 79 of the 100 greenest supercomputers in the world [234].

In order to leverage these computational resources, scientists have been faced with the challenge of rethinking the implementation of many simulation algorithms. Among these are particle-mesh simulation algorithms. In particular, the particle-in-cell (PIC) plasma simulation algorithm is the tool of choice for fully self-consistent simulation of nonlinear plasma physics where kinetic physics is important [57, 235]. PIC simulations have played an essential role in advancing the understanding of laser-plasma interactions, plasma-based acceleration, space physics, plasma astrophysics, and basic kinetic plasma physics [236, 121, 237]. In spite of the relative simplicity of the PIC algorithm, its implementation on GPU hardware is

complicated by the use of particles, which result in irregular patterns of memory access and movement. Coupled with evolving GPU capabilities and the growing number of available software paths to GPU acceleration, many questions remain about best practices and paths toward GPU acceleration of PIC codes.

GPU acceleration of PIC codes is an ongoing effort by the plasma simulation community. Early algorithms were constrained by limited device memory capacity and the need to circumvent the relatively poor performance of global atomic writes during current deposition on older hardware [33, 238, 239, 240, 241, 242, 243, 244, 245]. More recently, feature-rich, user-oriented PIC codes targeting modern GPU architectures have begun to emerge, including WarpX [246, 247, 248], HiPACE++ [82], VPIC [249], PIConGPU [250], and OSIRIS 2.0 [251] and OSIRIS 3.0 [252].

In this chapter, the approach taken to enable GPU acceleration of the fully electromagnetic PIC code OSIRIS 4.0 in Ref. [9] is presented. OSIRIS [220] is widely used by the plasma simulation community because of its maturity, speed and parallel scalability, and extensive suite of simulation capabilities, including customized field solvers for suppression of numerical dispersion [59, 60, 61], cartesian and quasi-3D geometries [222], analytic pushers including radiation reaction [65], monte-carlo modeling of Coulomb collisions [104], tile-based dynamic load balancing [253], particle-damping for laser-solid interactions [254], field ionization [105], real-time subgrid radiation calculation [255], and semiclassical prescriptions for modeling QED effects [256]. GPU acceleration was implemented in previous versions of OSIRIS, and was used in some physics studies. However, these versions were stand-alone, and due to the new data structures in OSIRIS 4.0 and developments in GPU hardware a new algorithm was needed. The new implementation, which improves upon previous versions, provides robust and performant acceleration of the base code via CUDA in all Cartesian geometries (1D, 2D, 3D). This is a step toward the larger software challenge of incorporating all the features listed above.

The outline of the remainder of this chapter is as follows. In Section 7.2, details of

the algorithm and software implementation in CUDA are discussed. Section 7.3 includes simulations run on NERSC’s Perlmutter system, which demonstrate (1) characterization of the absolute performance and weak scaling of the code on thermal plasmas, (2) capability of the code to model more complex physics problems such as Weibel filamentation and laser wakefield acceleration, and (3) 7x improved energy efficiency relative to the CPU algorithm on a node-to-node basis. Finally, Section 7.4 contains a discussion and perspectives on future development.

7.2 Methodology

7.2.1 GPU programming framework

A variety of software approaches have been taken to implement the PIC algorithm on GPUs, including those designed in CUDA [33, 238, 239, 240, 241, 242, 243, 245, 246, 257], HIP [246, 238], SYCL [246], OpenCL [244], Kokkos [249], AMRex [246, 82], Alpaka [250], and OpenACC [258]. OpenMP and RAJA are other approaches that have not been used. These frameworks have different levels of maturity and developer support, as well as differing implications for performance, supported GPU architectures, and implementation flexibility. The choice of which software approach to take is a significant and nontrivial developer decision that must consider these aspects of the available frameworks. For example, CUDA is the most mature general-purpose GPU programming language, is generally regarded to give the highest performance, and provides the greatest control over the GPU hardware. However, its greater level of control can come at the cost of greater quantity and complexity of code. This is especially true if cross-platform portability is required, since CUDA must be combined with other frameworks in order to support non-NVIDIA GPUs.

On the other end of the spectrum, a high-level approach abstracts hardware details, freeing application developers from the burden of supporting specialized, rapidly evolving systems. For example, OpenMP and OpenACC provide cross-platform portability through a

set of compiler directives. While perhaps the simplest to implement, these approaches offer the least amount of flexibility, hardware control, and performance. High-level abstraction layers, like Kokkos, are somewhere in between. These provide cross-platform portability and some degree of control over device memory, but to a lesser extent than CUDA.

Weighing the options led us to write our implementation in CUDA C through Fortran-C interoperability. This was informed by our goal of obtaining maximal performance. The low-level exposure of the code allowed us to optimize the particle-pushing kernels. The natural access to raw device memory pointers was also helpful for parts of the code with lower computational cost but greater algorithmic complexity, including particle sorting, management of the memory pool (discussed in Section 7.2.3), and buffered data movements. These aspects of the algorithm may have been significantly more difficult to implement in a higher-level framework. As a production code supporting a large user community, the stability and maturity of CUDA also weighed heavily in our decision. Restriction to NVIDIA hardware is not a significant limitation because many supercomputers in the modern supercomputing ecosystem use NVIDIA GPUs. Furthermore, a preliminary HIP version of our code demonstrated that translation from CUDA is relatively straightforward. This provides a reasonable path to supporting AMD GPUs as well, and thereby the vast majority of top supercomputers.

7.2.2 Tile-based domain decomposition

Most implementations of the PIC algorithm on GPUs have used some form of domain decomposition with more subdomains than GPUs. The resulting subdomains have been referred to in the literature as tiles, bins, clusters, and supercells. Here these subdomains are referred to as tiles. Tiles provide increased memory localization. In particular, they enable the use of shared memory as a developer-controlled cache by partitioning the computational domain into subdomains whose field and current arrays fit into shared memory. This gives higher performance for read, write, and atomic operations in both the field interpolation and

current deposit. Shared memory capacity determines the maximum tile size.

This GPU implementation is built on top of the existing tile data structure in OSIRIS, which was originally implemented to enable tile-based dynamic load balancing. The details of this implementation are discussed in Ref [253]. In essence, different MPI ranks can freely exchange tiles in order to ensure that the computational load is balanced. In addition to enabling the benefits for GPU performance previously discussed, the GPU implementation inherits the capacity for tile-based dynamic load balancing from the CPU code. This enables the GPU algorithm, like the CPU algorithm, to potentially run faster or with a lower memory footprint on load-imbalanced problems. The use of this tile data structure as the foundation of the algorithm is one of the main differences between this new implementation of GPU acceleration in OSIRIS and previous implementations in OSIRIS 2.0 [251] and OSIRIS 3.0 [252].

In this implementation, each MPI rank has access to one GPU, and as many shared-memory CPU cores as available on the compute node on which the code is running. For example, a Perlmutter GPU node has 64 CPU cores, and 4 GPUs. Therefore, a typical configuration would be 1 GPU and 16 CPU cores per MPI rank.

The use of tiles makes it more difficult to efficiently transfer memory between CPU and GPU. Naively calling CUDA memory copying operations within a loop over tiles results in a bottleneck due to function launch overhead. This issue is circumvented through the implementation of a memory transfer class which buffers unidirectional data movements of both particle and field data. The size of the buffer (typically several gigabytes) is specified by the user. The memory manager performs a data transfer only when the buffer is full or the source code explicitly flushes the buffer.

7.2.3 Particle chunk pool

The use of tiles also introduces the problem of inter-tile memory management. Naively allocating a fixed-size particle buffer for each tile leads to the possibility of buffer overflow as

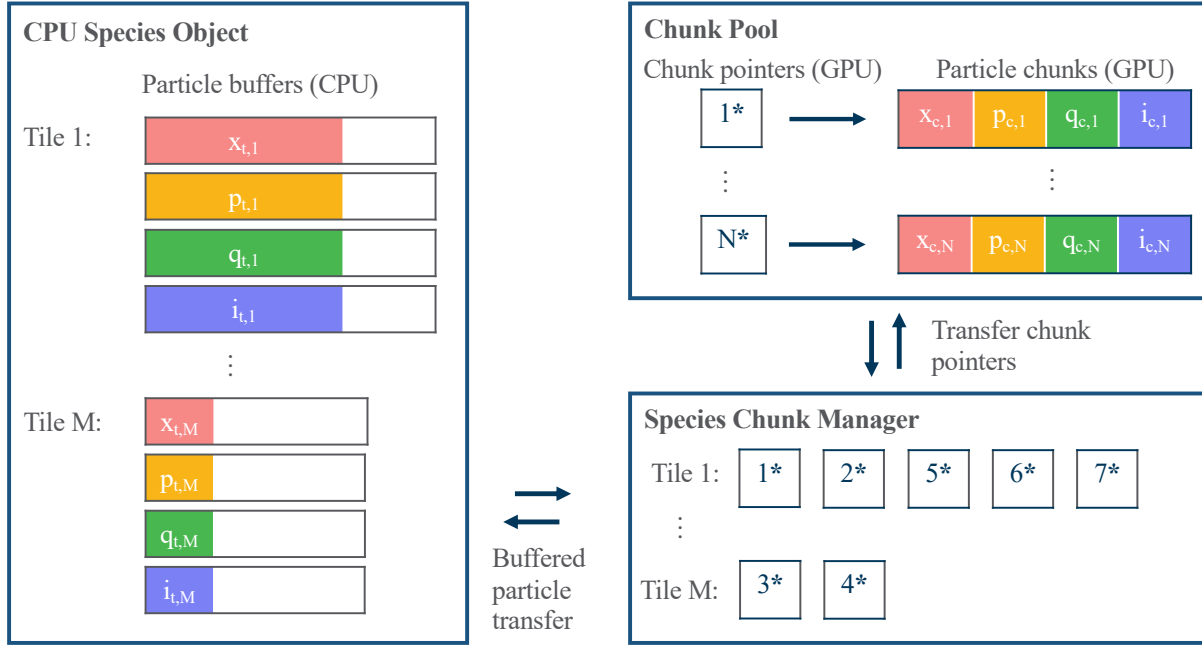


Figure 7.1: Schematic of the CPU memory layout, chunk pool data structure, and species chunk manager. The CPU species object stores contiguous arrays of particle position relative to nearest cell, momentum, charge, and cell index ($\mathbf{x}$, $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{ix}$) for each tile. The species chunk manager, whose state is maintained on both CPU and GPU, stores pointers to chunks of device memory for each tile. Chunks contiguously store all four data fields for a fixed number of particles. Each species chunk manager may transfer pointers to chunks to and from the chunk pool with nearly zero overhead, enabling arbitrary load-balancing between tiles with nearly zero overhead. Particle transfer between the CPU species object and species chunk manager is achieved through an object for buffered data movements, discussed in Section 7.2.2.

load fluctuates. Allowing the size of each tile's particle buffer to vary through reallocation can overcome this issue. However, reallocation of device memory is not possible within kernels, may not be possible if the GPU operates at full memory capacity, and can otherwise become a bottleneck.

These issues were addressed by using a memory pool of preallocated chunks of particle data, similar to Ref. [246]. The data structures are shown in Figure 7.1. Each MPI rank has its own chunk pool object, which allocates the majority of device memory as particle chunks at the beginning of the simulation. Each chunk contiguously stores the position, momentum, and charge for a fixed number of particles. The chunk pool stores pointers to these chunks in a circular buffer in device memory. Chunk pointers can be taken from or given back to the circular buffer through buffered device-to-device memory transfers. The state of the circular buffer is maintained on both the host and device so that chunks can be retrieved from the queue within either host functions or device kernels.

Each particle species has a species chunk manager object, which is responsible for ensuring that there is an appropriate number of chunks to store the particles on each tile. All species chunk managers on a given GPU share the same chunk pool and may store a different number of chunks for each tile. During each timestep, the species chunk managers transfer chunk pointers to and from the chunk pool so that each tile retains a specified number of empty chunks prior to the particle push. The number of empty chunks is chosen to be large enough to avoid overflow during the sort, but small enough to avoid under-utilization of device memory. When particles are sent to and from the host for communication or diagnostics, buffered particle transfers are used to move the data. Each CPU species object for each tile stores particles in one contiguous buffer instead of a memory pool. A CPU memory pool is not necessary because of the typically abundant levels of CPU memory.

The overhead of using particle chunks compared to larger contiguous blocks of memory is negligible. Within kernels that loop through particles, only a few additional variables and operations are required to also loop through the chunks. Furthermore, memory accesses of chunks within a kernel occur with the same performance as when accessing a larger contiguous block of memory. This is because by choosing the number of particles per chunk to be greater than or equal to the warp size, each warp sees contiguous memory. This was confirmed by tests with a large chunk size. Outside of kernels, the overhead of the species chunk managers

exchanging chunks with the chunk pool is negligible because only pointers to chunks are passed. Occasional reallocation of a species chunk manager’s chunk buffer for a particular tile is not a problem; these buffers can be much larger than necessary because they contain significantly less memory than the memory stored by the chunks themselves.

The particle chunk pool enables load to fluctuate between tiles with little wasted memory. When combined with dynamic load balancing for inter-GPU load imbalances, which ensures that total memory load per GPU remains roughly fixed, this enables simulations which nearly max out available system memory. This could be useful for running as large a simulation as possible given limited device memory resources, as in Ref. [259].

7.2.4 Particle pushing and current deposition

Particle pushing and current deposition is typically the most computationally expensive part of a PIC code. Thus, significant effort went into profiling and optimizing this part of the code, which is broken into two separate kernels. In the first kernel, which is referred to as `update_velocity`, electric and magnetic fields are interpolated onto the particle positions, then particle momenta are updated. In the second kernel, which is referred to as `advance_deposit`, particles positions are advanced, and a charge-conserving current deposition is performed.

While the `update_velocity` kernel is straightforward to parallelize — particles can be processed individually in parallel — `advance_deposit` is not; a simple loop in parallel over particles could result in a memory collision when depositing their current onto the grid. This is circumvented by depositing the current via atomic addition. On modern architectures, atomic addition of both single and double precision floats is natively supported and therefore fast enough that this is the typical solution [245, 246, 247, 248, 249, 250].

Both kernels saw a significant speedup when fields were explicitly loaded into shared memory. This includes both the electric and magnetic field arrays in `update_velocity`, and

the current array in `advance_deposit`. With fields and currents instead simply stored in global memory, both kernels (in particular `advance_deposit`) suffered from high latency due to an L2 Cache bottleneck, which was alleviated by using shared memory.

Best performance with shared memory came when using one CUDA thread block per tile with a block size of 512. Using more than one block added additional overhead related to shared memory without increasing occupancy. However, there may be situations where this would not give the best performance. For example, in problems with severe load imbalance, one MPI rank can sometimes be responsible for a small number of tiles (order 1), each with many more particles than average. With one block per tile, far fewer warps would be launched than is sufficient to achieve high device occupancy. In these situations, the benefits of minimizing data duplication could be outweighed by the benefits of launching multiple blocks per tile. This is a possible future direction.

In addition to using shared memory, kernels benefited from other optimizations. This applies particularly to `advance_deposit`, which has more complicated logic and flow control. Memory transactions (both to global and shared) were minimized, arithmetic was streamlined to minimize instructions executed, and register usage was optimized to ensure maximal warp occupancy. NVIDIA’s profiling tool, Nsight Compute, played an important role in this process.

The possibility was also explored of enhancing performance by optimizing the memory layout of particle data. GPU (and SIMD) programming guides would suggest that “array of structs” (AoS) layout for particle data is best on CPU architectures, while a “struct of arrays” (SoA) is best on GPUs due to coalesced memory access. For example, for N particles, each with position (x_i, y_i) , where $i \in [1, N]$ is the particle index, optimal performance on a GPU would be obtained by storing particle data as $(x_1, x_2, \dots, x_N, y_1, y_2, \dots, y_N)$, instead of $(x_1, y_1, x_2, y_2, \dots, x_N, y_N)$. This layout is used in Ref. [249]. In the code, particle data is stored in this manner. However, it was found that there was no measurable difference between SoA and AoS.

Finally, this algorithm makes use of a distinct particle initialization scheme compared to what is done for CPU simulations. Particles that are close together in simulation space are staggered in memory upon initialization. This results in a speedup in `advance_deposit` by reducing the likelihood of memory collisions in the current deposition. This is particularly important for simulations with a cold plasma (e.g. plasma based acceleration) where particle position evolves slowly after initialization. Benefit is also seen for warm plasmas over the first few 100 simulation iterations, until particle positions become sufficiently scrambled. A small slowdown occurs in `update_velocity` due to decreased data locality when interpolating fields onto particles, but this is outweighed by the speedup in `advance_deposit`.

In aggregate, the optimizations above resulted in a roughly 5–6x speedup relative to the initial implementation. As stressed above, this level of optimization may not have been possible had a higher-level approach using a framework such as OpenMP, OpenACC or Kokkos been taken. Nevertheless, these optimizations may well apply to other GPU programming frameworks.

7.2.5 Particle sorting and boundary conditions

The literature refers to spatial grouping of particles as particle sorting. This includes organizing particles by MPI rank, by tile, or by any other grouping of cells. Formally this is a bucket sort. The order of the particles within buckets typically doesn't matter, and one can make the additional assumption that particles are mainly presorted — only a fraction of the particles will move between buckets on any given timestep. Particle sorting is required when any form of domain decomposition is used, or in special cases such as particle-particle collisions [104, 260, 261].

Efficient particle-sorting in PIC codes on the GPU is non-trivial. Particle arrays must be compact; both in order to make efficient use of memory, and because a particle array with holes scattered throughout would result in poor performance during particle push and current deposition due to complicated control flow, warp divergence, and non-coalesced memory

access. In general, sorting is achieved by making use of some combination of atomic operations to resolve memory collisions, prefix scans and stream compaction to obtain ordered arrays, and temporary buffers to store moving particles and counting integers. Many algorithms have been proposed by others [33, 239, 240, 245, 246, 258, 262, 263]. Differences between algorithms in the literature arise because different codes face different constraints.

Our aim was to design a sorting algorithm with robust fault tolerance to memory overflow to handle the variety of physical systems modeled by the OSIRIS user community. Memory overflow can occur in a particle sorting kernel due to buffering of particles moving between tiles. This is prevented from occurring by using the chunk pool, as discussed in Section 7.2.3, and by minimizing the memory footprint of the sort algorithm altogether. The sort kernel was developed to be as performant as possible given these constraints. Ultimately, the choice to prioritize fault tolerance over performance was justified because performance was sufficient: kernel duration was dwarfed by CPU MPI communications of the particle data. The performance of the sort could potentially be improved by taking advantage of remote direct memory access (RDMA) via CUDA-Aware MPI.

In the remaining paragraphs of this section the algorithm is described in detail. The algorithm takes place in three steps. First, a kernel is called where threads loop over all particles on a given tile, identify those departing, and buffer them according to their destination. Particles whose destination tile is local (i.e., on the same MPI rank) are appended to the destination tile’s particle buffer directly, which is made up of globally accessible chunks. Particles crossing an MPI rank boundary or the simulation boundary (e.g., absorbing wall) are buffered in a dedicated array of chunks also stored in global memory. An additional buffer for each tile, `i_hole`, is required to store the array indices of departing particles, which are referred to as “holes.” In the current implementation, this buffer is of fixed size. But it could be implemented using chunks as well. Three integers are required to count the number of particles: departing a given tile; being received by a given tile locally; and crossing an MPI rank or simulation boundary. These are stored in global memory and are incremented

atomically.

Second, particles crossing an MPI rank or simulation boundary are copied from device to host. The host handles processing of particle boundary conditions (e.g. thermal-bath, open, or reflecting particle boundaries), as well as exchange of particles with other MPI ranks. Then, the new set of particles is copied from host back to device.

Finally, a kernel is called to compact the buffers for each tile: holes left by particles that have departed are filled so that particles are stored contiguously. Mainly, this kernel is embarrassingly parallel. Threads loop over particles (that previously were either received locally from other tiles, or copied from the host) and move them into holes specified by `i_hole`. However, special care must be taken for the case where the new number of particles on a given tile, `n_p_new`, is less than its old number of particles, `n_p_old`. This is because, in this case, some of the holes specified by `i_hole` are “invalid”, i.e. they have index greater than `n_p_new`. Since the holes in `i_hole` are in no particular order, an initial step is implemented in which `i_hole` is compacted into a smaller array containing only valid holes. This process is known as stream compaction and relies on parallel prefix scans [264]. A similar compaction step is required when moving particles with index between `n_p_new` and `n_p_old` into holes, since that part of the array contains the invalid holes that must be filtered out. No atomic operations are necessary in this kernel.

7.3 Results

In this section, examples of simulations run on NERSC’s Perlmutter system are presented. Each Perlmutter GPU node has four NVIDIA A100 GPUs with 40 GB of device memory and a single AMD EPYC 7763 CPU. Each CPU node has two AMD EPYC 7763 CPUs. Each AMD EPYC 7763 CPU has 64 cores.

In all of the results, the absolute throughput is defined as the total number of particle pushes divided by the total simulation time minus the time required to initialize the simulation.

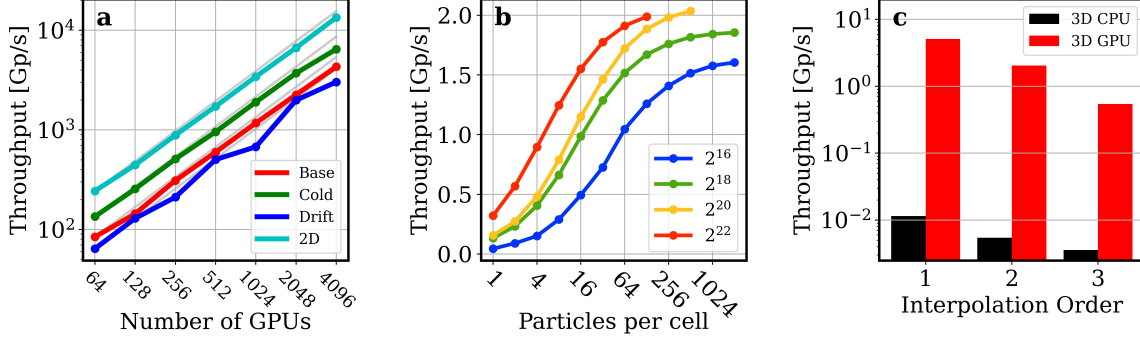


Figure 7.2: Comprehensive thermal plasma benchmarks run on Perlmutter. (a): Weak scaling tests for the cases described in the text, showing near-perfect scaling up to 4096 GPUs on Perlmutter. (b) Parameter scan showing dependence of throughput on particles per cell and grid sizes. (c) Comparison of absolute performance on one GPU and one core of one CPU for different interpolation orders. All simulations are run in 3D with quadratic interpolation and 512 particles per cell unless otherwise specified.

This metric is the average number of particles pushed per unit time when including the additional costs of particle communication, grid communication, and field evolution. No additional normalization is used if more than one GPU is used.

7.3.1 Thermal plasma benchmarks

Figure 7.2 shows several performance tests from simulations of thermal plasmas. Weak scaling is first considered, shown in Figure 7.2(a). To understand the effect of different configuration settings on performance, all of our simulations are derived from a single *base* case, which serves as a control. Weak scaling tests were run of the following computational problems, which can be extrapolated to understand a variety of use-cases:

- *Base*: 3D simulation of a uniform plasma using our GPU algorithm run on 64 GPUs. The plasma is warm ($u_{\text{th}} = 0.1m_e c$ in all directions) and is represented by 512 particles

per cell. Quadratic particle shapes are used. Single precision is used for both field and particle data. Note that particle positions are referenced with respect to cells, as depicted in Figure 7.1, so that the use of single precision does not limit accuracy for large position values. The box is fully periodic with no moving window. The grid dimensions are $512 \times 512 \times 256$ cells. The tile dimensions are $8 \times 8 \times 8$ cells. The resolution is $k_p \Delta x = 0.1$ in all directions, where $k_p \equiv \omega_p/c$, and $\omega_p \Delta t = 0.05$. The simulation is run for 1000 timesteps. One MPI rank is used per GPU and one CUDA thread block per tile with 512 threads.

- *Cold*: Same as *base* with $u_{\text{th}} = 0$. Particle communication vanishes aside from pings between neighboring ranks.
- *Drift*: Same as *cold* but with a moving window in the $\hat{x}_1$ direction. Particles are stationary but move relative to MPI partitions. This problem has the most particle communication and approximates the data movement for simulations of plasma-based acceleration.
- *2D*: A 2D version of *base*. The number of particles per cell and thermal velocity are the same. The grid size of 8192×8192 cells gives the same number of total grid points and total number of particles. The tile dimensions are 32×32 cells.

Figure 7.2(a) shows weak scaling tests for the four test problems. In each test, the grid dimensions are scaled proportionately to the number of MPI ranks used to decompose the problem. For ideal scaling, depicted by solid gray lines, the total throughput is proportional to the number of MPI ranks. In each case, weak scaling is nearly perfect up to 4096 GPUs, the highest power of two below the full system size of 6144. The *drift* case is slower than *cold* because of increased communication cost. It exhibits jitter due to variation in communication cost. *Cold* is faster than *base* due to the use of the staggered initialization discussed in Section 7.2.4 and reduced communication costs. Without the staggered initialization, *cold* is several times slower than *base*.

Figure 7.2(b) shows the dependence of the absolute throughput on the number of gridpoints and particles per cell. All simulations are run on one GPU with the same parameters as the 3D *base* simulation other than the number of gridpoints and particles per cell. The tile size is decreased for smaller grid sizes to maintain a sufficient number of blocks to saturate performance.

For a fixed total grid size, performance increases with particles per cell for two reasons. First, particle pushing contributes a greater fraction of the total simulation cost relative to the field solver and grid communication, and particle communication. Second, similarly, the kernels `update_velocity` and `advance_deposit` are both more efficient since they also have fixed costs, and particle pushing contributes a greater fraction of total kernel time when higher particles per cell is used. The `update_velocity` kernel has the fixed cost of loading fields into shared memory arrays. It also likely benefits from caching of field values. The `advance_deposit` kernel has the fixed cost of zeroing the shared-memory current array, and adding that array back to global memory after deposition. On the other hand, for a fixed number of particles per cell, performance decreases with decreasing grid size. This is because the relative cost of pushing to communication scales as the ratio of volume to surface area of the grid and because kernels are less efficient with smaller grid size.

While performance decreases with decreasing either particles per cell or gridpoints per GPU, the performance is still within a factor of two of the maximum with 2^{18} or more gridpoints and ~ 16 or more particles per cell. Note that the performance does not depend only on the total number of particles (i.e. particles per cell times grid points). For example, the cases with 2^{16} gridpoints and 128 particles per cell has the same number of particles as the case with 2^{20} gridpoints and 8 particles per cell, but has about double the performance.

Finally, Figure 7.2(c) shows the absolute performance of the GPU and CPU algorithms for different particle interpolation orders on a version of the *base* GPU case scaled to fit one GPU. The CPU code uses an optimized SIMD pusher [265]. The orders 1, 2, and 3 are linear, quadratic, and cubic interpolations. For these three cases, the GPU algorithm on one GPU

is about 300 times faster than the CPU algorithm on one core of one CPU. One CPU core was used (instead of all 64 cores) in order to simplify the comparison since both CPU and GPU tests are run with one MPI rank. The results in Figure 7.2(c) are tabulated along with 2D timings in Table 7.1. Note that the 2D CPU algorithm shares the same cost for 2nd and 3rd interpolation because of specific memory alignment requirements for the SIMD pusher.

	Interpolation Order		
	1	2	3
2D GPU	9.6	5.8	3.3
3D GPU	5.1	2.0	0.55
2D CPU	0.019	0.011	0.010
3D CPU	0.011	0.0054	0.0036
2D speedup	500	530	330
3D speedup	460	370	140

Table 7.1: Absolute throughput in gigaparticles per second for the GPU and CPU algorithms on one GPU and one core of one CPU with different interpolation orders. The benchmarks are measured on a thermal plasma.

7.3.2 Simulations of more complex plasma systems

Here the ability of the code to simulate commonly studied plasma systems with higher degrees of spatial inhomogeneity and more complex particle movement patterns is shown. The level of robustness of the code required to handle these problems is significantly higher than that required to simulate the thermal plasma cases shown in Section 7.3.1. Anecdotally, many bugs concerning edge cases outside particle pushing kernels were exposed in these cases which did not manifest for thermal problems. Furthermore, these problems stress the performance of the code because they depart from several factors which led to best performance in thermal plasma tests from Figure 7.2. Namely, they have lower particles per

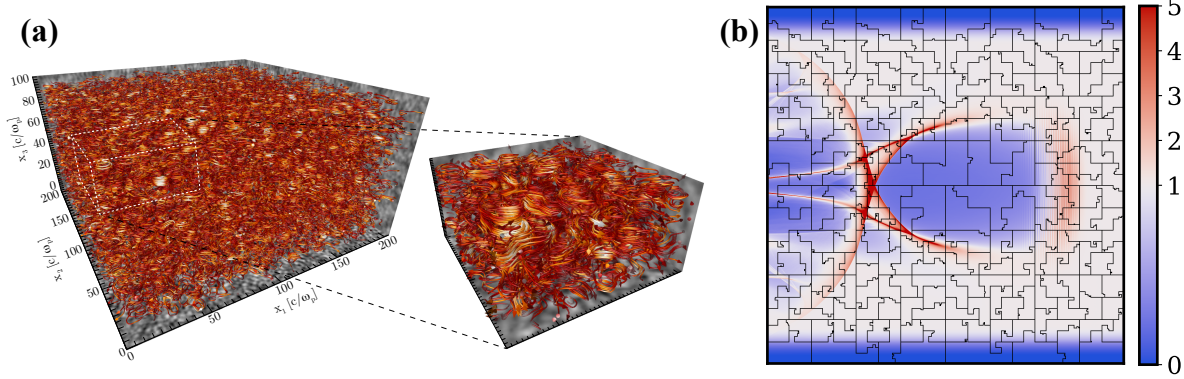


Figure 7.3: Examples of spatially inhomogeneous plasma simulations run on Perlmutter. (a): 3D simulation of the Weibel instability on 4096 GPUs, demonstrating the tangling of magnetic field lines. (b): charge density from a 2D simulation of LWFA in the nonlinear regime with dynamic load balancing. The black lines indicate the computational boundaries of different MPI ranks.

cell, higher communication levels, and greater degrees of spatial inhomogeneity.

Figure 7.3 shows two examples. Figure 7.3(a) shows tangling of magnetic filaments generated in a 3D simulation of the Weibel instability [266, 267, 268]. The simulation features two pairs of relativistically counterstreaming electron-positron plasmas with $u_{\text{th}} = 0.1$, $u_{\text{drift}} = 0.75$, and 32 particles per cell per species with quadratic interpolation. The grid dimensions are $2048 \times 2048 \times 1024$. The resolution is $k_p \Delta_x = 0.1 k_p$. The simulation was run on 4096 GPUs on Perlmutter up to $\omega_p t = 50$. The turbulence, relativistic streaming, and moving plasma structures with densities varying from 0 to $8n_0$ stress the code more than the thermal plasmas in Section 7.3.1. The code is robust to these stresses and the performance is 0.4 Gp/s, slightly less than 2 times lower than for an analogous thermal case (with 32 particles per cell, 2^{20} gridpoints per GPU, on 4096 GPUs) based on extrapolation from Figures 7.2(a) and (b).

Shown in Figure 7.3(b) is a 2D simulation with dynamic load balancing of a laser wakefield

accelerator (LWFA) with a frequency ratio of $\omega_0/\omega_p = 80$, corresponding to an 800 nm laser pulse incident on a plasma of density $2.7 \times 10^{17} \text{ cm}^{-3}$. The laser and plasma are matched according to the scaling in Ref.[269] with $a_0 = 4$ and $k_p w_0 = 4$. The simulation dimensions are (6400×9600) with $k_p \Delta_x = 0.0025$ in all directions and $\omega_p \Delta_t = 0.0015$. The simulation uses a moving window in the $\hat{x}_1$ direction. 64 particles per cell with quadratic interpolation are used for the plasma. The simulation was run on Perlmutter on 256 GPUs. The electron charge density is shown in color. The black lines outline the domain boundaries processed by different GPUs. The domains are formed by free exchange of tiles between different MPI ranks, as described in Section 7.2.2, to ensure load balance of simulation particles. Due to the particle communication associated with the moving window, this problem stresses the code similarly to the Weibel problem, but with a greater particle imbalance among GPUs. In this case, the performance is 0.2 Gp/s, roughly 8 times lower than for an analogous 2D thermal case (with 64 particles per cell, 2^{18} gridpoints per GPU, on 256 GPUs) based on extrapolation from Figures 7.2(a) and (b), and table 7.1. The discrepancy may be attributable to the fact that communication in this simulation is significantly higher relative to particle pushing (4 times higher) than in the thermal plasma benchmarks from Figures 7.2(a) and (b), limiting the accuracy of the extrapolation.

This example illustrates the functionality of dynamic load balancing in OSIRIS. Despite the existence of a particle load imbalance, the performance is roughly identical with and without dynamic load balancing. This is because the simulation is dominated by communication; load balancing can only lead to a speedup when the simulation is dominated by particle pushing. Nevertheless, we anticipate that this feature could provide a significant speedup on other problems, other systems, or as the code and hardware evolves (e.g., with the incorporation of CUDA-aware MPI, which could lower communication costs). Load balancing also ensures efficient use of limited device memory, which could enable running very large problems as discussed in Section 7.2.3.

The electron charge density is shown in color. The black lines outline the domain

boundaries processed by different GPUs. The domains are formed by free exchange of tiles between different MPI ranks, as described in Section 7.2.2, to ensure load balance of simulation particles.

A major motivation for the transition of supercomputers from CPU to GPU hardware is the higher energy efficiency of GPUs. However, the energy efficiency of specific applications may vary because of different patterns of computation and data movement. Here the energy efficiency of CPU and GPU implementations of OSIRIS are measured. To our knowledge, these are the first published measurements of energy consumption of the PIC algorithm.

A node-to-node comparison was run for a thermal plasma. The codes were used to simulate a thermal plasma just as in the *base* weak scaling test scaled to one GPU node. The CPU code was run with 128 MPI ranks and 128 cores on a Perlmutter CPU node, while the GPU code was run with 4 MPI ranks and 4 GPUs on a Perlmutter GPU node. Both the CPU and the GPU simulation were allowed to progress indefinitely and were terminated after 4 hours.

7.3.3 Energy efficiency

For both simulations, the total energy was measured using the Slurm `sacct` command, which reports energy consumption as measured by Cray’s Power Management System. The estimate provides a comprehensive measurement of energy consumption for the entire node. This includes power required for data movement, floating point operations, and baseline power needed to keep the node running. For GPU nodes, the measurement includes the energy consumption of the CPUs and the GPUs. Ref [270] provides more information on power consumption measurements on Perlmutter.

Performance on the GPU node was higher on a node-to-node basis: over the 4 hours, a factor of 13.7 times more simulation iterations were performed, higher than the value of 11.3 expected from extrapolating single-GPU and single-CPU benchmarks. The GPU node also

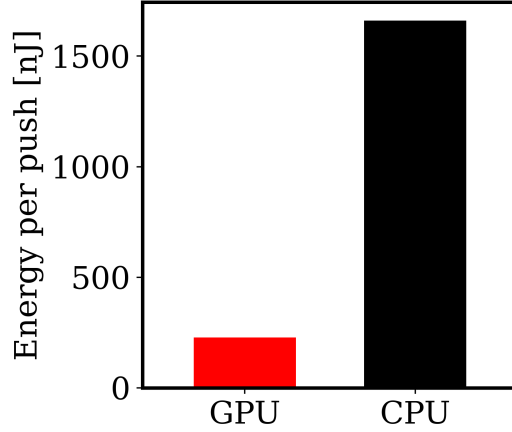


Figure 7.4: Energy consumption per particle push for the GPU algorithm and for the CPU algorithm. The comparison is between a simulation on a Perlmutter GPU node with 4 GPUs and a Perlmutter CPU node with 128 CPU cores. The GPU node is ~ 14 times faster than the CPU node and uses ~ 2 times more energy, leading to an overall energy efficiency ~ 7 times higher. Further detail of what is being measured is in the text.

consumed 1.9 times the energy of the CPU node. Dividing these numbers, the simulation was 7.2 times more energy-efficient when run on the GPU node. The results are shown in absolute units of nanojoules per push in Figure 7.4.

This simple test does not account for heating costs from inter-node communication or the potentially differing energy costs for creation and maintenance of CPU and GPU devices. Results may also vary significantly on different systems and hardware. Nonetheless, the test suggests that GPU architectures may be the more environmentally sustainable choice for PIC simulations. Alternatively, increased efficiency could result in greater demand, resulting in increased overall power consumption due to the Jevons paradox.

7.4 Discussion

In this chapter, the new implementation of GPU acceleration in the PIC code OSIRIS of Ref. [9] was presented. The code, written in CUDA, is built upon a tile-based domain decomposition in order to enable shared memory storage of grid quantities, enhance memory localization, and enable tile-based dynamic load balancing. For efficient memory usage when using tiles, a memory pool was implemented for particle data. Aspects of the optimization of the particle push were described in detail, along with details about the implementation of the particle sort. The pros and cons of using CUDA were discussed.

This chapter contains a comprehensive picture of the performance, capabilities, and energy efficiency of the code through several tests. A variety of thermal plasma simulations demonstrate strong absolute performance (e.g. 5.0 gigaparticles per second in 3D with linear interpolation) and excellent weak scaling up to 4096 GPUs on Perlmutter. Simulations of Weibel filamentation and laser wakefield acceleration illustrate the capacity of the code to run on more complex physical systems. Finally, node-to-node energy consumption measurements indicated that, on a node-to-node basis on Perlmutter, the GPU code is ~ 14 times faster and ~ 7 times more energy-efficient than the heavily optimized CPU code.

Within the implementation, several areas of improvement remain. The degraded performance when running with few particles per cell ($\lesssim 16$) is likely caused by bottlenecks which could be identified by NSight Systems and subsequently alleviated. CUDA-aware MPI could also be used for particle and field communication, which is currently a bottleneck in cases such as the *drifting* plasma in Figure 7.2.

Fault tolerance in the particle sort could also be improved. In the current implementation, fixed-size temporary buffers are allocated for to store the array indices of departing particles (`i_hole`). Overflowing these buffers causes a crash, and reallocation is not possible within the sort kernel. These buffers could be replaced by spare particle chunks from the memory pool. This would remove the need for preallocation of the fixed-size buffers and would make

any overflow impossible during the sort as long as the chunk pool does not become empty.

The performance of the algorithm for simulations using a moving window could be substantially improved by using stationary tiles which drift with respect to the moving window instead of moving tiles as in the current algorithm. This idea is illustrated in Fig. 7.5. This would make it so that no data movement occurs for a cold stationary plasma in a moving window. This would in principle improve the performance of the *drift* case in Fig. 7.2 to the level of the *cold* case. Once the computational domain has moved sufficiently far, tiles on the left side of the domain can be removed and appended to the right side containing freshly injected plasma. Parallelization works the same way as in the current algorithm. In Fig. 7.5, the load could be partitioned so that one GPU is assigned to tiles 1 and 2 and another GPU to tiles 3 and 4. This would work even with the tiles being spatially disjoint when using stationary tiles.

The final major area of improvement is GPU acceleration of other OSIRIS simulation modes and features, such as quasi-3D, QED, Monte Carlo collisions, and general relativity. The current implementation, which accelerates the base Cartesian simulation modes in 1D, 2D, and 3D, is the first step in accelerating the entire codebase.

In conclusion, the GPU implementation of OSIRIS provides excellent performance, supports a variety of use-cases, and provides backward compatibility with existing CPU code. This poses OSIRIS to continue to serve the PIC simulation community's computational demands. The details of this implementation may also be fruitful for the development of other feature-rich GPU-accelerated PIC codes.

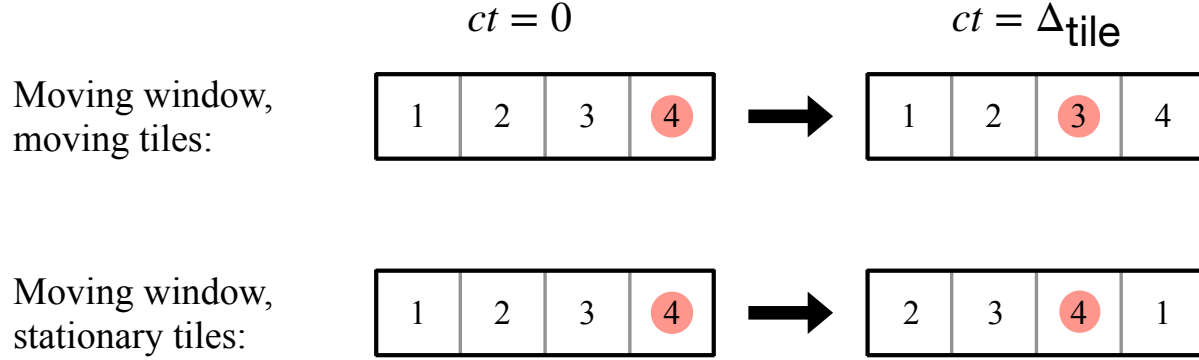


Figure 7.5: Pictorial explanation of an idea to improve performance for computational problems featuring a stationary plasma in a moving window, as in laser wakefield acceleration. The black box is the computational domain, moving from left to right, and the gray numbered boxes are tiles. In the current algorithm, the tiles move with the moving window. This results in a stationary plasma particle, shown in red, effectively moving across the tiles. In an improved algorithm, stationary tiles could be used. In this case, the tiles are stationary in space and drift with respect to the computational domain. This would in principle make the performance for drifting plasmas the same with and without a moving window. Once the computational domain passes the leftmost tile, that tile can be repositioned on the right side of the domain.

CHAPTER 8

Summary

This dissertation concerns computational modeling of laser wakefield acceleration (LWFA) with spatiotemporally structured laser pulses. In particular, the use of flying focus pulses to extend the acceleration length beyond the dephasing length was explored. These results are particularly relevant in the context of next-generation particle colliders, but are also relevant for other accelerator applications including x-ray free electron lasers, radiotherapy, and ion implantation. The work in this dissertation has led to several publications, including Refs. [7, 8, 9]. This chapter summarizes the main results of the dissertation and discusses ideas for future work.

Chapter 2 presents theory and computational method for LWFA in the blowout regime. This includes the key physical descriptions (single-particle, Vlasov, PGC approximation, quasistatic approximation), the key concepts (structure of the plasma wake, estimates for the acceleration length and energy gain, beam loading, emittance preservation), and simulation methods (PIC, PIC with the PGC approximation, and quasistatic PIC). While the chapter includes few original ideas, it does include some new perspectives and is certainly an original presentation of the material. It is written to streamline and clarify these topics. The author hopes that it will serve as a useful primer for others who are interested in understanding LWFA in the blowout regime.

Chapter 3 examines the scaling laws for the acceleration length and energy gain in the blowout regime of LWFA discussed in Chapter 2 for conventional laser pulses. This is important to understand both because of near-term experimental prospects as well as for

accurate comparison of a conventional pulse with new concepts involving spatiotemporal structure. In this study, the scaling laws for the acceleration length and energy gain in LWFA were tested by varying the dimensionless frequency ratio ω_0/ω_p for fixed dimensionless laser parameters of $a_0 = 4$, $k_p w_0 = 4$, and $\omega_p \tau_{\text{rise}} = \omega_p \tau_{\text{fall}} = 4$. By examining the evolution of the laser and wake, it was found that the laser guiding length satisfies $L_{\text{guide}} \approx 1.5L_d$ for $\omega_0/\omega_p \lesssim 50$, consistent with the assumptions of Lu et al (Ref. [1]). However, for $\omega_0/\omega_p \gtrsim 75$, the simulations suggest that $L_{\text{guide}} \approx 25z_R$, a less favorable scaling. The transition to this scaling also manifests in the properties of electron acceleration. We find that the Lu et al. scaling for the acceleration length $L_{\text{acc}} \approx L_d$ holds for $\omega_0/\omega_p \lesssim 150$, while $L_{\text{acc}} = 25z_R$ better models the data for $\omega_0/\omega_p \gtrsim 250$. Similarly, the Lu et al. scaling for the change in Lorentz factor $\Delta\gamma \approx \frac{2}{3} \left(\frac{\omega_0}{\omega_p} \right)^2 a_0$ holds for $\omega_0/\omega_p \lesssim 150$, while $\Delta\gamma \approx 25k_p z_R$ better models the data for $\omega_0/\omega_p \gtrsim 250$. The results are of interest for near-term experimental prospects with lasers in the 3–10 PW range at plasma densities of $10^{16} - 10^{17} \text{ cm}^{-3}$ at facilities such as ELI Beamlines. Future directions of this work are to confirm the accuracy of the simulations and to further explore the parameter space, including how to optimally shape the laser pulse. These are discussed in detail in Chapter 3.

Chapter 4 provides an original integral formulation for understanding linear laser propagation of laser pulses in linear dielectric. It extends calculations presented in Siegman's and Zangwill's textbooks [135, 136] to consider the full spatiotemporal evolution of the laser envelope, rather than just the transverse or longitudinal evolution. These calculations provide details about the linear propagation of laser pulses in plasma which are important to consider in the context of LWFA with flying focus pulses, especially with the ASTRL description used in Chapter 6. The approach in this chapter could easily be extended to consider higher-order effects of linear laser propagation in inhomogeneous plasmas, magnetized plasmas, and other linear dielectrics. In particular, extending the theory to plasma channels could provide insight into the coupling of flying focus pulses into a plasma column or plasma channel.

Chapter 5 presents the theory of Arbitrarily Structured Laser (ASTRL) pulses. This is a

general theoretical framework for the forward-design and mathematical description of various kinds of spatiotemporally structured light. The key idea is that many kinds of structured light can be described as a superposition of multiple unstructured laser pulses. The idea abstracts the flying focus concept, which can be seen as a superposition of pulses with different focal points. Though originally intended as a mathematical description of flying focus pulses for LWFA, this framework is much more general: it provides a framework for design and description of spatiotemporally structured light for laser-matter interactions in general. A clear direction for future research is to continue to conceptualize how ASTRL pulses can enable or enhance laser-matter interactions of interest. For example, some laser-plasma process can potentially be enhanced by using a pulse with evolving polarization or evolving Laguerre-Gaussian (ℓ, p) quantum numbers. In this vein, it is worth concentrating on applications that use superpositions of a handful (2-5) of laser pulses. Such applications could be feasible at experimental facilities in the near-term.

In Chapter 6, LWFA driven by an ASTRL flying focus is studied. This chapter contains the key result of this dissertation: quasistatic PIC simulations demonstrate that at a plasma density of $3 \times 10^{18} \text{ cm}^{-3}$, an ASTRL flying focus with a total energy of 150 J can transfer 40 GeV per electron to a 50-pC beam in a single 30-cm stage, corresponding to 50 dephasing lengths. The result was obtained for non-optimized parameters and was limited by simulation cost. This chapter also includes novel results on the application of ASTRL pulses to positron acceleration. Collectively, these results provide strong motivation for continued exploration of flying focus concepts for LWFA. Several future directions from this work stand out. First, one could explore the use of a neutral gas or plasma column in LWFA driven by a flying focus. This could increase the efficiency of the process by orders of magnitude by removing the need to preionize the plasma out to the initial spot size of the laser. Second, one could use machine learning methods to customize the parameters, delay, and shape of each pulse in order to optimize the resulting plasma wake. This could result in increases to the acceleration gradient, acceleration length, efficiency, and charge that can be loaded in the wake. The results in this

chapter show proof of principle but were not optimized. Finally, one can keep exploring the use of these methods for positron acceleration. This would require adjustment to the QPAD algorithm to ensure that it always converges within the on-axis electron filament.

In Chapter 7, the implementation of GPU acceleration in the particle-in-cell code OSIRIS is described. Though this algorithm was not used for the main simulations of this dissertation, these computational improvements may be beneficial for a variety of plasma simulations. Detailed performance benchmarks for thermal plasmas are presented, which demonstrate excellent weak scaling on NERSC’s Perlmutter supercomputer and high levels of absolute performance. The robustness of the code to model a variety of physical systems is demonstrated via simulations of Weibel filamentation and LWFA run with dynamic load balancing. Finally, measurements and analysis of energy consumption are provided that indicate that the GPU algorithm is ~ 14 times faster and is ~ 7 times more energy-efficient than the optimized CPU algorithm on a node-to-node basis. The described development addresses the PIC simulation community’s computational demands both by contributing a robust and performant GPU-accelerated PIC code and by providing insight on efficient use of GPU hardware. Several future directions are possible. First, the field solver can be implemented on the GPU. This would increase the performance of the field solver (which can be significant for LWFA runs using customized stencils which are often 16th-order longitudinally). It would also eliminate the need for copying field data back and forth every timestep. Implementing this in a way which preserves the full suite of available field solvers is a software challenge, which is why this was not implemented originally. More memory bottlenecks could be removed by implementing the particle and field boundary conditions on the GPU. Similarly, it would require a huge amount of effort to implement this while preserving all existing features. Other simulation features and modes (ionization, collisions, QED, quasi-3D, GR) could also be integrated with the GPU algorithm. A technical document written by the primary developers with suggestions on how to do this has been added to the main OSIRIS repository. These tasks could potentially be facilitated with the help of modern large language models. Finally,

performance on simulations involving a moving window (such as almost all LWFA simulations) could be significantly improved by using the method of moving tiles described in this chapter.

APPENDIX A

Detailed Guide to Running Lorentz-Boosted Simulations of LWFA

In this appendix we briefly review the concept of simulating LWFA and related processes in a Lorentz-boosted frame. So-called “Lorentz-boosted” simulations of laser wakefield acceleration are run in a Lorentz-boosted frame copropagating with the laser. The laser is doppler-redshifted and the plasma is Lorentz-contracted. This leads to a shorter light-crossing time and coarser longitudinal resolution in the boosted frame, yielding a computational speedup. Although Lorentz-boosted simulations turned out to not be a significant methodology in the research of this dissertation, getting them working in OSIRIS was a significant focus of the author’s doctoral research. This appendix explains a number of details involved in running Lorentz-boosted simulations, emphasizing their implementation in the PIC code OSIRIS. The material is adapted from a document jointly written with Kyle Miller, available in a README in the OSIRIS dev version as of September, 2024.

A.1 Introduction

Consider a particle-in-cell (PIC) simulation of a typical laser wakefield acceleration (LWFA) system in the lab frame. The largest degree of scale separation is typically given by the ratio of the final simulation time t_{max} to the timestep Δ_t . The latter is determined by the need to resolve the laser period, i.e., $\Delta_t \ll \lambda_0/c \equiv 2\pi/\omega_0$. Thus, the ratio of the largest to smallest scales is proportional to $\omega_0 t_{\text{max}}$. What would happen if we were to simulate the system in a

Lorentz-boosted frame with velocity β (and corresponding Lorentz factor $\gamma = 1/\sqrt{1 - \beta^2}$) copropagating with the laser in the z -direction? In this frame, the plasma is Lorentz-contracted by a factor of γ and has a drift velocity of $-\beta$, so that $t'_{\max} \sim t_{\max}/\gamma(1 + \beta)$. The laser wavelength is redshifted so that $\lambda'_0 = \gamma(1 + \beta)\lambda_0$. Thus, the temporal scale separation in the boosted frame is $\omega'_0 t'_{\max} = \gamma^2(1 + \beta)^2 \omega_0 t_{\max}$. Thus, because of the reduced temporal scale separation, one would expect a computational speedup on the order of $4\gamma^2$.

The Lorentz factor γ cannot be increased arbitrarily. Reasons for this will be discussed below. However, we will see that γ can potentially be increased for typical LWFA problems to $\gamma > 10$, which would yield a theoretical speedup of more than 400. So, what's stopping us from simulating everything in a boosted frame? There are several reasons. One challenge is the Numerical Cherenkov Instability (NCI), which, if unsuppressed, can lead to exponential growth of fields on timescales shorter than $t'_{\max}$. In addition, sometimes the physics of interest has a higher resolution requirement in the boosted frame (e.g., backward-going light, which is blueshifted); in principle the physics can be resolved with sufficient resolution, but this may limit or outweigh the benefit of boosting. In the boosted frame, the number of particles may also be too low to study statistical processes such as self-injection in LWFA. Finally, there are a number of complications for simulation initialization and diagnostics which must be considered.

The complications of the NCI, initialization, and diagnostics are all addressed in OSIRIS. Hence, Lorentz-boosted simulations of LWFA or other processes benefiting from boosting can be run. However, when running in this simulation configuration, it is important to understand the subtleties of these complications. The purpose of this appendix is to explain these subtleties.

A.1.1 What is the maximum γ that can be applied?

We begin by considering the maximum γ , because this gives a sense of the potential speedup that can be obtained with boosted simulations. Again, we consider the case of LWFA. For the NCI-mitigating field solvers in OSIRIS, the maximum γ is constrained by the CFL condition: Δ'_t will become limited by the transverse cell size $\Delta'_x = \Delta_x$ instead of by the longitudinal cell size Δ'_z as the boost velocity is increased. Thus, the maximum speedup occurs approximately in the frame where $\Delta'_z = \Delta'_x$. For a typical lab-frame resolution of $\Delta_z = \lambda_0/30$ and $\Delta_x = k_p^{-1}/10$, the maximum speedup will occur for a boost of $\gamma = \frac{1}{4} \left(\frac{k_0}{k_p} \right)$, which would yield a theoretical speedup of $\frac{1}{4} \left(\frac{k_0}{k_p} \right)^2$. Such a speedup could be very large for realistic laser parameters, potentially enabling simulations that would otherwise be impossible.

In practice, the speedup is generally significantly less than the heuristic theoretical speedup of $\frac{1}{4} \left(\frac{k_0}{k_p} \right)^2$. This is because the true speedup is lower when one performs a more detailed analysis of the theoretical speedup and also because of IO and other sources of simulation overhead. For example, at the time of writing (August 2024), replications of a simulation from Ref. [1] yield a heuristic theoretical speedup of ~ 300 , a more-accurate theoretical speedup of ~ 70 , and an actual speedup of ~ 20 .

A.1.2 When is a Lorentz-boosted simulation beneficial?

A necessary condition for the boosted frame to be useful is that the system timescale is decreased and the simulation timestep is increased in the boosted frame. Because of the constraint of the CFL condition, the transverse cell size must also be larger than the longitudinal cell size in the lab frame. In addition, backward-going light in the lab-frame will become blueshifted under Lorentz transformation; therefore, if backward-going light plays a significant role in the dynamics, the Lorentz factor will be constrained by its resolution. These restrictions constrain the problems for which a boosted-frame simulation leads to a speedup. LWFA simulations typically can benefit, but may not benefit if a high transverse

resolution is required. Some work has been done to permit small transverse cell sizes by developing solvers which do not require $c\Delta t < \Delta x_{\perp}$ (e.g., Ref. [95]). PWFA simulations typically do not benefit from boosting, but may under special circumstances, e.g., if a laser is included as in Ref. [271].

A.1.3 Brief review and current status in the field

A complete review of work on the problem is beyond the scope of this appendix, but here we include some brief details. The idea of carrying out PIC simulations of physics problems in different reference frames originated in studies of shocks. It was recognized in the early 1990s that simulation of LWFA in a boosted frame could lead to very large speedups [83]. Initial unpublished work found that the method worked in 1D Cartesian geometry. However, in 2D and 3D, the simulation fields were observed to be numerically unstable. Interest in the method resurged after publication of Ref. [84]. The numerical instability was eventually identified as the Numerical Cherenkov Instability (NCI) [88, 89]. The NCI had been first analyzed in a different context by B. Godfrey in Ref. [86]. There was not much progress in dealing with the NCI until the late 2000s, when some results were published in which aggressive current filters were applied [85]. Throughout the 2010s, various field solvers were developed which were designed to mitigate the NCI (and numerical dispersion more generally) [90, 91, 92, 93, 94, 95, 59, 96].

At this point, the NCI has largely been addressed. However, at the time of writing in 2025, Lorentz-boosted simulations of plasma-based acceleration remain under-utilized. Here are a number of the author’s guesses as to why this may be:

- Not all PIC codes include NCI-suppressing field solvers
- Not all PIC codes include algorithms for laser, beam, and plasma injection in a boosted frame
- Not all PIC codes support boosted diagnostics for lab-frame time-slice reconstruction

in a boosted frame

- The simulation initialization is significantly more complex than in the lab frame
- There can be subtle differences in the physics which require experience and familiarity with the method to understand and interpret

Although the NCI was historically the main bottleneck to getting boosted-frame plasma-based acceleration (PBA) simulations to run, these subtle details of initialization and diagnostics are still nontrivial. Almost all of the rest of the appendix discusses technical details of initialization and diagnostics.

A.2 Initialization of lorentz-boosted simulations

Here we give an overview of mathematical and computational considerations for various aspects of initializing Lorentz-boosted PIC simulations with an emphasis on features supported in OSIRIS. In all of what follows, the boost and laser propagation are in the $\hat{z}$ direction.

A.2.1 Field solvers

As discussed in Sec. A.1, the NCI induced by a quasi-neutral drifting plasma is the primary numerical challenge to running Lorentz-boosted simulations of PBA. OSIRIS provides several field solvers for suppression of the NCI. One example is the “bump” solver, which modifies the dispersion relation of electromagnetic waves directed along the axis at the specific wavenumber where the NCI occurs [59]. This solver was specifically designed to suppress the NCI. Another example is the “dual” solver, which was designed for general-purpose reduction of numerical dispersion for electromagnetic waves along the axis [61]. Although it was not specifically designed for suppression of the NCI, it was later discovered that it naturally suppresses the NCI even more effectively than the bump solver. At the time of writing, it is recommended to use the dual solver for boosted-frame simulations in OSIRIS. Reference [271] is the first

published work employing the dual solver in Lorentz-boosted simulations.

In the author’s experience, the default dual solver parameters are sufficient for the vast majority of cases. If the NCI is still suspected to be an issue, it may be able to be addressed by changing the solver parameters. The normal procedure for doing this is to use a script to compute and visualize the numerical dispersion relation of the solver. The dispersion relation depends on the solver parameters as well as the plasma density, boost velocity, timestep, and cell sizes. One can try experimenting with the parameters to find a field solver with a dispersion relation such that the NCI growth rate is sufficiently low. The script for doing this is not included here. This procedure should be unnecessary for the vast majority of users.

A.2.2 Boundary conditions

The author recommends to always run Lorentz-boosted simulations of PBA in a moving window. This is typically the most natural choice, since the Lorentz transformation of a moving window in the lab frame is a moving window in any boosted frame. In addition, the large plasma outflow can lead to numerical instability at the left wall when using a stationary box. Using a moving window addresses this problem because the outflow becomes causally disconnected from the simulation domain. For the transverse boundary conditions, thermal boundary conditions can be used. Leaving a small gap between the plasma and the transverse boundary can prevent numerical instability there.

A.2.3 Simulation box boundaries

The simulation box size can be determined by applying Lorentz transformations to the boundaries of the volume of spacetime simulated in the lab frame. In what follows, the boosted-frame and lab-frame coordinates will be respectively primed and unprimed. We will work everything out in the case that both simulations are run in a window moving at the speed of light, but the calculations can be repeated for a general case (although this

configuration is not recommended). We work in units where $c = 1$ for simplicity.

Suppose that the lab-frame simulation starts with $z \in [z_-, z_+]$, runs for $t \in [t_-, t_+]$, and runs with a moving window. (Typically, $t_- = 0$.) Assume also that the simulation can be initialized at any point outside the causal light cone of an interaction point (t_*, z_*) – for example, in an LWFA simulation, this is the point in spacetime at which the laser hits the plasma. Assume that the laser and the plasma can be analytically initialized outside of a light cone propagating with speed c from this point. It will become clear why this is important in the sections on initialization of laser pulses and particle beams.

What are the spatiotemporal boundaries of a Lorentz-boosted simulation to contain all the dynamics beyond the interaction point? We can find them algebraically by applying Lorentz transformations. First, what is the time t'_+ up to which the boosted-frame simulation must be run? The boundaries of the simulation in spacetime at the end of the lab-frame simulation are $(t_{b\pm}, z_{b\pm}) = (t_+, z_{\pm} + (t_+ - t_-))$. Applying a Lorentz transformation, the boosted coordinates are $(t'_{b\pm}, z'_{b\pm}) = (\gamma t_+ - \gamma\beta(z_{\pm} + (t_+ - t_-)), \gamma(z_{\pm} + (t_+ - t_-)) - \gamma\beta t_+)$. Of these values, the one associated with z_- has a larger t' . This t' is the point up to which the boosted-frame simulation must be run if we are to capture the entirety of the t_+ timeslice. Therefore, $t'_+ = \gamma(1 - \beta)t_+ - \gamma\beta(z_- - t_-)$.

Next, what is the time t'_- at which the boosted simulation must be started? This comes from the interaction point. In the boosted frame, the interaction point occurs at $(t'_*, z'_*) = (\gamma t_* - \gamma\beta z_*, \gamma z_* - \gamma\beta t_*)$. Thus, the simulation must be started at (or before) $t'_- = t'_* = \gamma t_* - \gamma\beta z_*$.

Finally, the box boundaries for the boosted simulation can be calculated from their known final positions. We know that, in the boosted frame, the box boundaries at time t' are $z'_{b'\pm} = z'_{\pm} + (t' - t'_-)$ for a box with $z' \in [z'_-, z'_+]$ and with t'_- given above. To determine the initial box boundaries in the boosted frame $z'_{\pm}$, we can equate the above spacetime coordinates corresponding to the boosted-frame box with the transformed coordinates of the lab-frame box given previously, or $(t', z'_{b'\pm}) = (t'_{b\pm}, z'_{b\pm})$. Solving this equation for the initial

boosted-frame box boundaries results in $z'_\pm = t'_- - t'_{b\pm} + z'_{b\pm} = t'_- + \gamma(1 + \beta)(z_\pm - t_-)$.

All this has been worked out in full generality. In the author's experience, the most convenient way to center the simulation is to choose $t_- = 0$ and $(t_*, z_*) = (0, 0)$ in the lab frame. When this is done, the box boundaries in the boosted frame simplify to $z'_\pm = \gamma(1 + \beta)z_\pm$, with $t'_- = 0$ and $t'_+ = \gamma(1 - \beta)t_+ - \gamma\beta z_-$. The box is therefore longitudinally expanded by a factor of $\gamma(1 + \beta)$. Note that $\gamma(1 - \beta) = 1/\gamma(1 + \beta)$, so that the first term in the t'_+ expression represents a Lorentz contraction of the simulation timescale. This is the term corresponding to the heuristic timescale estimate in the introduction from length contraction plus the drift. The more detailed analysis here shows that there is an additional $-\gamma\beta z_-$ term (note that $z_- < 0$) which adds a penalty to increasing γ .

The above analysis was worked out purely algebraically. It is also possible to visualize what the simulation boundary looks like in terms mappings of parallelograms in (t, z) space between the lab and boosted frame. Some relevant visualizations of these parallelograms can be found in Ref. [94]. The mapping of the parallelogram in the lab frame, starting and ending with constant t (i.e., $t = t_\pm$), maps to a tilted parallelogram in the boosted frame with non-constant t' at the beginning and end. We have computed what the coordinates of an enclosing parallelogram need to be (i.e., such that there is constant $t' = t'_\pm$) in the boosted frame such that the tilted parallelogram is surrounded by a non-tilted parallelogram.

A.2.4 Resolution: Δ'_z , Δ'_x , Δ'_t , and particles per cell

The transverse dimensions are not affected by Lorentz transformations. Thus, the dynamics take place over the same length scale in the boosted frame and require the same resolution as in the lab frame: $\Delta'_x = \Delta_x$.

As in the lab frame, the longitudinal cell size must be chosen to resolve the finest spatial scale. In the case of a laser, this is the doppler-shifted wavelength $\lambda'_0 = \gamma(1 + \beta)\lambda_0$. Thus, for a laser, one would naturally choose $\Delta'_z = \gamma(1 + \beta)\Delta_z$. Note that $\gamma(1 + \beta)$ is the same factor

for the overall expansion of the box longitudinally, so that the number of cells longitudinally is the same for the case of a laser.

The timestep is constrained by the required temporal resolution and the CFL condition. The former is implied by the latter in the case of PBA problems. As noted in the introduction, the CFL condition places a constraint on the maximum possible boost.

The number of macroparticles per cell, which effectively controls resolution of the particle distribution functions, is more subtle. The number of physical particles (not macroparticles) is a Lorentz invariant. The total physical charge is the number of physical particles times the charge per particle, independent of their motion. Thus, the total charge is a Lorentz invariant. However, the plasma is Lorentz contracted by a factor of γ . In addition, Δ_z is larger by a factor of $\gamma(1 + \beta)$. Thus, if the same number of macroparticles per cell is used as in the lab frame, then the total number of macroparticles is reduced by a factor of $\gamma^2(1 + \beta)$. This is an important consideration for processes such as self-injection of particles in LWFA, which may occur with few particle statistics in the lab frame.

The charge of each macroparticle, which is the total charge divided by the number of macroparticles, is also higher by the same factor of $\gamma^2(1 + \beta)$. Thus, each macroparticle has a higher response to fields. The fields themselves are also different because of the Lorentz transformation. For example, one can show that the focusing field $E_x - B_y$ will transform as $\gamma(1 + \beta)(E_x - B_y)$. Thus, the total focusing force $F_x = q(E_x - B_y)$ on the macroparticles will be $\sim \gamma^3(1 + \beta)$ times the value in the lab frame.

The above scalings assume that we use the same number of macroparticles in the boosted frame as in the lab frame. But one does not necessarily have to do this. How many should be used? In the author's experience, 32 particles per cell is typically sufficient for modeling LWFA in 2D and quasi-3D geometries in the lab frame. A similar rule of thumb is not known to the author for a simulation in the boosted frame. Although the answer is not clear in general, using 32 is a natural starting point. However, for simulating a particular physical process, there is a clear answer: a sufficient number of particles are required for numerical

convergence. One ought to increase the particles per cell until the results do not change. That is the number of particles that are required for any PIC simulation. Hopefully a general rule of thumb will emerge as the community makes more frequent use of this simulation method. Certain processes such as self-injection in LWFA may require more particles per cell in the boosted frame than in the lab frame.

A.2.5 Estimation of the simulation cost

Given the simulation box size and resolution parameters, the cost of the simulation in both the lab frame and boosted frame can be estimated. This can be done by counting the total number of particle pushes, assuming that particle pushing dominates the simulation cost and that there is no imbalance in the number of particles among different processors. Both of these assumptions are not true in practice, but this approximation is a useful point of reference.

Under these approximations, the cost of the simulation in both the lab frame and the boosted frame can be approximated as $\text{Cost} \propto N_{\text{ppc}} \left(\frac{L_t}{\Delta_t} \right) \left(\frac{L_z}{\Delta_z} \right) \left(\frac{L_x}{\Delta_x} \right)^{d_\perp}$, where N_{ppc} is the number of particles per cell, L_j denotes the total time or length in each dimension, and $d_\perp$ is the transverse dimensionality. From the above calculations, neither L_x nor Δ_x is changed, so the last factor contributes no speedup. L_z and Δ_z are changed by the same factor, so this factor also contributes no speedup. Thus, the theoretical speedup all comes from the particles per cell and the temporal factor and is given by $\left(\frac{N_{\text{ppc}}}{N'_{\text{ppc}}} \right) \left(\frac{\gamma(1+\beta)L_z}{\gamma(1-\beta)L_z - \gamma\beta z_-} \right)$. If the particles per cell are the same and the $\gamma\beta z_-$ term is neglected, this gives the heuristic speedup of $\gamma^2(1+\beta)^2$ computed in the introduction.

A.2.6 Background plasma initialization

The initialization of the plasma in the boosted frame must account for the Lorentz contraction of the plasma as well as the relativistic drift of the plasma. The former is easily handled

by changing the length of the plasma in the usual way. The charge density must also be increased by a factor of γ because the total physical charge is conserved.

The drift of the plasma adds a nontrivial complication when injecting from a moving wall. In the standard moving-wall injection algorithm of OSIRIS, new simulation particles are injected into new cells whenever the latter are added to the right boundary. When the particles are cold, they will not evolve before the injection of new particles during a later timestep. This is still approximately true if the velocity is nonrelativistic in the axial direction. However, when the axial velocity is relativistic, injected particles will evolve significantly by the time later particles are injected. In particular, if the window does not move on each timestep, then particles that are moving to the left may move such that the cell they were injected into is empty by the time new particles are injected into the next cell. This leads to a high-frequency longitudinal density modulation in what should be a uniform plasma. This problem occurs whenever particles are continuously injected with a relativistic axial velocity, independently of the plasma temperature. For example, the problem occurs both for a cold relativistically drifting plasma and for a non-drifting relativistic thermal Maxwellian.

In order to resolve this, the injector must account not just for whether it has crossed a sufficient number of cells for injection, but also the shift of the particles during the time between injections. This is accomplished by the “cathode” module in OSIRIS. The cathode must be used for injecting particles in boosted simulations when using a moving window. It is currently supported in quasi-3D and Cartesian geometries.

One final note about the plasma initialization: in the current implementation, ion particles must be included to neutralize the plasma. This is unnecessary in lab-frame simulations, where injecting electrons with no ions is equivalent to assuming a current-free neutralizing ion background. It may be possible to write a new current deposition algorithm that assumes a stationary drifting ion background so that the inclusion of ions in the boosted frame is not necessary. For the time being, the inclusion of ion particles leads to another factor of ~ 2 slowdown for the boosted simulation compared to a lab-frame simulation without ions. The

same number of particles per cell should be used for ions and electrons to avoid numerical artifacts.

A.2.7 Electromagnetic field initialization (including lasers)

Boosting changes two properties of the initial electromagnetic fields: (1) the spatiotemporal structure of the fields and (2) the strength of the fields. Here we discuss how both are accounted for in the simulation initialization. We focus on the case of laser initialization. For simplicity, we consider the case where the structure of the vacuum laser field (i.e., before entering plasma) in the lab frame is of the separable form $\mathbf{E}_\perp(t, z, \mathbf{x}_\perp) \approx E_0 B(z - t) C(z, \mathbf{x}_\perp) \exp(ik_0(z - t)) \hat{\mathbf{e}}$. Here $B(z - t)$ is the longitudinal profile, $C(z, \mathbf{x}_\perp)$ is a paraxial beam solution, $\exp(ik_0(z - t))$ is the fast phase factor, and $\hat{\mathbf{e}}$ is the laser polarization vector. This functional form is commonly used to initialize lab-frame laser fields in various simulation codes. Details on when this approximate solution is valid are discussed in Chapters 4 and 5. We will specifically consider the case that $C(z, \mathbf{x}_\perp)$ is a Gaussian beam with spot size w_0 and focal point at $(z, \mathbf{x}_\perp) = (0, \mathbf{0})$.

We begin by calculating the structure of such a field in the boosted frame. Applying a Lorentz transformation, the electric field in the boosted frame is given by $\mathbf{E}'_\perp(t', z', \mathbf{x}'_\perp) = \gamma(1 - \beta)E_0 B(\gamma(1 - \beta)(z' - t')) C(\gamma z' + \gamma\beta t', \mathbf{x}'_\perp) \exp(ik_0\gamma(1 - \beta)(z' - t')) \hat{\mathbf{e}}$. Thus, in the boosted frame, the duration of the pulse is increased by a factor of $1/\gamma(1 - \beta) = \gamma(1 + \beta)$. The amplitude of the transverse field is also reduced by a factor of $\gamma(1 - \beta)$. The laser wavenumber is decreased by a factor of $\gamma(1 - \beta)$ so that the wavelength is doppler-redshifted by a factor of $\gamma(1 + \beta)$.

The most notable change in the field structure comes from the paraxial factor $C(z, \mathbf{x}_\perp)$. In the lab frame, the spot size of the laser pulse is $w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R}\right)^2}$, where $z_R \equiv \frac{1}{2}k_0 w_0^2$ is the Rayleigh range. Therefore, the spot size of the laser pulse in the boosted frame at a given (t', z') is $w(z(t', z')) = w(\gamma z' + \gamma\beta t')$. Thus, in the lab frame, the spot size depends

only on space (z); in the boosted frame, it depends on both space (z') and time (t').

A few asides are worth mentioning about the laser field in the boosted frame. First, in the boosted frame, the pulse can be described as possessing spatiotemporal coupling, which is defined as the inability of $\mathbf{E}'_{\perp}$ to be expressed in a separable functional form (unlike in the lab frame). Second, we see that the point z' in spacetime at which the spot size is minimal, i.e., at which the pulse is in focus, now varies with t' . The focus occurs in the lab frame at $z = 0$, which in the boosted frame is the trajectory $z' = -\beta t'$. This means that the pulse has a time-varying focal point in the boosted frame. A time-varying focal point is equivalent to saying that the pulse is a so-called “flying-focus” pulse. This latter insight was first published in Ref. [272].

Returning to the initialization: in the lab frame, we ordinarily initialize the laser pulse at $t = 0$ (we take $t_- = t'_- = 0$ for simplicity in this discussion). The laser is initialized in vacuum and propagates into the plasma. What happens if we follow the same initialization approach at $t' = 0$ in the boosted frame? Suppose that the pulse tail occurs at $z = z_0 < 0$ for $t = 0$ in the lab frame, i.e., $B(z - t) = 0$ for $z < z_0$ at $t = 0$. In the boosted frame, the rear position of the laser occurs at $z' = z'_0 = \gamma(1 + \beta)z_0$. What is the spot size of the laser at $(t', z') = (0, z'_0)$? We compute $w(z(t', z')) = w_0 \sqrt{1 + \gamma^2(1 + \beta)^2 z_0^2 / z_R^2}$. Thus, the spot size will be significantly larger than w_0 when $\gamma(1 + \beta)z_0 > z_R$.

Why is this a problem? The transverse box size needs to fit the laser if it is initialized during a single timestep. Thus, if one is to fit the laser in the box at $t' = 0$, one would have to make transverse box size very large in order to fit the laser, even though no dynamics of interest take place outside the lab-frame upper transverse boundary. This increases the simulation cost.

Another option is to dynamically initialize the laser pulse, where only the part that is close to being in focus is injected. From the earlier calculation, this occurs in the vicinity of the trajectory $z' = -\beta t'$. One can either initialize all the field cells with $z' < -\beta t'$ according to the structure of the boosted laser field, or one can just initialize a narrow region

behind $z' = -\beta t'$, since cells to the left will eventually be overwritten later anyway. The latter algorithm has been referred to as a “moving wall injector.” This feature is currently implemented in OSIRIS. Note that such a moving-wall injector does not override the physics so long as it initializes the laser pulse fully outside the causal light cone of the interaction point (t'_*, z'_*) , i.e., all points (t', z') in the boosted frame such that $t' < t'_*$ and $z' - z'_* < t' - t'_*$. In WarpX and a previous OSIRIS implementation, this was done by propagating a slab of simulation particles along this trajectory and setting their momenta appropriately to generate the laser field. In the current version of OSIRIS, the fields are set directly. The moving-wall injector can be used to inject the pulse in the boosted frame along either a constant lab-frame z or t . Both options are possible in the current OSIRIS implementation. Constant t is recommended so that the initial condition can be verified to be correct by comparing the boosted diagnostics with a lab-frame simulation. This also results in the initial laser fields being visible at $t = 0$ in boosted diagnostics.

The final subtlety for laser field initialization is setting the axial field. Getting the axial field right is potentially more important in the boosted frame than in the lab frame. In the lab frame, the laser propagates out of the region of initialization and into the plasma, so that any numerical errors associated with not satisfying $\nabla \cdot E = 0$ at initialization exit the simulation. This is not true in the boosted frame, because the plasma drifts into the region where the laser was injected.

In practice, the axial field in the lab frame is normally computed in OSIRIS through a so-called “divergence correction.” In this method, the transverse field is set in some way, such as by using the above formula for the laser field. The axial field is then set by numerically integrating Gauss’s law in vacuum, starting from before the head of the laser pulse. This same procedure can be done in the boosted frame, but now must be done along the moving-wall boundary and at every timestep. This makes it more complicated to implement. For simulations where the contribution of E_z of the laser is negligible in the boosted frame, it can also be neglected in the boosted frame.

All of the above discussion applies to laser fields. What about for more-general electromagnetic fields? If it is possible to write the fields down analytically outside the light cone of the interaction point, then the same procedure can be followed to derive the structure of the fields in the boosted frame. The solution of using a moving wall should work for all cases in which the fields can be initialized fully in the box at $t = 0$ in the lab frame or at the plane $z = 0$ in the lab frame. However, a moving-wall injector may not always be necessary. For example, if a function is applied for external fields, it can be Lorentz transformed directly (amplitude and spatiotemporal dependence) without any other simulation changes.

A.2.8 Particle beam initialization

For Lorentz-boosted simulations including a particle beam, similar considerations apply as for a laser pulse. By the same reasoning, the beam spot size at the tail will be significantly larger than at the focal point if $\gamma(1 + \beta)z_0 > \beta^*$, where $\beta^* \equiv \sigma_x/\sigma_{x'}$ is the analogue of the Rayleigh range for a beam and z_0 is the location of the tail of the beam at $t = 0$ in the lab frame. Thus, the same issue with the transverse box size applies as for a laser when initializing the beam at $t' = 0$. Two possibilities are implemented in OSIRIS: injection using the standard algorithm at $t' = 0$ in OSIRIS or “thawing beam injection,” the analogue of a moving wall injector for a laser. In either case, the momentum, charge density, and length of the beam need to be adjusted from their values in the lab frame. These are discussed in the following subsections.

A.2.8.1 Standard injection

If the beam spot size at the tail and head are comparable at $t' = 0$, then the beam can be reasonably approximated using the standard OSIRIS initialization method. This involves simply increasing the beam duration by a factor of $\gamma(1 + \beta'_z\beta) = [\gamma(1 - \beta_z\beta)]^{-1}$ and decreasing the charge by the same factor. The beam is fully initialized at $t' = 0$.

A.2.8.2 Thawing beam injection

If the above condition does hold (i.e., the beam spot size varies significantly across the beam at $t' = 0$), a kind of moving-wall injector is needed. One possibility is to use a moving-wall injector analogous to the case of the moving-wall laser injector described above: simulation particles are injected and their corresponding electromagnetic fields are set at each timestep. However, the fact that the beam is constituted by simulation particles enables a simpler injection algorithm. This method may be described as a “thawing-beam” initialization. In this method, the beam is fully injected at $t' = 0$ with each axial slice initialized with the profile the beam has at $t = 0$ rather than at $t' = 0$. Thus, the beam is initialized at $t' = 0$ differently from how it should actually look if the evolution were fully self-consistent. The beam’s transverse evolution is neglected, i.e., the beam advances only according to its longitudinal momentum, until it has reached the lab-frame $t = 0$ timeslice (i.e., $t' \geq -\beta z'$). The full particle evolution is then unfrozen. Keeping the particles moving in the z -direction before thawing is analogous to the particle cathode: this prevents the problem of high-frequency density modulations if the particles are unfrozen from rest in the axial direction. Keeping the particles moving axially before unfreezing also enables the particle beam fields to be set self-consistently at $t' = 0$. This method of initialization is implemented in OSIRIS.

How should the beam be initialized when using the thawing-beam method? We must consider the initial particle locations, density, and momentum distribution in the boosted frame. Let us first consider the initial particle locations. Consider a slice of particles located at z_{inj} in the lab frame at $t = t_-$. In the boosted frame, the transverse coordinates are set to whatever they would be when the beam is in focus. What about the axial coordinate? In the thawing scheme, the axial position of the particles for $t' < t'_{\text{inj}}$ is given by $z'(t') = z'_0 + \beta'_z t'$, where z'_0 is the initial particle location that we must determine and β'_z is given from the usual velocity-addition formula as $\beta'_z = \frac{\beta_z - \beta}{1 - \beta\beta_z}$. Applying a Lorentz transformation to the lab-frame injection coordinates (t_-, z_{inj}) , we find that the injection in the boosted frame will occur at $t'_{\text{inj}} = \gamma t_- - \gamma\beta z_{\text{inj}}$ and $z'_{\text{inj}} = \gamma z_{\text{inj}} - \gamma\beta t_-$. Now, what does z'_0 need to be such

that the particles arrive at this $(t'_{\text{inj}}, z'_{\text{inj}})$ given the above trajectory? Equating $z'(t'_{\text{inj}}) = z'_{\text{inj}}$ allows us calculate $z'_0 = \gamma(1 + \beta'_z \beta) z_{\text{inj}} - \gamma(\beta'_z + \beta) t_-$. When $\beta'_z > 0$ (i.e., the boost does not reverse the propagation direction of the bunch), we see from the first term that the bunch is Lorentz-expanded by a factor of γ , with an additional factor of $(1 + \beta'_z \beta) \in [1, 2)$. From conservation of charge, we see that the charge density must be lower in the boosted frame by the same factor of $\gamma(1 + \beta'_z \beta)$, which can also be written as $[\gamma(1 - \beta_z \beta)]^{-1}$. All of these calculations hold for the initial particle locations, independently of whether the beam has a spread in its transverse or longitudinal momentum.

Note that, as in the case of a laser, the beam can be injected in the boosted frame at a constant lab-frame t or z . The above calculation was for a constant t . The calculations could be repeated for a constant z .

A.2.8.3 Beam momentum

We now consider how the momentum distribution should be initialized for a relativistic beam (with or without the thawing-beam injector). In general, the four-momentum of a particle $p^\mu = (p^0(\mathbf{p}), \mathbf{p})$ is a rank-1 Lorentz tensor, where $p^0(\mathbf{p}) \equiv \sqrt{1 + \mathbf{p}^2}$ is the Lorentz factor of the particle. Thus, in the boosted frame, we have $\mathbf{p}'_\perp = \mathbf{p}_\perp$ and $p'_z = \gamma p_z - \gamma p^0(\mathbf{p})$. (Note that γ is the Lorentz factor of the boost and $p^0(\mathbf{p})$ is the Lorentz factor of the particle). Thus, $\mathbf{p}_\perp$ is unaffected by the boost and should be sampled from the same distribution as in the lab frame. However, because of the appearance of $p^0(\mathbf{p})$ in the equation for p'_z , the distribution of p'_z now depends on the distribution of both the lab-frame momenta p_z and $\mathbf{p}_\perp$. This complicates the distribution function in the boosted frame. A robust way to mitigate this complication is to sample the particle momenta from the lab-frame distribution function and to subsequently apply a Lorentz transformation using the above formula to compute the new momenta in the boosted-frame.

An approximation for the distribution of p'_z can be derived in the limit that $\sigma_{p_z} \ll p_{z0}$ and $\sigma_{p_\perp} \ll p_{z0}$, where p_{z0} is the mean p_z for the beam in the lab frame. These approximations

typically hold in simulations of LWFA and PWFA. Such an approximation is useful both because it simplifies the initialization and also because it gives intuition about the structure of the beam in the boosted frame. Under these assumptions, we can approximate $p^0(\mathbf{p}) \approx p_z$. For each particle, we decompose $p_z = p_{z0} + \Delta p_z$. We can then approximate $p'_z \approx \gamma(1 - \beta)p_z = \gamma(1 - \beta)p_{z0} + \gamma(1 - \beta)\Delta p_z$. Thus, under this approximation, we have $p'_z = p'_{z0} + \Delta p'_z$, where $p'_{z0} = \gamma(1 - \beta)p_{z0}$ and the distribution for $\Delta p'_z$ is the same as it was before but scaled by $\gamma(1 - \beta)$. For example, if the distribution for Δp_z in the lab frame is a Gaussian with standard deviation σ_{p_z} , in the boosted frame it would have $\sigma_{p'_z} = \gamma(1 - \beta)\sigma_{p_z}$. The relative error in this approximation is roughly $(p^0(\mathbf{p}) - p_z)/p_z \approx \left(\frac{1 + \sigma_{p_\perp}^2}{p_{z0}^2}\right) \left(1 - 2\left(\frac{\sigma_{p_z}}{p_{z0}}\right)\right)$.

In the current OSIRIS implementation, the Lorentz transformation for the momentum distribution is not done automatically, so one must use the approximation in the previous paragraph. It would not be difficult to implement the more-robust algorithm for initializing the momentum distribution.

A.3 Boosted diagnostics

In the lab frame, we typically output simulation data at fixed timeslices of the simulation, e.g., $\omega_p t = 0, 10, 20, \dots$. A challenge with running boosted-frame simulations is that timeslices in the lab frame do not correspond to a single time in the boosted frame because of the spatial dependence of the Lorentz transformation, i.e., the fact that $t' = \gamma t - \gamma \beta z$. Simultaneity is not preserved under Lorentz transformations. Conversely, dumping data at constant timeslices in the boosted frame is not sufficient to reconstruct timeslices in the lab frame. (One could get around this by dumping the entire boosted simulation dataset and reconstructing the lab-frame timeslices in postprocessing, but this would be prohibitively expensive, both in memory and in IO cost, for most problems of interest.)

Why is it necessary to have lab-frame timeslices from a boosted simulation? Strictly speaking, it is not. The timeslices from the boosted frame contain interpretable physics. They

are also the more-useful diagnostics for debugging initialization or numerical issues. However, it is very useful to also have the capability of reconstructing lab-frame timeslices for two reasons: (1) this allows direct comparison with lab-frame simulations, which have now been stress-tested for several decades, and (2) scientists studying LWFA are far more familiar with observing the results in the lab frame. Lab-frame reconstructions may eventually become unnecessary as computational physicists hone the art of running and analyzing boosted-frame simulations. At the time of writing, they are necessary. It could also be useful to translate lab-frame data into boosted-frame data, both for comparison and for insight into the physics.

How can lab-frame timeslices be reconstructed during a simulation? The problem is analogous to the problem of injecting fields or particles at the lab-frame $t = 0$ timeslice within the boosted frame. In that case, a moving wall is used to initialize the simulation dynamically in spacetime. The analogous solution for boosted diagnostics is effectively to use a “moving-wall receptor,” instead of a moving-wall injector, to dynamically pull data out of the simulation for the reconstruction of a specified timeslice. We refer to such diagnostics as “boosted diagnostics.”

Boosted diagnostics are currently implemented for grid quantities (electromagnetic fields and covariant particle distribution function moments) and raw particle data in OSIRIS. The boosted diagnostics are supported in Cartesian geometries and in quasi-3D. Boosted slice diagnostics have also been implemented, which is important for reducing the cost of diagnostic output in 3D simulations. Additional boosted particle diagnostics (higher-order covariant moments of the distribution function) are not yet supported.

A.3.1 Code details

One of the challenges of boosted diagnostics that does not apply to the moving-wall injection algorithms is that the timeslices being reconstructed do not necessarily snap to the gridpoints in the boosted frame. Thus, interpolation onto the desired lab-frame timeslice is required in both space and time in the boosted simulation. To perform the interpolation, grid data from

the previous timestep is buffered. For a given lab-frame point (t, z) to be reconstructed, the four surrounding gridpoints in the boosted frame are used to linearly interpolate onto (t, z) . This data is then buffered. Because of memory constraints, all of the timeslices undergoing reconstruction cannot in general fit in memory. Thus, reconstructed data has to be dumped periodically. The dumping may occur before timeslices have been fully reconstructed. Thus, there are two decoupled dump frequencies: (1) the frequency of the lab-frame timeslices, which is determined by the necessary diagnostic resolution, and (2) the frequency of dumping of the buffered boosted diagnostics, which is determined by memory constraints.

All of this work is handled by two classes: in the code: a boosted grid manager class and a boosted diagnostic class. The boosted grid manager is responsible for handling coordinate conversions between the two frames and nothing more. Each boosted diagnostic object stores a pointer to a boosted grid manager object that informs which computations must be done at each timestep; the boosted diagnostic object handles buffering data from the simulation, performing, interpolation, and writing the data at specified intervals. The boosted grid manager will work with any γ , which is specified in the input deck.

It is important to point out that the spatiotemporal interpolation used is not consistent with the field solver used in the code. This would be a problem if the boosted diagnostics were coupled to the simulation, but since they are not, they should be close to the “true” simulation quantities (with quotes because these quantities are not actually attained in the simulation). With that said, the interpolation may be problematic for applications in which values of the simulation fields consistent with the numerical scheme are important, e.g., discovering equations from the data as in Ref. [273]. In such a case, more care could be taken to perform the spatiotemporal interpolation self-consistently with the field solver.

A.3.2 Other applications of boosted diagnostics

Aside from Lorentz-boosted simulations, there are other scenarios where boosted diagnostics may be beneficial. Two examples are a relativistic shock simulated in the shock frame or a

relativistic jet in its comoving frame. Boosted diagnostics may be insightful in these cases, for example if theoretical calculations are easier to carry out in the lab frame.

Another potentially useful application would be to forward-boost the data from a lab-frame LWFA simulation. This could be used to directly assess the quality of a given boosted-frame simulation if the lab-frame simulation is trusted.

Another useful feature made possible by boosted diagnostics is reporting of subsets of simulation data without boosting. For example, one can use the current implementation of the boosted diagnostics with $\gamma = 1$ to extract a high-resolution subvolume moving at arbitrary velocity. At present this would only work for subvolumes with full transverse dimensions, i.e., only a reduction in longitudinal spacing is possible; implementation of subvolume tracking with arbitrary transverse sizes would be possible using this framework.

REFERENCES

- [1] Wei Lu, M Tzoufras, C Joshi, FS Tsung, WB Mori, J Vieira, RA Fonseca, and LO Silva. Generating multi-gev electron bunches using single stage laser wakefield acceleration in a 3d nonlinear regime. *Physical Review Special Topics—Accelerators and Beams*, 10(6):061301, 2007.
- [2] Ricardo A Fonseca, Luis O Silva, Frank S Tsung, Viktor K Decyk, Wei Lu, Chuang Ren, Warren B Mori, S Deng, S Lee, T Katsouleas, et al. Osiris: A three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators. In *Computational Science—ICCS 2002: International Conference Amsterdam, The Netherlands, April 21–24, 2002 Proceedings, Part III 2*, pages 342–351. Springer, 2002.
- [3] Dustin H Froula, David Turnbull, Andrew S Davies, Terrance J Kessler, Dan Haberberger, John P Palastro, Seung-Whan Bahk, Ildar A Begishev, Robert Boni, Sara Bucht, et al. Spatiotemporal control of laser intensity. *Nature photonics*, 12(5):262–265, 2018.
- [4] Kyle G Miller. *Controlling Electron Acceleration in Underdense and Overdense Laser–Plasma Interactions to Generate X-Rays for Probing High Energy Density Material*. University of California, Los Angeles, 2022.
- [5] LM Gorbunov and VI Kirsanov. Excitation of plasma waves by an electromagnetic wave packet. *Journal of Experimental and Theoretical Physics*, 66:290, 1987.
- [6] Eric Esarey, Carl B Schroeder, and Wim P Leemans. Physics of laser-driven plasma-based electron accelerators. *Reviews of modern physics*, 81(3):1229–1285, 2009.
- [7] Jacob R Pierce, John P Palastro, Fei Li, Bernardo Malaca, Dillon Ramsey, Jorge Vieira, Kathleen Weichman, and Warren B Mori. Arbitrarily structured laser pulses. *Physical Review Research*, 5(1):013085, 2023.
- [8] Jacob R Pierce, Kyle G Miller, Fei Li, John P Palastro, and Warren B Mori. Laser wakefield acceleration driven by a discrete flying focus. *Physical Review Accelerators and Beams*, 28(10):101303, 2025.
- [9] Roman P Lee, Jacob R Pierce, Kyle G Miller, Maria Almanza, Adam Tableman, Viktor K Decyk, Ricardo A Fonseca, E Paulo Alves, and Warren B Mori. Acceleration of the particle-in-cell code osiris with graphics processing units. *Journal of Plasma Physics*, 91(1):E8, 2025.
- [10] CGR Geddes, K Nakamura, Guillaume R Plateau, Cs Toth, E Cormier-Michel, Eric Esarey, CB Schroeder, JR Cary, and Wim P Leemans. Plasma-density-gradient injection of low absolute-momentum-spread electron bunches. *Physical review letters*, 100(21):215004, 2008.

- [11] S Kalmykov, SA Yi, Vladimir Khudik, and Gennady Shvets. Electron self-injection and trapping into an evolving plasma bubble. *Physical review letters*, 103(13):135004, 2009.
- [12] Arthur Pak, KA Marsh, SF Martins, W Lu, WB Mori, and C Joshi. Injection and trapping of tunnel-ionized electrons into laser-produced wakes. *Physical review letters*, 104(2):025003, 2010.
- [13] X. L. Xu, F. Li, W. An, T. N. Dalichaouch, P. Yu, W. Lu, C. Joshi, and W. B. Mori. High quality electron bunch generation using a longitudinal density-tailored plasma-based accelerator in the three-dimensional blowout regime. *Physical Review Accelerators and Beams*, 20(11):111303, 11 2017.
- [14] TN Dalichaouch, XL Xu, Fei Li, Adam Tableman, FS Tsung, Weiming An, and WB Mori. Generating high quality ultrarelativistic electron beams using an evolving electron beam driver. *Physical Review Accelerators and Beams*, 23(2):021304, 2020.
- [15] Pisin Chen, J. M. Dawson, Robert W. Huff, and T. Katsouleas. Acceleration of electrons by the interaction of a bunched electron beam with a plasma. *Physical Review Letters*, 54(7):693–696, 2 1985.
- [16] C Joshi, S Corde, and WB Mori. Perspectives on the generation of electron beams from plasma-based accelerators and their near and long term applications. *Physics of Plasmas*, 27(7), 2020; and references therein.
- [17] Chandrashekhar Joshi, Warren B Mori, and Mark J Hogan. The positron arm of a plasma-based linear collider. *Nature Physics*, 21(6):885–894, 2025; and references therein.
- [18] CB Schroeder, F Albert, C Benedetti, J Bromage, D Bruhwiler, SS Bulanov, EM Campbell, NM Cook, B Cros, MC Downer, et al. Linear colliders based on laser-plasma accelerators. *Journal of Instrumentation*, 18(06):T06001, 2023.
- [19] Yujian Zhao, Weiming An, Xinlu Xu, Fei Li, Lance Hildebrand, Mark J Hogan, Vitaly Yakimenko, Chan Joshi, and Warren B Mori. Emittance preservation through density ramp matching sections in a plasma wakefield accelerator. *Physical Review Accelerators and Beams*, 23(1):011302, 2020.
- [20] Fei Li, Thamine N Dalichaouch, Jacob R Pierce, Xinlu Xu, Frank S Tsung, Wei Lu, Chan Joshi, and Warren B Mori. Ultrabright electron bunch injection in a plasma wakefield driven by a superluminal flying focus electron beam. *Physical Review Letters*, 128(17):174803, 2022.
- [21] Oscar Buneman. Dissipation of currents in ionized media. *Physical Review*, 115(3):503, 1959.

- [22] John Dawson. One-dimensional plasma model. *The Physics of Fluids*, 5(4):445–459, 1962.
- [23] Jay P Boris and Ramy A Shanny. *Proceedings: Fourth Conference on Numerical Simulation of Plasmas, November 2, 3, 1970*. Naval Research Laboratory, 1971.
- [24] Toshiaki Tajima and John M Dawson. Laser electron accelerator. *Physical review letters*, 43(4):267, 1979.
- [25] CE Clayton, KA Marsh, A Dyson, M Everett, A Lal, WP Leemans, R Williams, and C Joshi. Ultrahigh-gradient acceleration of injected electrons by laser-excited relativistic electron plasma waves. *Physical review letters*, 70(1):37, 1993.
- [26] Patrick Mora and Thomas M Antonsen, Jr. Kinetic modeling of intense, short laser pulses propagating in tenuous plasmas. *Physics of Plasmas*, 4(1):217–229, 1997.
- [27] Alexancer Pukhov and Jürgen Meyer-ter Vehn. Laser wake field acceleration: the highly non-linear broken-wave regime. *Applied Physics B*, 74(4):355–361, 2002.
- [28] Stuart PD Mangles, CD Murphy, Zulfikar Najmudin, Alexander George Roy Thomas, JL Collier, Aboobaker E Dangor, EJ Divall, PS Foster, JG Gallacher, CJ Hooker, et al. Monoenergetic beams of relativistic electrons from intense laser–plasma interactions. *Nature*, 431(7008):535–538, 2004.
- [29] CGR Geddes, Cs Toth, J Van Tilborg, E Esarey, CB Schroeder, D Bruhwiler, Ch Nieter, J Cary, and WP Leemans. High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding. *Nature*, 431(7008):538–541, 2004.
- [30] Jérôme Faure, Yannick Glinec, A Pukhov, S Kiselev, S Gordienko, E Lefebvre, J-P Rousseau, F Burgy, and Victor Malka. A laser–plasma accelerator producing monoenergetic electron beams. *Nature*, 431(7008):541–544, 2004.
- [31] Frank S Tsung, Ritesh Narang, Warren B Mori, Chandrasekar Joshi, RA Fonseca, and Luis O Silva. Near-gev-energy laser-wakefield acceleration of self-injected electrons in a centimeter-scale plasma channel. *Physical review letters*, 93(18):185002, 2004.
- [32] K Nakamura, B Nagler, Cs Tóth, CGR Geddes, CB Schroeder, E Esarey, WP Leemans, AJ Gonsalves, and SM Hooker. Gev electron beams from a centimeter-scale channel guided laser wakefield accelerator. *Physics of Plasmas*, 14(5), 2007.
- [33] George Stantchev, William Dorland, and Nail Gumerov. Fast parallel particle-to-grid interpolation for plasma pic simulations on the gpu. *Journal of Parallel and Distributed Computing*, 68(10):1339–1349, 2008.
- [34] Jorge Vieira and JT Mendonça. Nonlinear laser driven donut wakefields for positron and electron acceleration. *Physical Review Letters*, 112(21):215001, 2014.

- [35] WP Leemans, AJ Gonsalves, H-S Mao, K Nakamura, C Benedetti, CB Schroeder, Cs Tóth, J Daniels, DE Mittelberger, SS Bulanov, et al. Multi-gev electron beams from capillary-discharge-guided subpetawatt laser pulses in the self-trapping regime. *Physical review letters*, 113(24):245002, 2014.
- [36] AJ Gonsalves, K Nakamura, J Daniels, C Benedetti, C Pieronek, TCH De Raadt, S Steinke, JH Bin, SS Bulanov, J Van Tilborg, et al. Petawatt laser guiding and electron beam acceleration to 8 gev in a laser-heated capillary discharge waveguide. *Physical review letters*, 122(8):084801, 2019.
- [37] JP Palastro, JL Shaw, P Franke, D Ramsey, TT Simpson, and DH Froula. Dephasingless laser wakefield acceleration. *Physical review letters*, 124(13):134802, 2020.
- [38] C Caizergues, S Smartsev, V Malka, and C Thaury. Phase-locked laser-wakefield electron acceleration. *Nature Photonics*, 14(8):475–479, 2020.
- [39] Fei Li, Weiming An, Viktor K Decyk, Xinlu Xu, Mark J Hogan, and Warren B Mori. A quasi-static particle-in-cell algorithm based on an azimuthal fourier decomposition for highly efficient simulations of plasma-based acceleration: Qpad. *Computer Physics Communications*, 261:107784, 2021.
- [40] Wentao Wang, Ke Feng, Lintong Ke, Changhai Yu, Yi Xu, Rong Qi, Yu Chen, Zhiyong Qin, Zhijun Zhang, Ming Fang, et al. Free-electron lasing at 27 nanometres based on a laser wakefield accelerator. *Nature*, 595(7868):516–520, 2021.
- [41] Marie Labat, Jurjen Couperus Cabadağ, Amin Ghaith, Arie Irman, Anthony Berlioux, Philippe Berteaud, Frédéric Blache, Stefan Bock, François Bouvet, Fabien Briquez, et al. Seeded free-electron laser driven by a compact laser plasma accelerator. *Nature Photonics*, 17(2):150–156, 2023.
- [42] SK Barber, F Kohrell, CE Doss, K Jensen, C Berger, F Isono, Z Eisentraut, S Schröder, AJ Gonsalves, K Nakamura, et al. Greater than 1000-fold gain in a free-electron laser driven by a laser-plasma accelerator with high reliability. *Physical Review Letters*, 135(5):055001, 2025.
- [43] Zhan Jin, Masaki Kando, Yan-Jun Gu, Kai Huang, Nobuhiko Nakanii, Izuru Daito, Zhenzhe Lei, Shingo Sato, Hiroaki Sano, Toshiya Muto, et al. Optimized laser wakefield acceleration: Generating stable, high-energy, monoenergetic electron beams and demonstrating extreme-ultraviolet free-electron lasers. *Physical Review Research*, 8(1):013207, 2026.
- [44] James B Rosenzweig, B Breizman, T Katsouleas, and JJ Su. Acceleration and focusing of electrons in two-dimensional nonlinear plasma wake fields. *Physical Review A*, 44(10):R6189, 1991.

- [45] Kyle G Miller, Jacob R Pierce, Manfred V Ambat, Jessica L Shaw, Kale Weichman, Warren B Mori, Dustin H Froula, and John P Palastro. Dephasingless laser wakefield acceleration in the bubble regime. *Scientific Reports*, 13(1):21306, 2023.
- [46] Jonathan Zuegel, Ani Aprahamian, Eva Zurek, Franklin Dollar, and Antonino Di Piazza. Research Infrastructure: Midscale RI-1 (M1:DP): OMEGA-EP-Pumped Optical Parametric Amplifier Line (EP-OPAL) Facility Design. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2329970, 2024. Award Number: 2329970, accessed February 17, 2025.
- [47] University of Rochester. Optical Parametric Amplifier Line Flagship Experiment Selection Report. <https://nsf-opal.rochester.edu/documentation/>, 2024. Accessed February 17, 2025.
- [48] Viktor Konstantyn Decyk. Computer simulation study of waves in bounded plasma with application to lower hybrid heating. *Ph. D. Thesis*, 1977.
- [49] A Bruce Langdon and JM Dawson. *Investigations of a sheet model for a bounded plasma with magnetic field and radiation*. PhD thesis, Princeton University, 1969.
- [50] RW Hockney. Computer experiment of anomalous diffusion. *The Physics of Fluids*, 9(9):1826–1835, 1966.
- [51] John Myrick Dawson, CG Hsi, and R Shanny. A finite-size particle model for plasma simulation (report matt-719). Technical report, Princeton Univ., NJ (USA). Plasma Physics Lab.; General Electric Co . . . , 1969.
- [52] Charles K Birdsall and Dieter Fuss. Clouds-in-clouds, clouds-in-cells physics for many-body plasma simulation. *Journal of Computational Physics*, 3(4):494–511, 1969.
- [53] RL Morse and CW Nielson. Numerical simulation of the weibel instability in one and two dimensions. *The Physics of Fluids*, 14(4):830–840, 1971.
- [54] A Bruce Langdon and Barbara F Lasinski. Electromagnetic and relativistic plasma simulation models. *Controlled Fusion*, pages 327–366, 1976.
- [55] O Buneman, CW Barnes, JC Green, and DE Nielsen. Principles and capabilities of 3-d, em particle simulations. *Journal of Computational Physics*, 38(1):1–44, 1980.
- [56] A Bruce Langdon and Charles K Birdsall. Theory of plasma simulation using finite-size particles. *The Physics of Fluids*, 13(8):2115–2122, 1970.
- [57] Charles K Birdsall and A Bruce Langdon. *Plasma physics via computer simulation*. CRC press, 2004.

- [58] Jean-Luc Vay, Irving Haber, and Brendan B Godfrey. A domain decomposition method for pseudo-spectral electromagnetic simulations of plasmas. *Journal of Computational Physics*, 243:260–268, 2013.
- [59] Fei Li, Peicheng Yu, Xinlu Xu, Frederico Fiuza, Viktor K Decyk, Thamine Dalichaouch, Asher Davidson, Adam Tableman, Weiming An, Frank S Tsung, et al. Controlling the numerical cerenkov instability in pic simulations using a customized finite difference maxwell solver and a local fft based current correction. *Computer Physics Communications*, 214:6–17, 2017.
- [60] Xinlu Xu, Fei Li, Frank S Tsung, Thamine N Dalichaouch, Weiming An, Han Wen, Viktor K Decyk, Ricardo A Fonseca, Mark J Hogan, and Warren B Mori. On numerical errors to the fields surrounding a relativistically moving particle in pic codes. *Journal of Computational Physics*, page 109451, 2020.
- [61] Fei Li, Kyle G Miller, Xinlu Xu, Frank S Tsung, Viktor K Decyk, Weiming An, Ricardo A Fonseca, and Warren B Mori. A new field solver for modeling of relativistic particle-laser interactions using the particle-in-cell algorithm. *Computer Physics Communications*, 258:107580, 2021.
- [62] Kyle G Miller, John P Palastro, Jessica L Shaw, Fei Li, Frank S Tsung, Viktor K Decyk, C Joshi, and Warren B Mori. Accurate simulation of direct laser acceleration in a laser wakefield accelerator. *Physics of Plasmas*, 30(7), 2023.
- [63] J-L Vay. Simulation of beams or plasmas crossing at relativistic velocity. *Physics of Plasmas*, 15(5), 2008.
- [64] Adam V Higuera and John R Cary. Structure-preserving second-order integration of relativistic charged particle trajectories in electromagnetic fields. *Physics of Plasmas*, 24(5), 2017.
- [65] Fei Li, Viktor K Decyk, Kyle G Miller, Adam Tableman, Frank S Tsung, Marija Vranic, Ricardo A Fonseca, and Warren B Mori. Accurately simulating nine-dimensional phase space of relativistic particles in strong fields. *Journal of Computational Physics*, 438:110367, 2021.
- [66] D Turnbull, S Bucht, A Davies, D Haberberger, T Kessler, JL Shaw, and DH Froula. Raman amplification with a flying focus. *Physical Review Letters*, 120(2):024801, 2018.
- [67] J. P. Palastro, D. Turnbull, S.-W. Bahk, R. K. Follett, J. L. Shaw, D. Haberberger, J. Bromage, and D. H. Froula. Ionization waves of arbitrary velocity driven by a flying focus. *Physical Review A*, 97(3), mar 2018.
- [68] AJ Howard, D Turnbull, AS Davies, P Franke, DH Froula, and JP Palastro. Photon acceleration in a flying focus. *Physical review letters*, 123(12):124801, 2019.

- [69] D. Ramsey, P. Franke, T. T. Simpson, D. H. Froula, and J. P. Palastro. Vacuum acceleration of electrons in a dynamic laser pulse. *Physical Review E*, 102(4), oct 2020.
- [70] Philip Franke, Dillon Ramsey, Tanner T Simpson, D Turnbull, DH Froula, and JP Palastro. Optical shock-enhanced self-photon acceleration. *Physical Review A*, 104(4):043520, 2021.
- [71] Dillon Ramsey, Bernardo Malaca, Antonino Di Piazza, Martin Formanek, Philip Franke, DH Froula, Miguel Pardal, TT Simpson, Jorge Vieira, Kathleen Weichman, et al. Nonlinear thomson scattering with ponderomotive control. *Physical Review E*, 105(6):065201, 2022.
- [72] Martin Formanek, Dillon Ramsey, John P Palastro, and Antonino Di Piazza. Radiation reaction enhancement in flying focus pulses. *Physical Review A*, 105(2):L020203, 2022.
- [73] Martin Formanek, John P Palastro, Marija Vranic, Dillon Ramsey, and Antonino Di Piazza. Charged particle beam transport in a flying focus pulse with orbital angular momentum. *Physical Review E*, 107(5):055213, 2023.
- [74] Martin Formanek, John P Palastro, Dillon Ramsey, Stefan Weber, and Antonino Di Piazza. Signatures of vacuum birefringence in low-power flying focus pulses. *Physical Review D*, 109(5):056009, 2024.
- [75] John M Dawson. Particle simulation of plasmas. *Reviews of modern physics*, 55(2):403, 1983.
- [76] Charles K Birdsall and A Bruce Langdon. *Plasma physics via computer simulation*. CRC press, 2018.
- [77] Agustin F Lifschitz, Xavier Davoine, Erik Lefebvre, Jérôme Faure, Clément Rechatin, and Victor Malka. Particle-in-cell modelling of laser–plasma interaction using fourier decomposition. *Journal of Computational Physics*, 228(5):1803–1814, 2009.
- [78] Adam Davidson, Adam Tableman, Weiming An, Frank S Tsung, Wei Lu, Jorge Vieira, Ricardo A Fonseca, Luís O Silva, and Warren B Mori. Implementation of a hybrid particle code with a pic description in r – z and a gridless description in ϕ into osiris. *Journal of Computational Physics*, 281:1063–1077, 2015.
- [79] Chengkun Huang, VK Decyk, M Zhou, W Lu, WB Mori, JH Cooley, TM Antonsen Jr, B Feng, T Katsouleas, J Vieira, et al. Quickpic: a highly efficient fully parallelized pic code for plasma-based acceleration. In *Journal of Physics: Conference Series*, volume 46, page 190. IOP Publishing, 2006.
- [80] Weiming An, Viktor K Decyk, Warren B Mori, and Thomas M Antonsen Jr. An improved iteration loop for the three dimensional quasi-static particle-in-cell algorithm: Quickpic. *Journal of Computational Physics*, 250:165–177, 2013.

- [81] Fei Li, Weiming An, Frank S Tsung, Viktor K Decyk, and Warren B Mori. Integrating a ponderomotive guiding center algorithm into a quasi-static particle-in-cell code based on azimuthal mode decomposition. *Journal of Computational Physics*, 470:111599, 2022.
- [82] Severin Diederichs, Carlo Benedetti, Axel Huebl, Rémi Lehe, Andrew Myers, Alexander Sinn, J-L Vay, Weiqun Zhang, and Maxence Thévenet. Hipace++: A portable, 3d quasi-static particle-in-cell code. *Computer Physics Communications*, 278:108421, 2022.
- [83] WB Mori, VK Decyk, TC Katsouleas, DW Forslund, RD Ferraro, and E Horowitz. Full-scale numerical experiment to test a future plasma-based accelerator. *Submitted to NSF as part of the US High Performance Computing and Communications Program*, 1992. https://picksc.physics.ucla.edu/reports-notes/NSF_advanced_accel_gr_chall_92.pdf, accessed March 16, 2026.
- [84] J-L Vay. Noninvariance of space-and time-scale ranges under a lorentz transformation and the implications for the study of relativistic interactions. *Physical review letters*, 98(13):130405, 2007.
- [85] Samuel F Martins, RA Fonseca, Wei Lu, Warren B Mori, and LO Silva. Exploring laser-wakefield-accelerator regimes for near-term lasers using particle-in-cell simulation in lorentz-boosted frames. *Nature Physics*, 6(4):311–316, 2010.
- [86] Brendan B Godfrey. Numerical cherenkov instabilities in electromagnetic particle codes. *Journal of Computational Physics*, 15(4):504–521, 1974.
- [87] Brendan B Godfrey. Canonical momenta and numerical instabilities in particle codes. *Journal of Computational Physics*, 19(1):58–76, 1975.
- [88] Xinlu Xu, Peicheng Yu, Samuel F Martins, Frank S Tsung, Viktor K Decyk, Jorge Vieira, Ricardo A Fonseca, Wei Lu, Luis O Silva, and Warren B Mori. Numerical instability due to relativistic plasma drift in em-pic simulations. *Computer Physics Communications*, 184(11):2503–2514, 2013.
- [89] Brendan B Godfrey and Jean-Luc Vay. Numerical stability of relativistic beam multidimensional pic simulations employing the esirkepov algorithm. *Journal of Computational Physics*, 248:33–46, 2013.
- [90] Brendan B Godfrey and Jean-Luc Vay. Suppressing the numerical cherenkov instability in fdtd pic codes. *Journal of Computational Physics*, 267:1–6, 2014.
- [91] Brendan B Godfrey and Jean-Luc Vay. Improved numerical cherenkov instability suppression in the generalized pstd pic algorithm. *Computer Physics Communications*, 196:221–225, 2015.

- [92] Peicheng Yu, Xinlu Xu, Viktor K Decyk, Frederico Fiuza, Jorge Vieira, Frank S Tsung, Ricardo A Fonseca, Wei Lu, Luis O Silva, and Warren B Mori. Elimination of the numerical cerenkov instability for spectral em-pic codes. *Computer Physics Communications*, 192:32–47, 2015.
- [93] Peicheng Yu, Xinlu Xu, Adam Tableman, Viktor K Decyk, Frank S Tsung, Frederico Fiuza, Asher Davidson, Jorge Vieira, Ricardo A Fonseca, Wei Lu, et al. Mitigation of numerical cerenkov radiation and instability using a hybrid finite difference-fft maxwell solver and a local charge conserving current deposit. *Computer Physics Communications*, 197:144–152, 2015.
- [94] Peicheng Yu, Xinlu Xu, Asher Davidson, Adam Tableman, Thamine Dalichaouch, Fei Li, Michael D Meyers, Weiming An, Frank S Tsung, Viktor K Decyk, et al. Enabling lorentz boosted frame particle-in-cell simulations of laser wakefield acceleration in quasi-3d geometry. *Journal of Computational Physics*, 316:747–759, 2016.
- [95] Remi Lehe, Manuel Kirchen, Brendan B Godfrey, Andreas R Maier, and Jean-Luc Vay. Elimination of numerical cherenkov instability in flowing-plasma particle-in-cell simulations by using galilean coordinates. *Physical Review E*, 94(5):053305, 2016.
- [96] Manuel Kirchen, Remi Lehe, Soeren Jalas, Olga Shapoval, Jean-Luc Vay, and Andreas R Maier. Scalable spectral solver in galilean coordinates for eliminating the numerical cherenkov instability in particle-in-cell simulations of streaming plasmas. *Physical Review E*, 102(1):013202, 2020.
- [97] Wei Lu, Chengkun Huang, Miaomiao Zhou, Warren B Mori, and Thomas Katsouleas. Nonlinear theory for relativistic plasma wakefields in the blowout regime. *Physical review letters*, 96(16):165002, 2006.
- [98] W Lu, C Huang, M Zhou, M Tzoufras, FS Tsung, WB Mori, and T Katsouleas. A nonlinear theory for multidimensional relativistic plasma wave wakefields. *Physics of Plasmas*, 13(5), 2006.
- [99] Michail Tzoufras, Wei Lu, Frank S Tsung, Chengkun Huang, Warren B Mori, Thomas Katsouleas, Jorge Vieira, RA Fonseca, and Luis O Silva. Beam loading in the nonlinear regime of plasma-based acceleration. *Physical review letters*, 101(14):145002, 2008.
- [100] Nicholas A Krall, Alvin W Trivelpiece, and KR Symon. Principles of plasma physics. *IEEE Transactions on Plasma Science*, 2(3):196–196, 1974.
- [101] CD Decker, WB Mori, JM Dawson, and T Katsouleas. Nonlinear collisional absorption in laser-driven plasmas. *Physics of Plasmas*, 1(12):4043–4049, 1994.
- [102] MD Acciarri, C Moore, LP Beving, and SD Baalrud. When should pic simulations be applied to atmospheric pressure plasmas? impact of correlation heating. *Plasma Sources Science and Technology*, 33(3):035009, 2024.

- [103] Diogo D Carvalho, Pablo J Bilbao, Warren B Mori, Luis O Silva, and E Paulo Alves. Learning collision operators from plasma phase space data using differentiable simulators. *arXiv preprint arXiv:2601.10885*, 2026.
- [104] K Nanbu and Shigeru Yonemura. Weighted particles in coulomb collision simulations based on the theory of a cumulative scattering angle. *Journal of Computational Physics*, 145(2):639–654, 1998.
- [105] Suzhi Deng, F Tsung, S Lee, Wei Lu, WB Mori, Tom Katsouleas, P Muggli, BE Blue, CE Clayton, C O Connell, et al. Modeling of ionization physics with the pic code osiris. In *AIP Conference Proceedings*, pages 219–223. IOP INSTITUTE OF PHYSICS PUBLISHING LTD, 2002.
- [106] Michaël Touati, Romain Codur, Frank Tsung, Viktor K Decyk, Warren B Mori, and Luis O Silva. Kinetic theory of particle-in-cell simulation plasma and the ensemble averaging technique. *Plasma Physics and Controlled Fusion*, 64(11):115014, 2022.
- [107] Daniel F Gordon, WB Mori, and Thomas M Antonsen. A ponderomotive guiding center particle-in-cell code for efficient modeling of laser-plasma interactions. *IEEE transactions on plasma science*, 28(4):1135–1143, 2000.
- [108] WB Mori. The physics of the nonlinear optics of plasmas at relativistic intensities for short-pulse lasers. *IEEE Journal of Quantum Electronics*, 33(11):1942–1953, 2002.
- [109] WKH Panofsky and WA Wenzel. Some considerations concerning the transverse deflection of charged particles in radio-frequency fields. *Review of Scientific Instruments*, 27(11):967–967, 1956.
- [110] Thamine N Dalichaouch, XL Xu, Adam Tableman, Fei Li, Frank S Tsung, and Warren B Mori. A multi-sheath model for highly nonlinear plasma wakefields. *Physics of Plasmas*, 28(6), 2021.
- [111] Anton Golovanov, I Yu Kostyukov, A Pukhov, and V Malka. Energy-conserving theory of the blowout regime of plasma wakefield. *Physical review letters*, 130(10):105001, 2023.
- [112] Tom Katsouleas. Physical mechanisms in the plasma wake-field accelerator. *Physical Review A*, 33(3):2056, 1986.
- [113] CD Decker, WB Mori, K-C Tzeng, and T Katsouleas. The evolution of ultra-intense, short-pulse lasers in underdense plasmas. *Physics of Plasmas*, 3(5):2047–2056, 1996.
- [114] A Davidson, A Tableman, P Yu, W An, FS Tsung, W Lu, RA Fonseca, and WB Mori. Optimizing laser wakefield acceleration in the nonlinear self-guided regime for fixed laser energy. *arXiv preprint arXiv:1805.08761*, 2019.

- [115] T. Katsouleas, S. Wilks, P. Chen, J. M. Dawson, and J.J. Su. Beam loading in plasma accelerators. *Particle Accelerators*, 22:81–99, 1987.
- [116] S Van der Meer. Improving the power efficiency of the plasma wakefield accelerator. Technical report, CLIC-Note-3, 1985.
- [117] Michail Tzoufras, Wei Lu, Frank S Tsung, Chengkun Huang, Warren B Mori, Tom Katsouleas, Jorge Vieira, RA Fonseca, and LO Silva. Beam loading by electrons in nonlinear plasma wakes. *Physics of Plasmas*, 16(5), 2009.
- [118] T. N. Dalichaouch, X. L. Xu, A. Tableman, F. Li, F. S. Tsung, and W. B. Mori. A multi-sheath model for highly nonlinear plasma wakefields. *Physics of Plasmas*, 28(6):063103, June 2021.
- [119] Yujian Zhao. *Emittance preservation in a plasma wakefield accelerator*. University of California, Los Angeles, 2023; and references therein.
- [120] Kane Yee. Numerical solution of initial boundary value problems involving maxwell’s equations in isotropic media. *IEEE Transactions on antennas and propagation*, 14(3):302–307, 1966.
- [121] Tony D Arber, Keith Bennett, Christopher S Brady, Alistair Lawrence-Douglas, MG Ramsay, Nathan J Sircombe, Paddy Gillies, Roger G Evans, Holger Schmitz, Anthony R Bell, et al. Contemporary particle-in-cell approach to laser-plasma modelling. *Plasma Physics and Controlled Fusion*, 57(11):113001, 2015.
- [122] Rémi Lehe, Manuel Kirchen, Igor A Andriyash, Brendan B Godfrey, and Jean-Luc Vay. A spectral, quasi-cylindrical and dispersion-free particle-in-cell algorithm. *Computer Physics Communications*, 203:66–82, 2016.
- [123] J-L Vay, A Almgren, J Bell, L Ge, DP Grote, M Hogan, O Kononenko, R Lehe, A Myers, C Ng, et al. Warp-x: A new exascale computing platform for beam–plasma simulations. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 909:476–479, 2018.
- [124] Julien Derouillat, Arnaud Beck, Frédéric Pérez, Tommaso Vinci, M Chiaramello, Anna Grassi, M Flé, Guillaume Bouchard, I Plotnikov, Nicolas Aunai, et al. Smilei: A collaborative, open-source, multi-purpose particle-in-cell code for plasma simulation. *Computer Physics Communications*, 222:351–373, 2018.
- [125] Benjamin M Cowan, David L Bruhwiler, Estelle Cormier-Michel, Eric Esarey, Cameron GR Geddes, Peter Messmer, and Kevin M Paul. Characteristics of an envelope model for laser–plasma accelerator simulation. *Journal of Computational Physics*, 230(1):61–86, 2011.

- [126] T Silva, Anton Helm, J Vieira, R Fonseca, and LO Silva. On the use of the envelope model for down-ramp injection in laser-plasma accelerators. *Plasma Physics and Controlled Fusion*, 62(2):024001, 2019.
- [127] Davide Terzani and Pasquale Londrillo. A fast and accurate numerical implementation of the envelope model for laser-plasma dynamics. *Computer Physics Communications*, 242:49–59, 2019.
- [128] Francesco Massimo, Arnaud Beck, Julien Dérouillat, Mickael Grech, Mathieu Lobet, Frédéric Pérez, Imen Zemzemi, and Arnd Specka. Efficient start-to-end 3d envelope modeling for two-stage laser wakefield acceleration experiments. *Plasma Physics and Controlled Fusion*, 61(12):124001, 2019.
- [129] Francesco Massimo, Carlo Benedetti, Davide Terzani, Arnaud Beck, and Brigitte Cros. Modeling laser-wakefield accelerators using the time-averaged ponderomotive approximation in a lorentz boosted frame. *Plasma Physics and Controlled Fusion*, 67(6):065032, 2025.
- [130] B Rus, P Bakule, D Kramer, G Korn, JT Green, J N3vak, M Fibrich, F Batysta, J Thoma, J Naylor, et al. Eli-beamlines laser systems: status and design options. In *High-Power, High-Energy, and High-Intensity Laser Technology; and Research Using Extreme Light: Entering New Frontiers with Petawatt-Class Lasers*, volume 8780, pages 265–276. SPIE, 2013.
- [131] M Tzoufras, C Huang, JH Cooley, FS Tsung, J Vieira, and WB Mori. Simulations of efficient laser wakefield accelerators from 1 to 100gev. *Journal of Plasma Physics*, 78(4):401–412, 2012.
- [132] Thomas C Katsouleas, Warren B Mori, John M Dawson, and Scott C Wilks. Physics of plasmas with short pulse lasers. In *Femtosecond to Nanosecond High-Intensity Lasers and Applications*, volume 1229, pages 98–106. SPIE, 1990.
- [133] FS Tsung, C Ren, LO Silva, WB Mori, and T Katsouleas. Generation of ultra-intense single-cycle laser pulses by using photon deceleration. *Proceedings of the National Academy of Sciences*, 99(1):29–32, 2002.
- [134] M Tzoufras, FS Tsung, WB Mori, and AA Sahai. Improving the self-guiding of an ultraintense laser by tailoring its longitudinal profile. *Physical Review Letters*, 113(24):245001, 2014.
- [135] Anthony E Siegman. *Lasers*. University science books, 1986.
- [136] Andrew Zangwill. *Modern electrodynamics*. Cambridge University Press, 2013.

- [137] M.W. Beijersbergen, R.P.C. Coerwinkel, M. Kristensen, and J.P. Woerdman. Helical-wavefront laser beams produced with a spiral phaseplate. *Optics Communications*, 112(5):321–327, 1994.
- [138] G. A. Siviloglou, J. Broky, A. Dogariu, and D. N. Christodoulides. Observation of accelerating airy beams. *Phys. Rev. Lett.*, 99:213901, Nov 2007.
- [139] Narupon Chattrapiban, Elizabeth A. Rogers, David Cofield, III Wendell T. Hill, and Rajarshi Roy. Generation of nondiffracting bessel beams by use of a spatial light modulator. *Opt. Lett.*, 28(22):2183–2185, Nov 2003.
- [140] Y. Kato, K. Mima, N. Miyanaga, S. Arinaga, Y. Kitagawa, M. Nakatsuka, and C. Yamanaka. Random phasing of high-power lasers for uniform target acceleration and plasma-instability suppression. *Phys. Rev. Lett.*, 53:1057–1060, Sep 1984.
- [141] Tobias Brixner and G Gerber. Femtosecond polarization pulse shaping. *Optics letters*, 26(8):557–559, 2001.
- [142] Andrew M Weiner. Ultrafast optical pulse shaping: A tutorial review. *Optics Communications*, 284(15):3669–3692, 2011.
- [143] S Skupsky, RW Short, T Kessler, RS Craxton, S Letzring, and JM Soures. Improved laser-beam uniformity using the angular dispersion of frequency-modulated light. *Journal of Applied Physics*, 66(8):3456–3462, 1989.
- [144] J Hebling. Derivation of the pulse front tilt caused by angular dispersion. *Optical and Quantum Electronics*, 28(12):1759–1763, 1996.
- [145] Nihal Jhajj, Ilia Larkin, EW Rosenthal, Sina Zahedpour, JK Wahlstrand, and HM Milchberg. Spatiotemporal optical vortices. *Physical Review X*, 6(3):031037, 2016.
- [146] SW Hancock, S Zahedpour, A Goffin, and HM Milchberg. Free-space propagation of spatiotemporal optical vortices. *Optica*, 6(12):1547–1553, 2019.
- [147] S. W. Hancock, S. Zahedpour, and H. M. Milchberg. Mode structure and orbital angular momentum of spatiotemporal optical vortex pulses. *Phys. Rev. Lett.*, 127:193901, Nov 2021.
- [148] H Esat Kondakci and Ayman F Abouraddy. Diffraction-free pulsed optical beams via space-time correlations. *Optics Express*, 24(25):28659–28668, 2016.
- [149] H. Esat Kondakci and Ayman F. Abouraddy. Diffraction-free space–time light sheets. *Nature Photonics*, 11(11):733–740, oct 2017.
- [150] H Kondakci and Ayman F Abouraddy. Optical space-time wave packets having arbitrary group velocities in free space. *Nature communications*, 10(1):1–8, 2019.

- [151] A. Sainte-Marie, O. Gobert, and F. Quéré. Controlling the velocity of ultrashort light pulses in vacuum through spatio-temporal couplings. *Optica*, 4(10):1298, oct 2017.
- [152] Tanner T. Simpson, Dillon Ramsey, Philip Franke, Navid Vafaei-Najafabadi, David Turnbull, Dustin H. Froula, and John P. Palastro. Nonlinear spatiotemporal control of laser intensity. *Opt. Express*, 28(26):38516–38526, Dec 2020.
- [153] Tanner T. Simpson, Dillon Ramsey, Philip Franke, Kathleen Weichman, Manfred Virgil Ambat, David Turnbull, Dustin H. Froula, and John P. Palastro. Spatiotemporal control of laser intensity through cross-phase modulation. *Opt. Express*, 30(6):9878–9891, Mar 2022.
- [154] A. Di Piazza. Unveiling the transverse formation length of nonlinear compton scattering. *Physical Review A*, 103(1), jan 2021.
- [155] Daniel Cruz-Delgado, Stephanos Yerolatsitis, Nicolas K Fontaine, Demetrios N Christodoulides, Rodrigo Amezcua-Correa, and Miguel A Bandres. Synthesis of ultrafast wavepackets with tailored spatiotemporal properties. *Nature Photonics*, 16(10):686–691, 2022.
- [156] Lu Chen, Wenqi Zhu, Pengcheng Huo, Junyeob Song, Henri J Lezec, Ting Xu, and Amit Agrawal. Synthesizing ultrafast optical pulses with arbitrary spatiotemporal control. *Science Advances*, 8(43):eabq8314, 2022.
- [157] Yaron Silberberg. Quantum coherent control for nonlinear spectroscopy and microscopy. *Annual review of physical chemistry*, 60:277–292, 2009.
- [158] DM Villeneuve, SA Aseyev, P Dietrich, M Spanner, M Yu Ivanov, and PB Corkum. Forced molecular rotation in an optical centrifuge. *Physical Review Letters*, 85(3):542, 2000.
- [159] T Brixner, G Krampert, T Pfeifer, R Selle, G Gerber, M Wollenhaupt, O Graefe, C Horn, D Liese, and T Baumert. Quantum control by ultrafast polarization shaping. *Physical review letters*, 92(20):208301, 2004.
- [160] Martin Aeschlimann, Michael Bauer, Daniela Bayer, Tobias Brixner, F Javier García de Abajo, Walter Pfeiffer, Martin Rohmer, Christian Spindler, and Felix Steeb. Adaptive subwavelength control of nano-optical fields. *Nature*, 446(7133):301–304, 2007.
- [161] Halina Rubinsztein-Dunlop, Andrew Forbes, Michael V Berry, Mark R Dennis, David L Andrews, Masud Mansuripur, Cornelia Denz, Christina Alpmann, Peter Banzer, Thomas Bauer, et al. Roadmap on structured light. *Journal of Optics*, 19(1):013001, 2016.
- [162] Kazuhiko Misawa. Applications of polarization-shaped femtosecond laser pulses. *Advances in Physics: X*, 1(4):544–569, 2016.

- [163] C. Benedetti, C. B. Schroeder, E. Esarey, and W. P. Leemans. Quasi-matched propagation of high-intensity, ultra-short laser pulses in plasma channels. *AIP Conference Proceedings*, 1507(1):252–257, 2012.
- [164] I. V. Igumenshchev, D. H. Froula, D. H. Edgell, V. N. Goncharov, T. J. Kessler, F. J. Marshall, R. L. McCrory, P. W. McKenty, D. D. Meyerhofer, D. T. Michel, T. C. Sangster, W. Seka, and S. Skupsky. Laser-beam zooming to mitigate crossed-beam energy losses in direct-drive implosions. *Physical Review Letters*, 110(14), apr 2013.
- [165] DH Froula, TJ Kessler, IV Igumenshchev, R Betti, VN Goncharov, H Huang, SX Hu, E Hill, JH Kelly, DD Meyerhofer, et al. Mitigation of cross-beam energy transfer: Implication of two-state focal zooming on omega. *Physics of Plasmas*, 20(8):082704, 2013.
- [166] Xiaoxia Huang, Dongxia Hu, Wei Zhou, Wanjun Dai, Xuewei Deng, Qiang Yuan, Qihua Zhu, and Feng Jing. Optical zooming based on focusing grating in direct drive icf. *Optics Express*, 24(19):22051–22060, 2016.
- [167] Zhangyu Zhou, Changjun Min, Haixiang Ma, Yuquan Zhang, Xi Xie, Hao Zhan, and Xiaocong Yuan. Time-varying orbital angular momentum in tight focusing of ultrafast pulses. *Opt. Express*, 30(8):13416–13433, Apr 2022.
- [168] Laura Rego, Kevin M Dorney, Nathan J Brooks, Quynh L Nguyen, Chen-Ting Liao, Julio San Román, David E Couch, Allison Liu, Emilio Pisanty, Maciej Lewenstein, et al. Generation of extreme-ultraviolet beams with time-varying orbital angular momentum. *Science*, 364(6447):eaaw9486, 2019.
- [169] Kevin M Dorney, Laura Rego, Nathan J Brooks, Julio San Román, Chen-Ting Liao, Jennifer L Ellis, Dmitriy Zusin, Christian Gentry, Quynh L Nguyen, Justin M Shaw, et al. Controlling the polarization and vortex charge of attosecond high-harmonic beams via simultaneous spin–orbit momentum conservation. *Nature photonics*, 13(2):123–130, 2019.
- [170] Les Allen, Marco W Beijersbergen, RJC Spreeuw, and JP Woerdman. Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes. *Physical review A*, 45(11):8185, 1992.
- [171] H He, NR Heckenberg, and H Rubinsztein-Dunlop. Optical particle trapping with higher-order doughnut beams produced using high efficiency computer generated holograms. *Journal of Modern Optics*, 42(1):217–223, 1995.
- [172] KT Gahagan and GA Jr Swartzlander. Optical vortex trapping of particles. *Optics Letters*, 21(11):827–829, 1996.

- [173] PA Prentice, MP MacDonald, TG Frank, Alfred Cuschieri, GC Spalding, Wilson Sibbett, PA Campbell, and Kishan Dholakia. Manipulation and filtration of low index particles with holographic laguerre-gaussian optical trap arrays. *Optics Express*, 12(4):593–600, 2004.
- [174] Robert M Lorenz, J Scott Edgar, Gavin DM Jeffries, Yiqiong Zhao, David McGloin, and Daniel T Chiu. Vortex-trap-induced fusion of femtoliter-volume aqueous droplets. *Analytical chemistry*, 79(1):224–228, 2007.
- [175] Severin Fürhapter, Alexander Jesacher, Stefan Bernet, and Monika Ritsch-Marte. Spiral phase contrast imaging in microscopy. *Optics Express*, 13(3):689–694, 2005.
- [176] Severin Fürhapter, Alexander Jesacher, Stefan Bernet, and Monika Ritsch-Marte. Spiral interferometry. *Optics letters*, 30(15):1953–1955, 2005.
- [177] Christian Maurer, Alexander Jesacher, Stefan Bernet, and Monika Ritsch-Marte. What spatial light modulators can do for optical microscopy. *Laser & Photonics Reviews*, 5(1):81–101, 2011.
- [178] Alois Mair, Alipasha Vaziri, Gregor Weihs, and Anton Zeilinger. Entanglement of the orbital angular momentum states of photons. *Nature*, 412(6844):313–316, 2001.
- [179] Mehul Malik, Manuel Erhard, Marcus Huber, Mario Krenn, Robert Fickler, and Anton Zeilinger. Multi-photon entanglement in high dimensions. *Nature Photonics*, 10(4):248–252, 2016.
- [180] K Dholakia, NB Simpson, MJ Padgett, and L Allen. Second-harmonic generation and the orbital angular momentum of light. *Physical Review A*, 54(5):R3742, 1996.
- [181] José T Mendonça, Bo Thidé, and Holger Then. Stimulated raman and brillouin backscattering of collimated beams carrying orbital angular momentum. *Physical review letters*, 102(18):185005, 2009.
- [182] S Ali, JR Davies, and JT Mendonca. Inverse faraday effect with linearly polarized laser pulses. *Physical review letters*, 105(3):035001, 2010.
- [183] Genevieve Gariépy, Jonathan Leach, Kyung Taec Kim, Thomas J Hammond, Eugene Frumker, Robert W Boyd, and Paul B Corkum. Creating high-harmonic beams with controlled orbital angular momentum. *Physical review letters*, 113(15):153901, 2014.
- [184] Jorge Vieira, Raoul MGM Trines, Eduardo Paulo Alves, RA Fonseca, JT Mendonça, Robert Bingham, Peter Norreys, and LO Silva. Amplification and generation of ultra-intense twisted laser pulses via stimulated raman scattering. *Nature communications*, 7(1):1–7, 2016.

- [185] Yin Shi, Jorge Vieira, RMGM Trines, Robert Bingham, BF Shen, RJ Kingham, et al. Magnetic field generation in plasma waves driven by copropagating intense twisted lasers. *Physical Review Letters*, 121(14):145002, 2018.
- [186] R Nuter, Ph Korneev, and VT Tikhonchuk. Raman scattering of a laser beam carrying an orbital angular momentum. *Physics of Plasmas*, 29(6):062101, 2022.
- [187] Dustin H. Froula, David Turnbull, Andrew S. Davies, Terrance J. Kessler, Dan Haberberger, John P. Palastro, Seung-Whan Bahk, Ildar A. Begishev, Robert Boni, Sara Bucht, Joseph Katz, and Jessica L. Shaw. Spatiotemporal control of laser intensity. *Nature Photonics*, 12(5):262–265, mar 2018.
- [188] W. J. Tomlinson. Wavelength multiplexing in multimode optical fibers. *Appl. Opt.*, 16(8):2180–2194, Aug 1977.
- [189] James R Leger. External methods of phase locking and coherent beam addition of diode lasers, 1993.
- [190] T.Y. Fan. Laser beam combining for high-power, high-radiance sources. *IEEE Journal of Selected Topics in Quantum Electronics*, 11(3):567–577, 2005.
- [191] Arnaud Brignon. *Coherent laser beam combining*. John Wiley & Sons, 2013.
- [192] Shian Zhou, Frank W. Wise, and Dimitre G. Ouzounov. Divided-pulse amplification of ultrashort pulses. *Opt. Lett.*, 32(7):871–873, Apr 2007.
- [193] Enrico Seise, Arno Klenke, Jens Limpert, and Andreas Tünnermann. Coherent addition of fiber-amplified ultrashort laser pulses. *Optics express*, 18(26):27827–27835, 2010.
- [194] Enrico Seise, Arno Klenke, Sven Breitskopf, Marco Plötner, Jens Limpert, and Andreas Tünnermann. Coherently combined fiber laser system delivering 120 μ j femtosecond pulses. *Optics letters*, 36(4):439–441, 2011.
- [195] Louis Daniault, Marc Hanna, Laurent Lombard, Yoann Zaouter, Eric Mottay, Didier Goular, Pierre Bourdon, Frédéric Druon, and Patrick Georges. Coherent beam combining of two femtosecond fiber chirped-pulse amplifiers. *Optics letters*, 36(5):621–623, 2011.
- [196] Arno Klenke, Enrico Seise, Jens Limpert, and Andreas Tünnermann. Basic considerations on coherent combining of ultrashort laser pulses. *Opt. Express*, 19(25):25379–25387, Dec 2011.
- [197] Marco Kienel, Arno Klenke, Tino Eidam, Martin Baumgartl, Cesar Jauregui, Jens Limpert, and Andreas Tünnermann. Analysis of passively combined divided-pulse amplification as an energy-scaling concept. *Opt. Express*, 21(23):29031–29042, Nov 2013.

- [198] Arno Klenke, Steffen Hädrich, Tino Eidam, Jan Rothhardt, Marco Kienel, Stefan Demmler, Thomas Gottschall, Jens Limpert, and Andreas Tünnermann. Gw peak-power fiber chirped-pulse-amplification system. *Opt. Lett.*, 39(24):6875–6878, Dec 2014.
- [199] Tong Zhou, John Ruppe, Cheng Zhu, I-Ning Hu, John Nees, and Almantas Galvanauskas. Coherent pulse stacking amplification using low-finesse gires-tournois interferometers. *Optics express*, 23(6):7442–7462, 2015.
- [200] Ihsan Fsaifes, Louis Daniault, Severine Bellanger, Matthieu Veinhard, Jerome Bourderionnet, Christian Larat, Eric Lallier, Eric Durand, Arnaud Brignon, and Jean-Christophe Chanteloup. Coherent beam combining of 61 femtosecond fiber amplifiers. *Opt. Express*, 28(14):20152–20161, Jul 2020.
- [201] Dan Wang, Qiang Du, Tong Zhou, Derun Li, and Russell Wilcox. Stabilization of the 81-channel coherent beam combination using machine learning. *Optics express*, 29(4):5694–5709, 2021.
- [202] Arohi Jain, Jiayang Yan, Jacob R Pierce, Tanner T Simpson, Mikhail Polyanskiy, William Li, Marcus Babzien, Mark Palmer, Michael Downer, Roman Samulyak, et al. Collider-quality electron bunches from an all-optical plasma photoinjector. *arXiv preprint arXiv:2503.09557*, 2025.
- [203] Lokenath Debnath and Firdous Ahmad Shah. *Wavelet Transforms and Their Applications*. Birkhäuser Boston, 2015.
- [204] K. Y. Kim, J. H. Glowia, A. J. Taylor, and G. Rodriguez. Terahertz emission from ultrafast ionizing air in symmetry-broken laser fields. *Opt. Express*, 15(8):4577–4584, Apr 2007.
- [205] Matthew R Edwards, Victor T Platonenko, and Julia M Mikhailova. Enhanced attosecond bursts of relativistic high-order harmonics driven by two-color fields. *Optics Letters*, 39(24):6823–6826, 2014.
- [206] Qiwen Zhan. Cylindrical vector beams: from mathematical concepts to applications. *Advances in Optics and Photonics*, 1(1):1–57, 2009.
- [207] Xi-Lin Wang, Yongnan Li, Jing Chen, Cheng-Shan Guo, Jianping Ding, and Hui-Tian Wang. A new type of vector fields with hybrid states of polarization. *Optics Express*, 18(10):10786–10795, 2010.
- [208] Giovanni Milione, H. I. Sztul, D. A. Nolan, and R. R. Alfano. Higher-order poincaré sphere, stokes parameters, and the angular momentum of light. *Physical Review Letters*, 107(5), jul 2011.

- [209] Darryl Naidoo, Filippus S Roux, Angela Dudley, Igor Litvin, Bruno Piccirillo, Lorenzo Marrucci, and Andrew Forbes. Controlled generation of higher-order poincaré sphere beams from a laser. *Nature Photonics*, 10(5):327–332, 2016.
- [210] Carl B Schroeder, Eric Esarey, CGR Geddes, Carlo Benedetti, and Wim P Leemans. Physics considerations for laser-plasma linear colliders. *Physical Review Special Topics—Accelerators and Beams*, 13(10):101301, 2010.
- [211] T Mehrling, J Grebenyuk, FS Tsung, Klaus Floettmann, and Jens Osterhoff. Transverse emittance growth in staged laser-wakefield acceleration. *Physical Review Special Topics—Accelerators and Beams*, 15(11):111303, 2012.
- [212] S Steinke, J Van Tilborg, C Benedetti, CGR Geddes, CB Schroeder, J Daniels, KK Swanson, AJ Gonsalves, K Nakamura, NH Matlis, et al. Multistage coupling of independent laser-plasma accelerators. *Nature*, 530(7589):190–193, 2016.
- [213] Alexander Debus, Richard Pausch, Axel Huebl, Klaus Steiniger, René Widera, Thomas E Cowan, Ulrich Schramm, and Michael Bussmann. Circumventing the dephasing and depletion limits of laser-wakefield acceleration. *Physical Review X*, 9(3):031044, 2019.
- [214] Tanner Nutting, Nicholas Ernst, Alexander GR Thomas, and Karl Krushelnick. Transversely pumped laser driven particle accelerator. *arXiv preprint arXiv:2501.11825*, 2025.
- [215] Aaron Liberman, Anton Golovanov, Slava Smartsev, Sheroy Tata, Igor A. Andriyash, Salome Benracassa, Eitan Y. Levine, Eyal Kroupp, and Victor Malka. First direct observation of a wakefield generated with structured light, 2025.
- [216] Spencer W. Jolly, Olivier Gobert, Antoine Jeandet, and Fabien Quéré. Controlling the velocity of a femtosecond laser pulse using refractive lenses. *Opt. Express*, 28(4):4888–4897, Feb 2020.
- [217] J. J. Pigeon, P. Franke, M. Lim Pac Chong, J. Katz, R. Boni, C. Dorrer, J. P. Palastro, and D. H. Froula. Ultrabroadband flying-focus using an axiparabola-echelon pair. *Opt. Express*, 32(1):576–585, Jan 2024.
- [218] J. L. Shaw, M. V. Ambat, K. G. Miller, R. Boni, I. LaBelle, W. B. Mori, J. J. Pigeon, A. Rigatti, I. Settle, L. Mack, J. P. Palastro, and D. H. Froula. Path to a single-stage, 100-gev electron beam via a flying-focus-driven laser-plasma accelerator, 2025.
- [219] T. Katsouleas. Physical mechanisms in the plasma wake-field accelerator. *Phys. Rev. A*, 33:2056–2064, Mar 1986.
- [220] R A Fonseca, L O Silva, F S Tsung, V K Decyk, W Lu, C Ren, W B Mori, S Deng, S Lee, T Katsouleas, and J C Adam. OSIRIS: A Three-Dimensional, Fully Relativistic Particle in Cell Code for Modeling Plasma Based Accelerators. In Peter M A Sloot,

- Alfons G Hoekstra, C J Kenneth Tan, and Jack J Dongarra, editors, *Computational Science — ICCS 2002: International Conference Amsterdam, The Netherlands, April 21–24, 2002 Proceedings, Part III*, pages 342–351. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.
- [221] AG Khachatryan, A Irman, FA Van Goor, and K-J Boller. Femtosecond electron-bunch dynamics in laser wakefields and vacuum. *Physical Review Special Topics—Accelerators and Beams*, 10(12):121301, 2007.
 - [222] A. Davidson, A. Tableman, W. An, F. S. Tsung, W. Lu, J. Vieira, R. A. Fonseca, L. O. Silva, and W. B. Mori. Implementation of a hybrid particle code with a PIC description in r-z and a gridless description in ϕ into OSIRIS. *Journal of Computational Physics*, 281:1063–1077, 2015.
 - [223] Fei Li, Kyle G. Miller, Xinlu Xu, Frank S. Tsung, Viktor K. Decyk, Weiming An, Ricardo A. Fonseca, and Warren B. Mori. A new field solver for modeling of relativistic particle-laser interactions using the particle-in-cell algorithm. *Computer Physics Communications*, 258:107580, jan 2021.
 - [224] S. Morshed, T. M. Antonsen, and J. P. Palastro. Efficient simulation of electron trapping in laser and plasma wakefield acceleration. *Physics of Plasmas*, 17(6), June 2010.
 - [225] A. Pukhov, Z.-M. Sheng, and J. Meyer ter Vehn. Particle acceleration in relativistic laser channels. *Physics of Plasmas*, 6:2847–2854, 7 1999. This is the main paper that derives the betatron motion equation.
 - [226] Qianqian Su, Jeffrey Larson, Thamine N Dalichaouch, Fei Li, Weiming An, Lance Hildebrand, Yujian Zhao, Viktor Decyk, Paulo Alves, Stefan M Wild, et al. Optimization of transformer ratio and beam loading in a plasma wakefield accelerator with a structure-exploiting algorithm. *Physics of Plasmas*, 30(5), 2023.
 - [227] Bruno Silveira Nunes, Samara Prass dos Santos, Roger Pizzato Nunes, C Bontoiu, MS Alva-Sánchez, Ricardo Elgul Samad, Nilson Dias Vieira Jr, Guoxing Xia, Javier Resta-López, and Alexandre Bonatto. Bayesian optimization of laser wakefield acceleration in the self-modulated regime (sm-lwfa) aiming to produce molybdenum-99 via photonuclear reactions. *arXiv preprint arXiv:2407.20087*, 2024.
 - [228] Pan-Fei Geng, Min Chen, and Zheng-Ming Sheng. Efficient muon acceleration in laser wakefields driven by single or combined laser pulses. *Physics of Plasmas*, 31(2), 2024.
 - [229] J. P. Palastro, B. Malaca, J. Vieira, D. Ramsey, T. T. Simpson, P. Franke, J. L. Shaw, and D. H. Froula. Laser-plasma acceleration beyond wave breaking. *Physics of Plasmas*, 28(1):013109, January 2021.

- [230] Alexander Rainville, Mathew Whittlesey, Christopher Pasquale, Yanwen Jing, Mingshu Chen, Siyun Chen, Hanzhang Pei, John Ruppe, Tong Zhou, Qiang Du, Zhigang Zhang, Guoqing Chang, Franz X. Kärtner, and Almantas Galvanauskas. Near-complete extraction of maximum stored energy from large-core fibers using coherent pulse stacking amplification of femtosecond pulses. *Optica*, 11(11):1540–1548, Nov 2024.
- [231] John D Owens, Mike Houston, David Luebke, Simon Green, John E Stone, and James C Phillips. Gpu computing. *Proceedings of the IEEE*, 96(5):879–899, 2008.
- [232] André R Brodtkorb, Trond R Hagen, and Martin L Sætra. Graphics processing unit (gpu) programming strategies and trends in gpu computing. *Journal of Parallel and Distributed Computing*, 73(1):4–13, 2013.
- [233] Song Huang, Shucai Xiao, and Wu-chun Feng. On the energy efficiency of graphics processing units for scientific computing. In *2009 IEEE International Symposium on Parallel & Distributed Processing*, pages 1–8. IEEE, 2009.
- [234] TOP500. Top500 june 2024 list. <https://top500.org/lists/top500/2024/06/>, 2024. Accessed: 2024-06-17.
- [235] Roger W Hockney and James W Eastwood. *Computer simulation using particles*. crc Press, 2021.
- [236] J Van Dijk, GMW Kroesen, and Annemie Bogaerts. Plasma modelling and numerical simulation. *Journal of Physics D: Applied Physics*, 42(19):190301, 2009.
- [237] Kenichi Nishikawa, Ioana Duțan, Christoph Köhn, and Yosuke Mizuno. Pic methods in astrophysics: Simulations of relativistic jets and kinetic physics in astrophysical systems. *Living Reviews in Computational Astrophysics*, 7(1):1, 2021.
- [238] Heiko Burau, Renée Widera, Wolfgang Hönig, Guido Juckeland, Alexander Debus, Thomas Kluge, Ulrich Schramm, Tomas E Cowan, Roland Sauerbrey, and Michael Bussmann. Picongpu: a fully relativistic particle-in-cell code for a gpu cluster. *IEEE Transactions on Plasma Science*, 38(10):2831–2839, 2010.
- [239] Rejith George Joseph, Girish Ravunnikutty, Sanjay Ranka, Eduardo D’Azevedo, and Scott Klasky. Efficient gpu implementation for particle in cell algorithm. In *2011 IEEE International Parallel & Distributed Processing Symposium*, pages 395–406. IEEE, 2011.
- [240] Xianglong Kong, Michael C Huang, Chuang Ren, and Viktor K Decyk. Particle-in-cell simulations with charge-conserving current deposition on graphic processing units. *Journal of Computational Physics*, 230(4):1676–1685, 2011.
- [241] Viktor K Decyk and Tajendra V Singh. Adaptable particle-in-cell algorithms for graphical processing units. *Computer Physics Communications*, 182(3):641–648, 2011.

- [242] Francesco Rossi, Pasquale Londrillo, Andrea Sgattoni, Stefano Sinigardi, and Giorgio Turchetti. Towards robust algorithms for current deposition and dynamic load-balancing in a gpu particle in cell code. In *AIP Conference Proceedings*, volume 1507, pages 184–192. American Institute of Physics, 2012.
- [243] Guangye Chen, Luis Chacón, and Daniel C Barnes. An efficient mixed-precision, hybrid cpu-gpu implementation of a nonlinearly implicit one-dimensional particle-in-cell algorithm. *Journal of Computational Physics*, 231(16):5374–5388, 2012.
- [244] Sergey Bastrakov, R Donchenko, Arkady Gonoskov, Evgeny Efimenko, A Malyshev, Iosif Meyerov, and Igor Surmin. Particle-in-cell plasma simulation on heterogeneous cluster systems. *Journal of Computational Science*, 3(6):474–479, 2012.
- [245] Viktor K Decyk and Tajendra V Singh. Particle-in-cell algorithms for emerging computer architectures. *Computer Physics Communications*, 185(3):708–719, 2014.
- [246] Andrew Myers, A Almgren, Ligia Diana Amorim, J Bell, Luca Fedeli, Lixin Ge, Kevin Gott, David P Grote, M Hogan, Axel Huebl, et al. Porting warpx to gpu-accelerated platforms. *Parallel Computing*, 108:102833, 2021.
- [247] J-L Vay, A Huebl, A Almgren, LD Amorim, J Bell, L Fedeli, L Ge, K Gott, DP Grote, M Hogan, et al. Modeling of a chain of three plasma accelerator stages with the warpx electromagnetic pic code on gpus. *Physics of Plasmas*, 28(2):023105, 2021.
- [248] Luca Fedeli, Axel Huebl, France Boillod-Cerneux, Thomas Clark, Kevin Gott, Conrad Hillairet, Stephan Jaure, Adrien Leblanc, Rémi Lehe, Andrew Myers, et al. Pushing the frontier in the design of laser-based electron accelerators with groundbreaking mesh-refined particle-in-cell simulations on exascale-class supercomputers. In *2022 SC22: International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, pages 25–36. IEEE Computer Society, 2022.
- [249] Robert Bird, Nigel Tan, Scott V Luedtke, Stephen Lien Harrell, Michela Taufer, and Brian Albright. Vpic 2.0: Next generation particle-in-cell simulations. *IEEE Transactions on Parallel and Distributed Systems*, 33(4):952–963, 2021.
- [250] Erik Zenker, René Widera, Axel Huebl, Guido Juckeland, Andreas Knüpfer, Wolfgang E Nagel, and Michael Bussmann. Performance-portable many-core plasma simulations: Porting picongpu to openpower and beyond. In *High Performance Computing: ISC High Performance 2016 International Workshops, ExaComm, E-MuCoCoS, HPC-IODC, IXPUG, IWOPH, P³MA, VHPC, WOPSSS, Frankfurt, Germany, June 19–23, 2016, Revised Selected Papers 31*, pages 293–301. Springer, 2016.
- [251] Xianglong Kong, MC Huang, and C Ren. Preliminary results on gpu acceleration of the pic simulation code osiris using cuda. In *APS Division of Plasma Physics Meeting Abstracts*, volume 51, pages JP8–138, 2009.

- [252] Adam Ryan Tableman. *Kinetic Plasma Simulation: Meeting the Demands of Increased Complexity*. University of California, Los Angeles, 2019.
- [253] Kyle G Miller, Roman P Lee, Adam Tableman, Anton Helm, Ricardo A Fonseca, Viktor K Decyk, and Warren B Mori. Dynamic load balancing with enhanced shared-memory parallelism for particle-in-cell codes. *Computer Physics Communications*, 259:107633, 2021.
- [254] Kyle G Miller, Joshua May, Frederico Fiuza, and Warren B Mori. Extended particle absorber for efficient modeling of intense laser–solid interactions. *Physics of Plasmas*, 28(11):112702, 2021.
- [255] M Pardal, A Sainte-Marie, A Reboul-Salze, RA Fonseca, and J Vieira. Radio: An efficient spatiotemporal radiation diagnostic for particle-in-cell codes. *Computer Physics Communications*, 285:108634, 2023.
- [256] Marija Vranic, Thomas Grismayer, Joana Luis Martins, Ricardo A Fonseca, and Luis O Silva. Particle merging algorithm for pic codes. *Computer Physics Communications*, 191:65–73, 2015.
- [257] Viktor K Decyk. Skeleton particle-in-cell codes on emerging computer architectures. *Computing in Science & Engineering*, 17(2):47–52, 2015.
- [258] Farah Hariri, Trach-Minh Tran, Andreas Jocksch, Emmanuel Lanti, J Progsch, Peter Messmer, Stephan Brunner, Claudio Gheller, and Laurent Villard. A portable platform for accelerated pic codes and its application to gpus using openacc. *Computer Physics Communications*, 207:69–82, 2016.
- [259] Nigel Tan, Robert Bird, Guangye Chen, and Michela Taufer. Optimize memory usage in vector particle-in-cell (vpic) to break the 10 trillion particle barrier in plasma simulations. In *Computational Science–ICCS 2021: 21st International Conference, Krakow, Poland, June 16–18, 2021, Proceedings, Part II 21*, pages 452–465. Springer, 2021.
- [260] Tomonor Takizuka and Hirotada Abe. A binary collision model for plasma simulation with a particle code. *Journal of computational physics*, 25(3):205–219, 1977.
- [261] E Paulo Alves, Warren B Mori, and Frederico Fiuza. Numerical heating in particle-in-cell simulations with monte carlo binary collisions. *Physical Review E*, 103(1):013306, 2021.
- [262] Andreas Jocksch, Farah Hariri, Trach-Minh Tran, Stephan Brunner, Claudio Gheller, and Laurent Villard. A bucket sort algorithm for the particle-in-cell method on manycore architectures. In *Parallel Processing and Applied Mathematics: 11th International Conference, PPAM 2015, Krakow, Poland, September 6-9, 2015. Revised Selected Papers, Part I 11*, pages 43–52. Springer, 2016.

- [263] Philipp Mertmann, Denis Eremin, Thomas Mussenbrock, Ralf Peter Brinkmann, and Peter Awakowicz. Fine-sorting one-dimensional particle-in-cell algorithm with monte-carlo collisions on a graphics processing unit. *Computer Physics Communications*, 182(10):2161–2167, 2011.
- [264] Guy E Blelloch. Prefix sums and their applications. Technical Report CMU-CS-90-190, School of Computer Science, Carnegie Mellon University Pittsburgh, PA, USA, November 1990.
- [265] Ricardo A Fonseca, Jorge Vieira, Frederico Fiúza, Asher Davidson, Frank S Tsung, Warren B Mori, and Luis O Silva. Exploiting multi-scale parallelism for large scale numerical modelling of laser wakefield accelerators. *Plasma Physics and Controlled Fusion*, 55(12):124011, 2013.
- [266] Erich S Weibel. Spontaneously growing transverse waves in a plasma due to an anisotropic velocity distribution. *Physical Review Letters*, 2(3):83, 1959.
- [267] Luis O Silva, RA Fonseca, JW Tonge, JM Dawson, WB Mori, and MV Medvedev. Interpenetrating plasma shells: near-equipartition magnetic field generation and nonthermal particle acceleration. *The Astrophysical Journal*, 596(1):L121, 2003.
- [268] Ricardo A Fonseca, Luis O Silva, John W Tonge, Warren B Mori, and John M Dawson. Three-dimensional weibel instability in astrophysical scenarios. *Physics of Plasmas*, 10(5):1979–1984, 2003.
- [269] W. Lu, M. Tzoufras, C. Joshi, F. S. Tsung, W. B. Mori, J. Vieira, R. A. Fonseca, and L. O. Silva. Generating multi-GeV electron bunches using single stage laser wakefield acceleration in a 3D nonlinear regime. *Physical Review Special Topics - Accelerators and Beams*, 10(6):061301, 6 2007.
- [270] Zhengji Zhao, Ermal Rrapaj, Sridutt Bhalachandra, Brian Austin, Hai Ah Nam, and Nicholas Wright. Power analysis of nersc production workloads. In *Proceedings of the SC’23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis*, pages 1279–1287, 2023.
- [271] Kyle G Miller, Jacob R Pierce, Fei Li, Brandon K Russell, Warren B Mori, Alexander GR Thomas, and John P Palastro. Photon acceleration of high-intensity vector vortex beams into the extreme ultraviolet. *Communications Physics*, 8(1):229, 2025.
- [272] D Ramsey, A Di Piazza, M Formanek, P Franke, DH Froula, B Malaca, WB Mori, JR Pierce, TT Simpson, J Vieira, et al. Exact solutions for the electromagnetic fields of a flying focus. *Physical Review A*, 107(1):013513, 2023.
- [273] E Paulo Alves and Frederico Fiuza. Data-driven discovery of reduced plasma physics models from fully kinetic simulations. *Physical Review Research*, 4(3):033192, 2022.